

Predicting Penang Road Accidents Influences: Time Series Regression Versus Structural Time Series

Noor Wahida Md Junus^{1*}, Mohd Tahir Ismail² and Zainudin Arsad²

¹Department of Mathematics, Sultan Idris Education University, Tanjung Malim – 35900, Malaysia;
noor_wahida@fsmi.upsi.edu.my

²School of Mathematical Sciences, Universiti Sains Malaysia, Minden – 11800, Malaysia;
m.tahir@usm.my, zainudin.arsad@usm.my

Abstract

Road accidents become a major cause of death in recent years. The trends of road accidents keep increasing from year to year. Therefore, this paper will compare two time series methods that are time series regression and structural time series of modeling road accidents in Penang from year 2001 until 2013. There are eight explanatory variables used to model the monthly number of road accidents. In the structural time series, level component is allowed to vary over time, while seasonal components are fixed. Results show that both models have a different view in term of the relationship and the factor contributing to road accidents. However the advantage goes to structural time series method in fitting the best model.

Keywords:

1. Introduction

Rapid growth of population and development of economies in most developing countries indirectly has created the additional burden to road safety. Along with the advantage of this progress indirectly has increase the vehicle usage among people and increase the volume of traffic that is the main contributor of road accidents. Malaysia is one of develop countries that contribute the highest number of road fatalities per 100,000 population among the ASEAN countries¹. Penang which are the third largest economic contributor among the state in Malaysia have no exception in contributing the increasing number of road accidents. Various initiative has been set up in ways to reduce this unexpected events. It is including conspicuity programme, road enhancement programme, road behavioural change and accident prevention and reduction programme². However, the figure of road accidents occurrence were keep increasing year by year. Therefore the study aim to investigate the possible factor that influence road accidents occurrence in Penang

state of Malaysia by comparing two statistical method that is time series regression and structural time series.

Several reason were related to the road accidents occurrence. The structure of the road, vehicle, environment and road user attitude were among possible factor that influence the occurrence. Hence, several methods including statistical and mathematical methods were introduce too investigate this possible influential. The earliest study of road accidents can be found in Smeeds³ that relate the vehicle registration and population with the number of road accidents. Later on, Andreassen⁴ found that the origin of Smeed's formul cannot be applied universally to all countries. In addition this approach has been applied many other reaseacher such as Valli⁵ that compare Smeeds and Andreassen equation and Nasarudin et al.⁶ that compare the performance of Smeeds methods with linear regression.

Linear regression is the most simple model that relate dependent with independent variable that gave satisfactory performance in model development. Some of the recent studies that use this prediction techniques is

*Author for correspondence

Zlatoper⁷ that relates the annual series of total motor vehicle death and occupant death and pedestrian death with few explanatory variable. The study founds that income, driving speed, vehicle size and type of driving were significantly related to the motorcycle death. In addition using similar method Desai and Patel⁸ relate the hourly traffic volume with the average of hourly fatal accidents and total accidents. In the same way, Ali and Bakheit⁹ compare the performance of linear regression with artificial neural network. The study founds that artificial neural network were closed to the actual value compared to regression model.

Generalized linear model is more suitable for count data. As a number of road accidents is considered as count data, some of the study prefer to use this approach. The earliest Malaysia road accidents study that is by Radin Umar et al.¹⁰ applied poisson regression with the aim to determine the effectiveness of introducing the running head light campaign. This study found that this intervention has reduced the conspicuity related accidents by about 29%. Recently Sarani et al.¹¹ use the similar method in determining the effectiveness of enforcement of rear seat belt. In their study, it is found that rear seat belt intervention has successfully reduce the number of people getting severe and slight injuries. Other researcher that use the similar method is Greibe¹², Harnen et al.¹³, Abusini et al.¹⁴ and Abdul Manan et al.¹⁵.

On the other hands, application of those approach as mentioned above may not suitable for series data. Therefore, Box Jenkin time series analysis such as Autoregressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA) model were more preferred in modeling road accidents occurrence as in¹⁶⁻¹⁸. In addition, incorporated both cross sectional and series data, longitudinal or panel data analysis were more preferred in modeling road accidents study as in¹⁸⁻²². Commonly, achieving stationarity condition is one of the compulsory rule in application of time series analysis. However, stationarity condition in real series data may not always be easy to achieve. Detrending and deseasonalizing process may a good way of achieving stationarity, but some important information during the process of detrending and deseasonalizing may be loss.

