Jurnal
Teknologi

Full Paper

CLIMATE AND FESTIVAL EFFECTS ON PENANG ROAD ACCIDENTS

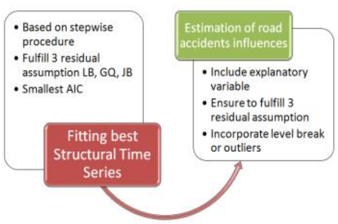
Noor Wahida Md Junus[a,b*], Mohd Tahir Ismail[b], Zainudin Arsad[b]

[a]Mathematics Department, Sultan Idris Education University
[b]School of Mathematical Sciences, Universiti Sains Malaysia

Article history
Received
25 October 2015
Received in revised form
14 December 2015
Accepted
9 February 2016

*Corresponding author
noor_wahida@fsmt.upsi.edu.my

Graphical abstract



Abstract

Road accidents have become the fifth main cause of death in Malaysia in 2008 as reported by the Department of Statistics. The causes and trends should be investigated to prevent reoccurrence in the future. The purpose of this study is to identify the pattern of occurrence of road accidents and subsequently investigate the climate and festival effects on road accidents in Penang based on structural time series analysis. Structural time series analysis offers the possibility of discovering the stochastic behaviour of road accidents. The climate, festival, and intervention effects were incorporated in investigating their influences on the occurrence of road accidents. The study found that road accidents in Penang can be represented by a stochastic level with a fixed seasonal and were influenced by the climate and the intervention effects. The study should be enhanced by applying the model to another state with other relevant variables, such as economic factor and school holiday effect.

Keywords: Road accidents influences; climate effect; festival effects;structural time series; stochastic behaviour

1.0 INTRODUCTION

Traffic accidents are such a nightmare to all the road users. The Department of Statistics in 2008 reported that traffic accidents have become the fifth main cause of death. Reducing the number of traffic accidents is the aim of developing countries, including Malaysia. Penang is the third largest economic contributor among the states in Malaysia is no exception in contributing to the increasing number of road accidents. Various efforts are implemented to reduce the number of road accidents. In the early 1980s, road accidents had been a subject of interest by many researchers.

Numerous researches have been conducted to identify the factors that influence the increasing number of road accidents. Accident frequency can be considered as count data. Some researchers tend to employ generalised linear modelling, such as the Poisson and negative binomial model [1]–[5]. On the other hand, some researchers tend to employ the most simplest cross-sectional model, that is, a linear regression analysis [6]–[8].However, the assumption of this method may not satisfy for series data. It is suggested that time series analysis, such as the Box and Jenkins analysis, be used. Researchers who applied this method were [9], [10], and [11]. In [9] Scott, compared the performance of Box-Jenkins and the linear regression approach in explaining the reduction of road accident frequency in Great Britain. The study found that in terms of simplicity, regression is more preferable, but the Box and Jenkins approach fits the series better. On the other hand, [10] studied the effect of the recent economic crisis and the motorcycle safety programme on motorcycle-related accidents. In [11] Yaacob et al. examined the effectiveness of the implementation of new intervention policies, such as Ops Sikap, towards road accidents in Malaysia.

Meanwhile, for cross-sectional time series data, the most popular method employed is panel data analysis. This analysis is the combination of cross-sectional data and time series data. This model has been applied by [12] to determine the effectiveness road infrastructure in reducing fatalities and injuries. On the other hand, [13] incorporated meteorological variable and traffic exposure as a potential contributor in predicting the effect of environmental factor on road safety. The study found that Poisson integer value autoregressive gave a higher likelihood ratio compared to other competing model and the weather conditions were significantly related to the crash count. Other researchers who used the similar method are [14]–[16], who applied it on Malaysian road accident series.

Commonly, achieving stationarity series condition is compulsory before performing time series analysis. The process of achieving stationarity condition involved detrending and deseasonalizing the series. Important information regarding the series might be lost during the process of detrending and deseasonalizing. On the other hand this problem might be solved by using the Structural time series approach. This approach does not account the stationarity condition but it is model the trend and seasonal series. Structural time series has been used as a model and forecast in a variety of applications, such as financial time series, macroeconomic time series, and many others areas, such as medicine, biology, engineering, and marketing. The advantage of modelling an unobserved component, such as trend and seasonal, and allowing it to vary overtime made structural time series the most possible method to understand the pattern of occurrences of road accidents and at the same time identify the most influential factor contributing to road accident occurrences.

However, the application of structural time series in the road safety industry is very rare, especially in Malaysia. The earliest application of structural time series in road safety can be found in [17]. Therefore, the objective of this paper is to identify the pattern of road accident occurrence in Penang, Malaysia, by using structural time series analysis and subsequently to investigate the factor that contributes to this increasing number of road accidents. The rest of this paper will discuss the analysis approach used in this study followed by the estimation procedure, result and finally the discussion and conclusion of the study.

2.0 STRUCTURAL TIME SERIES APPROACH

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, which can be defined as sums of trend μ_t, seasonal, γ_t and irregular, ε_t components.

$$y_t = \mu_t + \gamma_t + \varepsilon_t, \qquad \varepsilon_t \sim NID(0,\sigma_\varepsilon^2), \qquad t = 1,\ldots,T \tag{1}$$

Instead of using a fixed time series component as in classical decomposition, the structural time series approach allowed the time series component to move in a stochastic manner. The stochastic trend component is specified as

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

where v_t is the slope of trend component, μ_t. The irregular (ε_t), level (ξ_t), and slope disturbances(η_t) are mutually uncorrelated. In addition, the slope component can be excluded when it is appropriate. Excluding the slope component yields the local level model as follows:

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim \mathrm{NID}(0,\sigma_\varepsilon^2), \\ \mu_t &= \mu_{t-1} + \xi_t, & \xi_t &\sim \mathrm{NID}(0,\sigma_\xi^2) \end{aligned} \tag{3}$$

Instead of treating all the components into a stochastic process, the time series component can be treated as either deterministic or stochastic, whichever is appropriate. Fixing the variance disturbance to zero made the time series component corresponding to the variance disturbance reduced in deterministic form. Some special level and trend specifications that were used in this study were in Table 1.

Some series were influences of the seasonal effect. Incorporating the seasonal effect as a stochastic process can be in dummy form or trigonometric form. In this study, a dummy seasonal form was preferred, and the stochastic seasonal can be written as follows

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

where s denotes the number of seasons, and ω_t denotes the seasonal disturbance term that is independent of all disturbance terms. In contrast with the level and the slope components that only require one state equation, seasonal equation will require an $s-1$ state equation. Since the study involved monthly data, eleven state equations are needed. The first equation is as defined above, while another ten remaining equations are identities that can be interpreted as follows

$$\gamma_{s,t} = \gamma_{s-1,t} \tag{5}$$

Similar to level and slope components, the seasonal component also can be treated as deterministic whenever appropriate by fixing the seasonal disturbance, $\sigma_\omega^2 = 0$.

Table 1 Special level and trend specification

Level Without Slope	s_e	s_h	s_V
Deterministic Level	*	0	
Local Level	*	*	
Random Walk	0	*	
With Slope	s_e	s_h	
Deterministic Linear Trend	*	0	0
Local Level with Drift	*	*	0
Local Linear Trend	*	*	*
Smooth Trend	*	0	*

*Indicate any positive value

2.1 Adding Explanatory and Intervention Variable

Relating the particular series with the time series component may not be enough to describe the pattern of certain time series. It is because some of the series may be influenced by some other external factor. Incorporating the influences of an external factor in structural time series to investigate their effect on particular series is developed by adding explanatory and intervention variables into the measurement equation as follows:

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

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

2.2 Statistical Treatment

The structural time series approach is based on the linear Gaussian model, which is often referred as the state space model. The state space form relates an observed variable to an unobserved component representing the various time series components, such as level, trend, and seasonal. The observed variable, y_t, is related to the unobserved component, α_t, through a measurement equation as written below.

$$y_t = Z_t \alpha_t + \varepsilon_t \qquad \varepsilon_t \sim \text{NID}(0, H_t) \tag{7}$$

Meanwhile, the unobserved component follows the first order autoregressive process as given by the following state equation:

$$\alpha_t = T_t \alpha_{t+1} + R_t \eta_t \qquad \eta_t \sim \text{NID}(0, Q_t) \tag{8}$$

where ε_t and η_t are vector of disturbances, and Z_t, T_t, H_t, R_t and Q_t are system matrices that need to be estimated and usually treated as constant. The dimensions of vector and system matrices are shown as in Table 2.

Table 2 Dimension of state space model

Vector		Matrix	
	$p \times 1$	Z_t	$p \times m$
	$m \times 1$	T_t	$m \times m$
	$p \times 1$	H_t	$p \times p$
s_V	$r \times 1$	R_t	$m \times r$
		Q_t	$r \times r$

Appropriate choice of vectors Z_t, α_t and η_t and of the matrices T_t, H_t, R_t and Q_t of the structural time series model can be derived as a special case of (7) and (8). The formulation based on the state space model is important for estimation purposes since this is the routine mechanical procedure for updating estimates based on the Kalman filter. Details on the Kalman filter estimation procedure can be obtained from [18].

2.3 Statistical Treatment

The stepwise fashion is used to identify the best fitted structural time series of road accidents model. The best structural time series model is chosen based on the smallest Akaike Information Criterion (AIC). Similar to other regression and time series analysis, it is customary to diagnose the estimated residual to fulfill certain assumptions. The three assumptions for disturbances, ε_t, that need to be fulfilled are independence, homoscedasticity, and normality, which are diagnosed using Ljung-Box (LB), Goldfeld-Quandt (GQ), and Jarque-Bera (JB) tests, respectively. In addition, if the normality assumption is not satisfied, an auxiliary residual is plotted to ensure the occurrence of any outliers and structural break. Further details on the residual diagnostic procedure is discussed in [19] and [20].