For that reason structural time series were more preferable. Structural time series is another time series analysis classes that does not account the stationarity condition and will be model all the time series component instead

of removing it as in Box and Jenkin analysis. 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 in²³. Therefore, the effectiveness of this approach were ensure by comparing it with time series regression that is another model that were preferred by the researcher. 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. Methodology

This section will discussed general description on the methodology of the study and the data that have been used in this study.

2.1 Time Series Regression

Regression technique is normally used to investigate the relationship between variables. The variables used in regression analysis can be categorized as dependent and independent variables. The dependent variable is variable of interest that is normally influence the independent variables. While the regression technique was developed for cross sectional data, the technique has been employed for analysis involving time series variables. The general form of the time series regression involving k variable can be written as in the Equation (1).

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_k x_{kt} + \varepsilon_t \quad (1)$$

Where y_t is variable of interest, and $x_{1t}, x_{2t}, \dots, x_{kt}$ are the possible factor that might affect the dependent variable y_t and ε_t is an error term. The regression model above may take two forms; a simple regression which involves only one independent variable and multiple regressions involving more than one independent variable. Also, the

regression model may also include the lag of dependent and independent variables.

The parameters of the model are estimated by using Ordinary Least Square (OLS) method and a few assumptions pertaining the model and error term are examined. Durbin Watson (DW) test statistic is used to examine the serial correlation in the error term and lag of dependent variable or independent variable is included in the model as remedy to the serial correlation problem. Multicollinearity is tested by using Variance Inflation Factor (VIF) where the multicollinearity is said to exist if the VIF is greater than 10. For complete discussion of time series regression technique refer to^{24,25}.

2.2 Structural Time Series

Traditionally, the representation of time series data can be treated as sum of trend, seasonal and irregular components.

$$y_t = \mu_t + \gamma_t + \varepsilon_t \quad t = 1, \dots, n \quad (2)$$

Where y_t denote the observation of interest, μ_t are linear deterministic trend consisting of level and slope component, γ_t and ε_t are seasonal and irregular component which are also often treated deterministically. Application of this formulation representing time series process has a limitation since the model does not allowed the components to evolve over time²³. The structural time series model allows the components of time series process to vary overtime.

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

$$y_t = Z_t \alpha_t + \varepsilon_t \quad \varepsilon_t \sim NID(0, H_t) \quad (3)$$

Meanwhile, the unobserved component follow first order autoregressive process as given by the following state equation.

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t \quad \eta_t \sim NID(0, Q_t) \quad (4)$$

Where ε_t and η_t are vector of disturbances and Z_t , T_t , R_t , H_t , and Q_t system matrices which need to estimated and usually treated as constant. The $(n \times 1)$

observation y_t , consist of n observations and α_t is unobserved state vector $(m \times 1)$. The $(n \times 1)$ irregular vectors of, ε_t has a zero mean and $(n \times n)$ covariance matrix H_t . The $(n \times m)$ matrix Z_t link the observed vector y_t with the unobservable state vector α_t and the measurement equation may consist of regression variable. The $(m \times m)$ transition matrix T_t , determines the dynamic evolution of state vector. The state disturbance Z_t is $(m \times r)$ matrix while $(r \times 1)$ disturbance vector η_t has zero mean and covariance matrix Q_t . The observation and state disturbance ε_t and η_t are assumed to be uncorrelated with each other at all time period.

By appropriate choices of vector α_t the matrices Z_t and T_t , wide range of time series model can be written in the state space form of (3) and (4). The time series model in (2) can be written in state space form to allow evolvement in each of its components. The structural time series model with local level with drift (deterministic slope) and local linear trend with seasonal (say period s) are written as (5) and (6) below.

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \lambda_t + \xi_t \end{aligned} \quad (5)$$

$$\begin{aligned} y_t &= \mu_t + \gamma_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \lambda_{t-1} + \xi_t \\ \lambda_t &= \lambda_{t-1} + w_t \\ \gamma_t &= -\gamma_{1,t} - \gamma_{2,t} \dots - \gamma_{s,t} + \omega_t \\ \gamma_{s,t} &= \gamma_{s-1,t} \end{aligned} \quad (6)$$

To investigate the effect of the selected independent variable, the explanatory variable and intervention variable can be added to the measurement equation of the model. The effect of selected independent variable x_1, x_2, \dots, x_k can be investigated based on the following measurement equation.

$$\begin{aligned} y_t &= \mu_t + \gamma_t + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \dots + \beta_k x_{k,t} + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \lambda_{t-1} + \xi_t \\ \lambda_t &= \lambda_{t-1} + w_t \\ \gamma_t &= -\gamma_{1,t} - \gamma_{2,t} \dots - \gamma_{s,t} + \omega_t \\ \gamma_{s,t} &= \gamma_{s-1,t} \end{aligned} \quad (7)$$

The structural time series in (5), (6), (7) can be estimated using iterative procedure of Kalman filter technique. Details on Kalman filter estimation procedure

can be obtained from²⁶. The step wise fashion is used to develop 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 others regression and time series analysis, it is customary to diagnose the estimated residual to fulfill certain assumption. 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. For comparison purposes the similar test were applied for time series regression.

2.3 Prediction Accuracy

Prediction accuracy for both models will be evaluated by using three famous loss function measure that is Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). The equations for each measure are given as in Equation (8).

$$\begin{aligned}
 RMSE &= \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}}, \\
 MAPE &= \frac{\sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|}{n}, \\
 MAE &= \frac{\sum_{t=1}^n |y_t - \hat{y}_t|}{n}
 \end{aligned} \tag{8}$$

Where y_t and \hat{y}_t actual observed value and predicted value respectively, and n is the number of predicted values.

2.4 Data Considered

The data considered in this study include monthly number of Road Accidents (RDAC) from January 2001 until December 2013. This data are obtained from Royal Malaysia Police Department referred as an occurrence 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 police²⁷. The monthly number of road accidents datasets act as dependent variable for this study.

Independent variables for this study will include the climatic variable, oil price, law enforcement, and school holiday. Data for climatic variable were obtained from Malaysian Meteorological Department and Department of Environmental. The climate factors that were considered in this study include monthly average amount of rainfall (in milliliter) (MRAIN), number of rainy days (DRAIN), monthly average temperature (in degree Celsius) (TEMP) and air pollution index (API) in Penang. The climatic variables had been most commonly used in previous work in road safety modeling literature^{16,20,22,28,29}. While, the data for crude oil price (in Malaysian Ringgit per Barrel) (OILP) are access from The World Bank website. The price is simple average of three spot prices that is Date Brent, West Texas Intermediate and Dubai Fateh.

The other data that also have been considered is the enforcement of the road safety and new traffic legislation. The data that was included is enforcement of rear seat belt law (BELT) and Ops Sikap (OPSKP). The compulsory of rear seat belt traffic legislation were enforced since January 2009. Ops Sikap or Ops attitude is traffic safety operation carried out by Royal Malaysia Police to nurture the safety awareness on all roads in Malaysia during festive seasons. It is began during December 2001 and only operated during main festival season such Eid ul-Fitr, Deepavali and Chinese New Year. Both variable are presented as dummy variable where “0” represent before enforcement and “1” represent after enforcement.

Another variable that will be included in this study are school holiday (SCHOL) one of the calendar effect variables. The SCHOL variable is included with the aim to look whether school holiday have an impact on road accident occurrence. Utusan Malaysia 2012, claimed that road fatalities increase up to 18-20 per day on school holiday compared to not more than 15 death on school days³⁰. In contrast²⁸ claimed that during school holiday traffic volume in Australia reduce 3.4% of the main daily volume compared to school days. Similar with law and legislation variable, SCHOL also presented by dummy variable that “1” if the school holiday is occurred and “0” if no school holiday.

3. Model Estimation Results

This section discusses the results of model estimation using both time series regression and structural time series

model. Recall that in this study the dependent variable, y_t is the number of road accidents (RDAC) in Penang state, while the independent variable are average temperature (TEMP), average rainfall (RAIN), crude oil price (OILP), enforcement of rear seat belt law (BELT), operation of Ops SIKAP (OPSKP) and school holiday (SCHOOL).