3.0 DATA DESCRIPTION

Most of the variable interests for road safety research comprise of the number of injuries, number of casualties, and frequency of road accidents. In this study, monthly frequency or the monthly number of road accident occurrences in Penang is considered as a dependent variable or the subject of interest. The number of road accidents was obtained from Royal Malaysia Police (RMP). RMP has defined road accidents as follows:

"The occurrence of the accidents on the public or private road due to the negligence or omission by any party concerned (on the aspect of road users conduct, maintenance of vehicle and road condition) or due to environmental factor (excluding natural disaster) resulting in collision (including out of control cases and collision or victim in vehicle against object inside or outside the vehicle eg: bus passenger) which involved at least a moving vehicle, structure or animal and is recorded by the police"

There are seven variables that were considered as possible factors that influence the road accident occurrence. It is includes of four climate factor, one festival effect and two intervention or legislation factor. The climate factors that are considered in this study include of monthly average of the amount of rainfall (in millilitre) (MRAIN), number of rainy days (DRAIN), monthly maximum temperature (in degrees Celsius) (TEMP), and air pollution index (API) for Penang. Majority of the data were based on the Monthly Statistical Bulletin and Compendium of Environmental Statistics, which are published by the Department of Statistics Malaysia, while some were obtained from the main source that is responsible for compiling the environmental data in Malaysia, that is the Department of Meteorology. Daily rainfall was considered if the amount of rainfall recorded is equal or exceeds 0.1mm. Penang has at least five rainy days per month. The detailed description of climate effect from 2001–2013 was in Table 3.

Table 3 Descriptive statistics of climate effect variable

Variable	Min	Max	Mean	Std. Dev
MRAIN	50.9	1544.5	379.84	228.420
DRAIN	5.0	29	14.99	5.155
TEMP	29.9	33.9	31.90	0.773
API	30	121	65	16.153

API is the value that was calculated based on the average concentration of each air pollutant, such as SO2, NO2, CO, O3, and PM10. Air pollutant with the highest concentration accounted will determine the API. Typically, concentration of a fine particulate matter (PM10) is the highest compared with other pollutants, and this determines the API. The API can be categorised as good if the index is between 0 and 50, moderate if the index is between 51 and 100, unhealthy if the index is between 101 and 200, very unhealthy if the index falls between 201 and 300, and hazardous if the index is more than 300. Around 2001 to 2004, Penang has recorded the highest air pollution index. It is expected, since Penang is known as an industrial state.

Some of the climate effects included in this study, such as amount of rainfall, number of rainy days, temperature, and air pollution index, were already used in road safety modelling literature, such as in[9], [13], [15], [16], and [21]–[23]. It was reviewed that this factor has some influences on road accident occurrence as mentioned by [24], which stated that visibility can be reduced to 50 meters during heavy rain as it also does in snow and thick fog. On the other hand, extreme temperature tends to cause harmful effects on driver's performance, road infrastructure, and vehicle component.

A festival celebration usually related to the religious activities invites more road accidents. Traffic suddenly becomes heavier during this season, with people taking advantage of travelling to their hometown to visit their relatives. Such festivals that invite heavier traffic volume are Chinese New Year, Eid ul-Fitr, and Deepavali. Unfortunately, the dates of celebration are not fixed every year, and it is moved from year to year. A study by [1] has incorporated these festivals in measuring their effect on motorcycle accidents. They applied a dummy variable to represent this event. It is coded "0" to represent non festival season and "1" to represent festival season. It is quite relevant since the study was based on weekly data. Since this study used monthly data, incorporating festival variable in terms of a dummy variable will be biased. In this case, this study applied one weight variable for a moving holiday as suggested by [25]. However, in this study, festival variable were name as return to village (RTV) only considered three main festivals, that is, Chinese New Year, Eid ul-Fitr, and Deepavali.

The other data that also has been considered in the study is the enforcement of road safety and new traffic legislation. The data included is the enforcement of rear seat belt law (BELT) and safety operation (SAFE). The compulsory rear seat belt traffic legislation was enforced on January 2009. The representation for this variable will be put into a dummy variable where "0" represents the period before rear seat belt law was enforced and "1" the period after the enforcement, which is from January 2009.

Table 4 An Example of SAFE coding

Year	Duration	Month	Code
2001	9 Dec–23 Dec	12	1
2002	5 Feb–19 Feb	2	1
	29Nov–13 Dec	11	2/15
		12	13/15
2003	25 Jan–8 Feb	1	7/15
		2	8/15
	18 Nov–2Dec	11	13/15
		12	2/15

SAFE is a traffic safety operation carried out by Royal Malaysia Police to nurture safety awareness on all roads in Malaysia during festival seasons such as OPS Sikap or Operation of Attitude and OPS Selamat or Operation of Safety. The SAFE variable has been used by [11] to examine the effect of SAFE on road accidents in Malaysia. The study implements the dummy variable "0" to represent no SAFE and "1" to represent SAFE operation. However, it is felt that this notation was not relevant if the dates of SAFE involve two consecutive months. In that case, this study suggests the use of the weight variable for SAFE where the representation of the SAFE variable is based on the number of days of operation carried out per month divided by the total number of days the operation is carried out, which is 12 to15 days. For example, in 2002, the Ops Sikap III, which is begun from 29 November until 13 December, shows that the operation involves two days in November and 13 days in December. In that case, the rate of operation carried out on November equals 0.13 and on December equals 0.87. While other months that do not involve SAFE are coded as "0" to represent no SAFE. The illustration of the data coding is in Table 4. Each variable considered in this study covers the period of January 2001 until December 2013.