3.1 Road Accidents Model with Time Series Regression

The result of the estimation of the time series regression model using OLS method is given in Table 1. Trend and seasonal variable is included as to consider time variable. Trend variable is time variable, $t=1, \dots, 156$ and seasonal variable is dummy of month variable where month 12 as the reference or as the benchmark.

Table 1 shows the estimated model fit the model fairly to the data with the value of R^2 of 93.5%. Multicollinearity

Table 1. Estimation result of the multiple linear regression

	Coefficient	Significance	VIF
Constant	7.2970	0.0000*	
Trend	0.0040	0.0000*	16.379
MRAIN	0.0084	0.3288	3.663
DRAIN	0.0013	0.3118	4.801
TEMP	0.0108	0.1506	3.839
API	-0.0008	0.0012*	1.975
OILP	-6.2×10^{-5}	0.4850	7.466
BELT	-0.0944	0.0000*	5.763
OPSKP	0.0186	0.0305*	1.498
SCHOL	-0.0079	0.4522	3.087
Seasonal_1	0.0229	0.1448	2.128
Seasonal_2	-0.0819	0.0000*	3.066
Seasonal_3	0.0107	0.5427	2.725
Seasonal_4	-0.0181	0.3794	3.700
Seasonal_5	-0.0140	0.4206	2.630
Seasonal_6	-0.0257	0.1271	2.452
Seasonal_7	0.0137	0.4756	3.185
Seasonal_8	0.0338	0.0383*	2.285
Seasonal_9	-0.0401	0.0204*	2.549
Seasonal_10	-0.0128	0.0016*	3.340
Seasonal_11	-0.0511	0.0007*	2.193

DW=1.4911 $R^2=0.9351$

*Indicate significance at 10% significance level

does not exist since the values of VIF are less than 10 except for trend variable. However the DW statistic of 1.467 shows a serial correlation in the estimated residual as the value is lower than lower limit of 1.55. The lag of the dependent variable is included in the model as remedial to the serial correlation. The result of the new estimated model with lag dependent is given in Table 2. The trend variables were excluded as it was show the multicollinearity was exist.

Table 2 shows that the serial correlation problems occur in previous model was removed with the inclusion of lag dependent variable. The value of R^2 show that approximately 91% total variation of road accidents occurrence were explained by the selected variable. The estimated model shows that an increase in the rainfall, number of rainy day, temperature and the oil price results in an increase in the number of road accidents. The finding that the amount of rainfall increases the number of

Table 2. Estimation result of the multiple linear regression with lag

	Coefficient	Significance	VIF
Constant	1.7732	0.0028*	
MRAIN	0.0098	0.3289	3.6748
DRAIN	0.0016	0.2901	5.0013
TEMP	0.0121	0.1723	3.8410
API	-0.0001	0.7145	1.9462
OILP	0.0003	0.0006*	5.1758
BELT	0.0239	0.0340*	2.4498
OPSKP	0.0159	0.1142	1.5155
SCHOL	-0.0080	0.5174	3.0730
Seasonal_1	-0.0236	0.2129	2.1042
Seasonal_2	-0.1308	0.0000*	3.0621
Seasonal_3	0.0172	0.4114	2.7640
Seasonal_4	-0.0727	0.0030*	3.7017
Seasonal_5	-0.0540	0.0083*	2.5814
Seasonal_6	-0.0646	0.0011*	2.3800
Seasonal_7	-0.0091	0.6815	3.1321
Seasonal_8	-0.0202	0.2818	2.2322
Seasonal_9	-0.1018	0.0000*	2.6079
Seasonal_10	-0.0243	0.2922	3.3714
Seasonal_11	-0.0879	0.0000*	2.2749
LRA_1	0.7143	0.0000*	6.1847

$R^2=0.9102$ DW=2.5660

*Indicate significance at 10% significance level

road accidents is as expected since more accidents tend to occur during rainy days.

In addition, surprisingly the estimated model suggests that the number of road accidents also increases after the enforcement of rear seat belt law and as the operation of Ops Sikap took place during the festival period. Meanwhile the model indicates that the number of road accidents decreases with the increases in API and during school holiday. Table 2 shows that only enforcement of rear seatbelt law and increase in oil price has significant effect on the number of road accidents in Penang.

Figure 1 shows the fitted value obtained from the time series regression and the actual values of the number of road accidents. Generally the plot shows that, the fitted values change very similar to the observed values. While, Figure 2 show the residual plot of time series regression model. The plot show there is serial correlation in the residual.