4.0 ESTIMATION RESULT

Developing the structural time series model for road accidents in Penang involves two stages of estimation. In the first stage, estimation of the best fitted structural time series is carried out. It is followed by investigating the influence of climate, festival, and intervention policies on road accident occurrence based on the best model fitted in the first stage.

Fitting the best structural time series for road accidents in Penang is estimated based on the stepwise method. The estimation begins by fitting the road accidents series with the deterministic level model. This model fixed the level disturbance to zero and yielded the estimation results as in table 5.

Table 5 Estimation for deterministic level model

Variance Disturbance	value	q-ratio	
Level, σ_ξ^2	0.0000	0.0000	
Irregular, σ_ε^2	0.0183	1.000	
Residual Diagnostic	**Value**	**p-value**	**Assumption Satisfied**
Independence, LB(6)	307.46	<0.0001	-
Homoscedasticity, GQ	3.9262	<0.0001	-
Normality, JB	6.4532	0.0401	-
AIC	-3.9315		

Since all residual assumptions were not satisfied, estimation was preceded by fitting the road accident series to the local level model, and estimation was as in table 6. When the component level was allowed to vary overtime, the irregular component was lower than the previous model, and it seemed that the observation disturbance was much closer to random values. However, referring to the Ljung-Box test at lags 6 the heterocedasticity test was significant, which indicates that the local level model still was not the best fitted model for Penang road accident series. On the other hand, the model has improved compared to previous models.

Table 6 Estimation for stochastic level model

Variance Disturbance	value	q-ratio	
Level, σ_ξ^2	0.0002	0.0890	
Irregular, σ_ε^2	0.0018	1.0000	
Residual Diagnostic	**Value**	**p-value**	**Assumption Satisfied**
Independence, LB(6)	18.829	0.0026	-
Homoscedasticity,GQ	0.6929	0.9032	+
Normality, N	12.574	0.0019	-
AIC	-5.9585		

The estimation proceeds by adding the slope component to the local level model, which yields estimation for local linear trend model as in table 7. The estimation shows that slope component was fixed with the most important residual assumption, which is that residual autocorrelation was not satisfied.

Table 7 Estimation for local linear trend model

Variance Disturbance	value	q-ratio	
Level, σ_{ξ}^{2}	7.2×10^{-5}	0.0386	
Slope, σ_{ν}^{2}	0.0000	0.0000	
Irregular, σ_{ε}^{2}	0.0019	1.0000	
Residual Diagnostic	**Value**	**p-value**	**Assumption Satisfied**
Independence, LB(6)	13.025	0.0112	-
Homoscedasticity,GQ	0.7208	0.8772	+
Normality, N	12.942	0.0015	-
AIC	-5.9268		

The model also shows that observation disturbance in this model has increased little bit compared with the previous models. On the other hand, this model indicates that adding the slope component has not improved the model. In that case, the estimation was preceded by adding the seasonal component to the local level model by excluding the slope components. The estimation results were as in table 8. The estimated model indicated that level and slope allowed to vary overtime with the observation disturbance seem to be close to the independent random values. In addition, it was reflected by the Ljung-Box test at both nonseasonal and seasonal lags that it is no longer significant. At this stage, the local level with seasonal model was the best model compared to other structural time series that have been estimated. However, referring to the seasonal disturbance value that is too small indicated that the seasonal component has rarely changed overtime.

Table 8 Estimation for local level seasonal model

Variance Disturbance	value	q-ratio	
Level, σ_{ξ}^{2}	0.0002	0.1850	
Seasonal, σ_{ω}^{2}	4.8×10^{-8}	5.2×10^{-5}	
Irregular, σ_{ε}^{2}	0.0009	1.0000	
Residual Diagnostic	**Value**	**p-value**	**Assumption Satisfied**
Independence, LB(6)	2.8086	0.5820	+
Independence, LB(12)	13.678	0.1882	+
Homoscedasticity,GQ	1.0545	0.4270	+
Normality, N	0.82316	0.6626	+
AIC	-5.6890		

To ensure the best fitted model of the road accident series, the local level with fixed seasonal must be fitted. The estimate of the model was as in table 9. By fixing the seasonal component, the estimated model shows little changes, but the AIC value has improved better than in previous models. In that case, the local level with fixed seasonal model fits road accidents much better.

Table 9 Estimation for local level fixed seasonal model

Variance Disturbance	value	q-ratio	
Level, σ_{ξ}^{2}	0.0002	0.1830	
Seasonal, σ_{ω}^{2}	0.0000	0.0000	
Irregular, σ_{ε}^{2}	0.0009	1.0000	
Residual Diagnostic	**Value**	**p-value**	**Assumption Satisfied**
Independence, LB(6)	2.9156	0.5720	+
Independence, LB(12)	13.747	0.1849	+
Homoscedasticity,GQ	1.0587	0.4221	+
Normality, N	0.8374	0.6579	+
AIC	-5.7017		

After all, to ensure the best fitted structural time series for Penang road accidents, the model once again fits by adding a trend component into the model that yields local linear trend with seasonal as in table 10. Even though all the residual assumption has been satisfied, adding the slope component to the model does not improve the model. It clearly can be seen through the irregular disturbance term that shows an increasing value compared to local level seasonal and local level with fixed seasonal. Also, the AIC value in the estimated model was higher than in previous models. After all, from this analysis, it clearly indicated that the best model to fit the Penang road accidents series is local level with fixed seasonal.

Table 10 Estimation for local linear trend seasonal model

Variance Disturbance	value	q-ratio	
Level, σ_{ξ}^{2}	9.8×10^{-5}	0.1010	
Slope, σ_{ν}^{2}	0.0000	0.0000	
Seasonal, σ_{ω}^{2}	4.6×10^{-8}	4.7×10^{-5}	
Irregular, σ_{ε}^{2}	0.0010	1.0000	
Residual Diagnostic	**Value**	**p-value**	**Assumption Satisfied**
Independence, LB(6)	1.1338	0.7689	+
Independence, LB(12)	10.634	0.3016	+
Homoscedasticity,GQ	1.0923	0.3817	+
Normality, N	0.3714	0.8305	+
AIC	-5.6452		

Figure 1 shows the stochastic level estimate for Penang road accidents. The observation was lower and above the estimated level. While referring to the seasonal component plot for this model, Figure 2 shows that the seasonality of the number of road accidents in Penang was fixed, and the lowest number of road accidents usually occurred during February while the highest number of road accidents was in August. In addition, the irregular component (Figure 3) clearly indicates that observation disturbances of the model were independently distributed. Figure 4 illustrates the stochastic level combined with the deterministic seasonal component for the road accident model in Penang. This plot shows that the estimated value recovered quite well with the stochastic level and the fixed seasonal model.

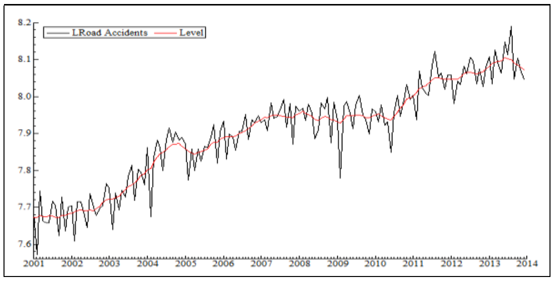


Figure 1 Stochastic level component

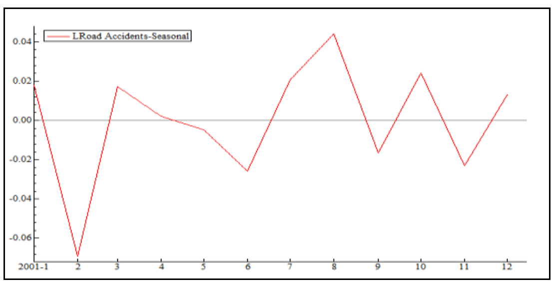


Figure 2 Fixed seasonal component for January–December 2001

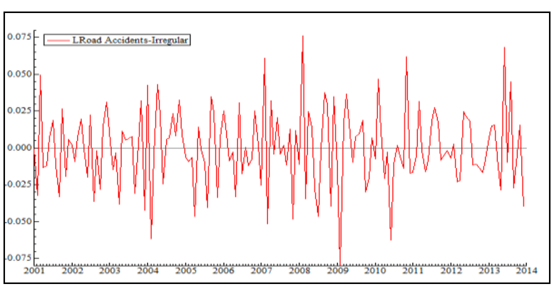


Figure 3 Irregular components

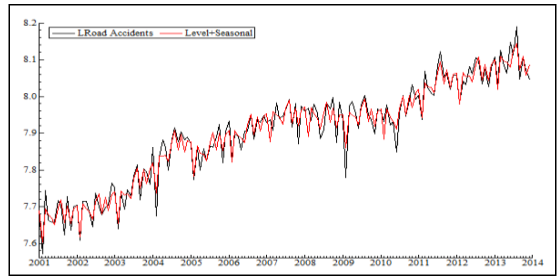


Figure 4 Stochastic level combined with fixed seasonal

Even the univariate model looks like the time series component has explained the model quite well, the value of R^2 is quite small, which is only 38.3 percent. In that order, the explanatory and intervention variables were added, and the estimated model was in Table 11.