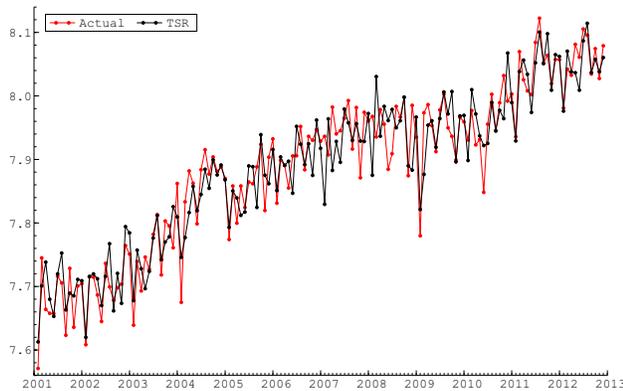


Figure 1. Estimated model of time series regression.

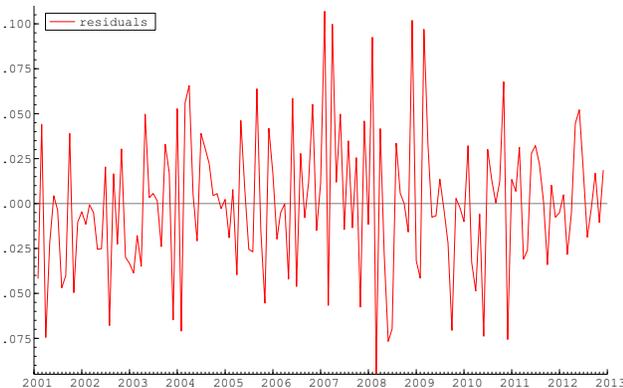


Figure 2. Estimated residual of time series regression model.

3.2 Road Accidents Model with Structural Time Series

The structural time series models are developed by using stepwise procedure. The analysis begins with simple univariate local level model, followed by local linear trend model and finally adding the seasonal component with both models. All the model were estimated without the independent and lag dependent variables. The variance disturbances are estimated by using the maximum likelihood estimation approach, while the state components are estimate using the Kalman filter technique. Table 3 presents the estimated variance disturbance and the main diagnostic test for each structural time series.

Referring to Table 3 the first model develop that is local level model show that the estimated residual does not fulfilled the residual assumption of serial autocorrelation and equal variance. Adding the slope component that is yield local linear trend model also show no improvement. The AIC value of the model is increasing compared to local level model. Besides, the estimated residual assumption of independence once again does not fulfilled. Since the slope component does not improve the model, local level with seasonal were develop. This model allow level and seasonal component to vary overtime while excluding the slope component. This model show little improvement where all the estimated residual have fullfild the structural time series residual assumption. However the seasonal disturbance value is too small that indicate that seasonal component rarely change overtime. In that case the similar model is develop by fixing the seasonal disturbance. The result show that the model was improved as the AIC value, were among the lowest with all the estimated residual were satisfied the residual assumption. To ensure the local level with fixed seasonal model is the best model to present Penang road accidents, the slope component once again were incorporated. Although the estimated residual satisfied all the assumption, adding the slope component does not show any improvement as the AIC value show some increment.

Therefore the best model to represent the monthly number of road accidents is local level with deterministic seasonal. This model allowed level component to vary overtime while the seasonal component is fixed. Figure 3 shows that the level of road accidents is increased from overtime with the true observations are above and below the estimated level. Due to the deterministic assumption of seasonal component, Figure 4 shows a seasonal

Table 3. Estimation result of univariate model of structural time series

		Univariate Structural Time Series Model				
		Local Level	Local Linear Trend	Local Level Seasonal	Local Level Fixed Seasonal	Local Linear Trend Seasonal
Variance Disturbance	Level	0.0002	7.2×10^{-5}	0.0002	0.0002	9.8×10^{-5}
	Slope		0.0000			0.0000
	Seasonal			4.8×10^{-8}	0.0000	4.6×10^{-8}
	Irregular	0.0018	0.0019	0.0009	0.0009	0.0010
Residual Diagnostic	LB (12)	0.0000*	0.0000*	0.1882	0.1849	0.3016
	GQ	0.0019*	0.8772	0.4270	0.4221	0.3817
	JB	0.9032	0.0015*	0.6626	0.6579	0.8305
AIC		-5.9585	-5.9268	-5.6890	-5.7017	-5.6452

*values of the diagnostic are the p-value of the test statistic which indicate significance at 10% significance level.