The estimated model shows that after adding the explanatory variable, the irregular variance disturbance has reduced to 0.0008. This estimate clearly indicated that incorporating the explanatory variable has accounted the residual in the estimated univariate model previously. The irregular component in the current model was illustrated in Figure 5. The residual in early 2004 has disappeared, but there are few larger residuals that still have no improvement. Referring to the residual diagnostic, Table 11 shows that the normality residual assumption was not satisfied at 5% significant level. In [20], Commandeur and Koopman stated that this assumption is the least important and might be because of the outliers and structural break. To ensure the occurrence of outliers and level break, the auxiliary residual that has smoothed estimates for irregular and level disturbances for road accidents in Penang is plotted in Figure 6. This figure shows that the highest peak in smooth estimates for irregular disturbance plot that determines the outliers is in February 2008 and in the level disturbance plot that determines the level break possibly occurs in November 2010. Based on this information, a new model was developed by incorporating the outliers and structural break, and the model can be summarised as in Table 12.

Table 11 Parameter estimation with explanatory variable

Variance Disturbance	Coefficient	q-ratio
Level, σ_ξ^2	0.0002	0.2638
Seasonal, σ_ω^2	0.0000	0.0000
Irregular, σ_ε^2	0.0008	1.0000
Residual Diagnostic	**Coefficient**	**p-value**
Independence, LB(6)	3.6252	0.4591
Independence, LB(12)	13.898	0.1777
Homoscedasticity,GQ	1.0833	0.3948
Normality, JB	1.3456	0.5103

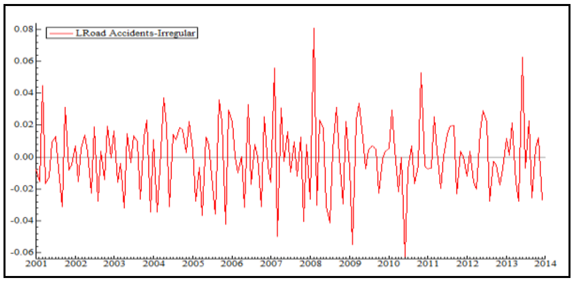


Figure 5 Irregular component after adding explanatory variable

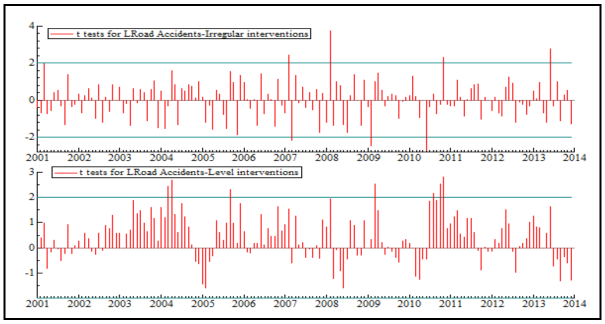


Figure 6 Standardised smoothed observation disturbance(top) and standardised level disturbances (bottom)

Incorporating level break and outlier shows that the level disturbance has reduced, while irregular disturbance shows not much difference with the model without outliers and level break. Figure 7 shows the irregular component of the model. The difference of randomness between Figure 3, Figure 5, and Figure 7 is noticeable. A large residual in the previous model has disappeared, and the observation disturbance was close to the random values. It is proven by the diagnostic residual that all assumptions have been satisfied and subsequently assures that the current model is the best model to represent road accident series in Penang state.

The current estimated model shows that Penang has positively related to all variables incorporated except the API, BLKG, and BELT, which have a negative relationship with road accident occurrence. The result contradicts with previous research by [26], which found that TEMP has a negative relationship with Penang road accidents. It is assumed that it might be because, in the previous study, other climate effects, such as API and DRAIN, that may influence indirectly the Penang road accidents were not incorporated. However, road accidents in Penang are only significantly related to the MRAIN, TEMP, and OPSKP. The result is expected, as there is an increase in the amount of rainfall in Penang; the number of road accidents will increase up to 1 percent. The result agrees with the study found by [14]–[16], which show the highest correlation between road accidents and the amount of rainfall.

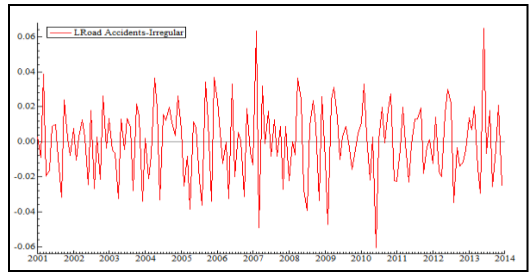


Figure 7 Irregular components with outliers and level break

Table 12 Parameter estimation with outliers and level break

Variance Disturbance	Coefficient	q-ratio
Level, σ_{ξ}^2	0.0001	0.214
Seasonal, σ_{ω}^2	0.0000	0.000
Irregular, σ_{ε}^2	0.0007	1.000
Parameter	**Coefficient**	**p-value**
Outlier 2008(2)	0.1403	0.00006
Level break 2010(11)	0.0909	0.00125
MRAIN	0.0122	0.07199
DRAIN	0.0013	0.19943
TEMP	0.0229	0.00195
API	-3.0×10^{-5}	0.91094
BLKG	-0.0061	0.52229
OPSKP	0.0358	0.00516
BELT	-0.0293	0.27324
Residual Diagnostic	**Coefficient**	**p-value**
Independence, LB(6)	1.1024	0.8939
Independence, LB(12)	7.4820	0.6793
Homoscedasticity,GQ	0.8889	0.6499
Normality, JB	1.6651	0.4349

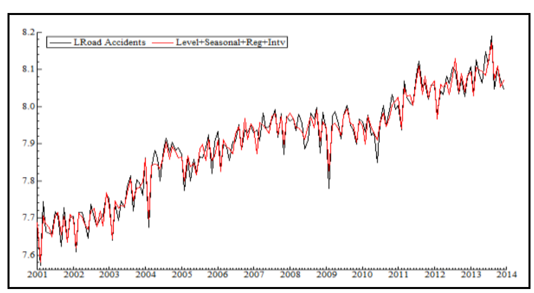


Figure 8 Combined stochastic level, fixed seasonal, explanatory, and intervention effect

In addition, it is shown that increase in temperature in Penang will increase the number of road accidents by approximately 1 percent. It agrees with [27], which states that a road accident is more likely to occur during hot temperature since, during this time, driving performance can possibly worsen because of psychological and physiological effects of ambient temperature. In terms of the operation of OPSKP, the study shows that this intervention failed to decrease the number of road accidents. This result agrees with the previous study by [26] and is expected since the Ops Sikap is only implemented during the festival seasons, when the traffic volume become much higher, as Malaysians take full advantage of going back to their hometown, which has contributed to a higher number of road accidents. At the same time, this result denies the study by [11], which states that the implementation of OPSKP reduces the number of road accidents.

On the other hand, Penang road accidents were related to the outliers in the month of February in 2008 and the level break around November 2010. Interpretation of this variable may be quite crucial since the event that falls on that month may be unknown. Outliers that fall in February 2008 may be related to the long weekend holiday that fall on Chinese New Year that falls on 7 and 8 February, which is on Thursday and Friday. It is assumed that the traffic is quite heavier during this period because citizens have at least four days of weekend holiday. Level break in November 2010 may be related to the double festival celebration, that is, Deepavali and Eid ul-Adha, as well as the year-end school holiday. Figure 8 shows the estimated model of Penang road accident series. It is shown that all observations were recovered quite well with the estimated value. R^2 increases up to 57.8 percent indicates that 57.8 percent total variation of Penang road accidents was explained by climate, festival, and intervention effects, as well as outliers and level break.

5.0 CONCLUSION

The study applied structural time series analysis in the investigation of climate, festival, and intervention effects on road accident occurrences in Penang state. The structural time series is used, as this method offers possibilities to model time series component in terms of the stochastic component. The study found that Penang road accident series can be respresented by local level with a fixed seasonal model. This model allowed the level component to vary overtime while seasonal component rarely changed overtime. Road accidents were more likely to occur during August and quite low during February. Increasing road accidents in Penang were related to the increasing amount of rainfall, the temperature, and the operation of Ops Sikap.

However, this study has a few limitations since it is developed only for the case of Penang state, and it may not reflect the scenario of road accidents in other states in Malaysia. Also, the model may not accurately represent the number of road accidents, as the data only represented reported cases of road accidents while there might be more unreported cases. Further investigation may include other states with other relevant variables, such as economic factor and holiday effect.

Acknowledgement

The authors would express their greatest appreciation to the School of Mathematical Science USM, for funding this study. The authors also wish to thank Royal Malaysia Police, Malaysia Meteorologist Department and Department of Statistics Malaysia for providing the data for this study.

References

[1] Radin Umar, R. S., G. M. Murray, and L.H. Brian. 1996. Modelling of Conspicuity- Related Motorcycle Accidents in Seremban and Shah Alam, Malaysia. *Accident Analysis and Prevention.* 28(3): 325–332.

[2] Sarani, R., S. M. S. Allyana, M. Jamilah, and S.V. Wong. 2012. Predicting Malaysian Road Fatalities for Year 2020. Department Of Statistics, Kuala Lumpur.

[3] Greibe, P. 2003. Accident Prediction Models for Urban Roads. *Accident Analysis and Prevention.* 35(2003): 273-285.

[4] Harnen, S., S. V. Wong, and W. I. W. Hashim. 2006. Motorcycle Accident Prediction Model for Junctions on Urban Roads in Malaysia. *Advances in Transportation Studies an International Journal. 8: 31-40*

[5] Abusini, S. 2013. The Effect of Road Characteristics on Motorcycle Accident in Batu East Java Indonesia. *International Conference on Mathematical Sciences and Statistics, (ICSMS) 2013*. Kuala Lumpur, Malaysia. 5-7 February 2013. 241–246.

[6] Zlatoper, T. J. 1984. Regression Analysis of Time Series Data on Motor Vehicle Death in the United States. *Journal of Transport Economics and Policy.* 18(3): 263–273.

[7] Ali, G.A. and C.S. Bakheit. 2011. Comparative Analysis and Prediction of Traffic Accidents in Sudan Using Artificial Neural Netwrok and Statistical Method. *Proceedings of the 30th Southern African Transport Conference (SATC)*. Pretoria, South Africa. 11-14 July 2011. 202–214.

[8] Desai, M. M., and A.K. Patel. 2011. Road Accidents Study Based On Regression Model: A Case Study of Ahmedabad City. *National Conference On Recent Trends In Engeneering & Technology*. Gujarat, India. 13-14 May 2011. 1–8.

[9] Scott, P. P. 1986. Modelling Time Series of British Road Accidents Data. *Accident Analysis and Prevention.* 18(2):109–117.

[10] Law, T.H., R.S. Radin Umar, S. Zulkaurnain, and S. Kulanthayan. 2005. Impact of the Effect of Economic Crisis and the Targeted Motorcycle Safety Programme on Motorcycle-Related Accidents, Injuries and Fatalities in Malaysia. *International Journal of Injury Control and Safety Promotion.* 12(1): 9–21

[11] Wan Yaacob, W. F., W. Z. Wan Husin, N. Abd Aziz, and N.I. Nordin. 2011. An Intervention Model of Road Accidents: The Case of OPS Sikap. *Journal of Applied Science.* 11(7): 1105–1112.

[12] Noland, R. B. 2003. Traffic Fatalities and Injuries: The Effect of Changes In Infrastructure and Other Trends. *Accident Analysis and Prevention.* 35 (4): 599–611.

[13] [13] Brijs, T., D. Karlis, and D. Wets. 2008. Studying The Effect of Weather Conditions On Daily Crash Counts Using

A Discrete Time-Series Model. *Accident Analysis and Prevention.* 40(3): 1180–90.

[14] Wan Yaacob, W.F., M.A. Lazim, and Y.B. Wah. 2010. Evaluating Spatial and Temporal Effects of Accidents Likelihood using Random Effects Panel Count Model. *International Conference on Science and Social Research (CSSR)*. Kuala Lumpur, Malaysia. 5 - 7 December 2010. 960–964.

[15] Wan Yaacob, W.F., M.A. Lazim, and Y.B. Wah. 2011. Applying Fixed Effect Panel Count Model to Examine Road Accident Occurence. *Journal of Applied Science.* 11(7): 1185–1191.

[16] Wan Yaacob, W. F., M.A. Lazim, and Y.B. Wah. 2012. Modeling Road Accidents using Fixed Effects Mode: Conditional versus Unconditional Model. *Proceeding of the World Congress on Engineering* (WCE). London, UK. 4-6 July 2012. 218-223

[17] Harvey, A. C., and J. Durbin. 1986. The Effects of Seat Belt Legislation on British Road Casualties: A Case Study in Structural Time Series Modelling. *Journal of Royal Statistical Society.*149 (3): 187–227.

[18] Durbin, J., and S.J. Koopman. 2012. *Time Series Analysis by State Space Method*. 2nd edition. London: Oxford University Press.

[19] Koopman, S. J., A. C. Harvey, J. A. Doornik, and N. Shephard. 2009. *Structural Time Series Analyser, Modeller and Predictor STAMP 8.2*. London: Timberlake Consultants Ltd.

[20] Commandeur, J. J. F., and S. J. Koopman. 2007. *An Introduction to State Space Time Series Analysis*. United State: Oxford University Press.

[21] Abbas, K.A. 2004. Traffic Safety Assessment and Development of Predictive Models for Accidents On Rural Roads In Egypt. *Accident Analysis and Prevention.* 36(2004): 149–163.

[22] Keay, K., and I. Simmonds. 2006. Road Accidents and Rainfall In a Large Australian City. *Accident Analysis and Prevention.* 38 (2006): 445–454.

[23] Amin, M. S. R., A. Zareie, and L. E. Amador-Jiménez. 2014. Climate Change Modeling And The Weather-Related Road Accidents In Canada. *Transportation Research Part D: Transport and Environment.* 32(2014): 171–183.

[24] SWOV. 2012. SWOV Fact Sheet- The Influence of Weather on Road Safety. Leidschendam, Netherlands. [Online]. From:http://www.swov.nl/rapport/Factsheets/UK/FS_Influence_of_weather.pdf. [Acessed on 7 April 2015].

[25] Shuja, N., M. A. Lazim, and Y. B. Wah. 2007.Moving Holiday Effects Adjustment for Malaysian Economic Time Series. *Journal of the Department of Statistics, Malaysia.* 1(2007):36-50.

[26] Md Junus, N. W., M. T. Ismail, and Z. Arsad. 2014. Road Accidents Model: Time Series Regression versus Structural Time Series. The *3rd International Conference on Computer Science and Computational Mathematics*. Langkawi, Malaysia. 8-9 May 2014. 32–39.

[27] Mohd Jawi, Z., M.H. Md Isa, R. Sarani, W.S. Voon, and M. A. F. Sadullah. 2009. *Weather as a Road Safety Hazard in Malaysia - An Overview*, MRev 03/2009, Kuala Lumpur: Malaysian Institute of Road Safety Research.