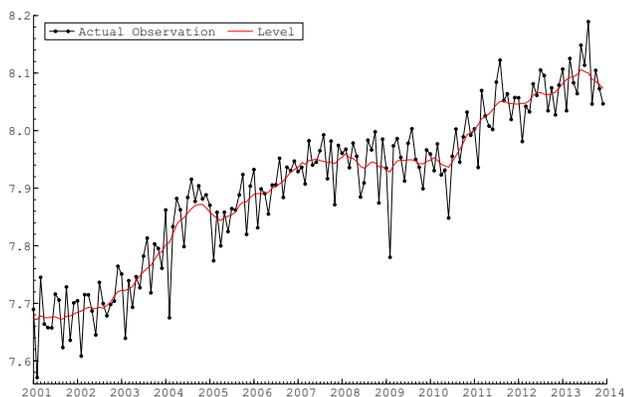


Figure 3. Estimated level component of structural time series model.

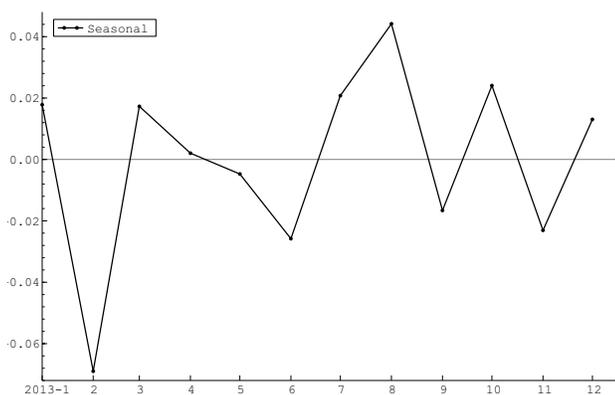


Figure 4. Estimated seasonal component of structural time series model.

component of the road accidents for year 2012. The plot show shows that the higher and lower road accidents occur in August and February respectively.

3.2.1 Adding Explanatory and Intervention Variable

Since the aim of this study is to investigate the most influential factors that contribute to road accident, this lead to consideration of adding the univariate model above with few explanatory and intervention variables. The best univariate model to represent the monthly number of road accidents is local level with the deterministic seasonal model. By adding explanatory and intervention variable the estimation can be simplified as in Table 4. The lag of dependent variable also added to make the model comparable with the time series regression model.

Table 4 shows the estimated structural time series model for the number of road accidents when accounting for the eight selected explanatory variable. For comparison purposes the estimated coefficients in the time series regression model are reproduced. Similar to time series regression model, the result from the structural time series model show negative effects of an increase in the API and during the school holiday and positive effect of an increase in the rainfall, number of rainy day, temperature and during the OPSKP operation on the number of road accidents. Unlike the estimated regression model, the estimated structural time series model shows the negative effect of an increase in oil price and the enforcement of rear seatbelt. The structural time series model provides better reflection on the effect of enforcement of rear seatbelt law that reduces the number of road accidents in Malaysia. These two variables are also found to be significant at the 10% level in time series regression.

Table 4. Estimation of structural time series with explanatory variable

	Time Series Regression		Structural Time Series	
	Coef	Sig.	Coef	Sig.
Constant	1.7732	0.0028*	9.2140	0.0000
MRAIN	0.0098	0.3289	0.0079	0.2503
DRAIN	0.0016	0.2901	0.0017	0.1125
TEMP	0.0121	0.1723	0.0191	0.0142*
API	-0.0001	0.7145	-0.0001	0.8550
OILP	0.0003	0.0006*	-0.0001	0.5975
BELT	0.0239	0.0340*	-0.0469	0.1629
OPSKP	0.0159	0.1142	0.0213	0.0043*
SCHOL	-0.0080	0.5174	-0.0114	0.1621
Seasonal_1	-0.0236	0.2129	-0.0023	0.8534
Seasonal_2	-0.1308	0.0000*	-0.0180	0.0148*
Seasonal_3	0.0172	0.4114	0.0065	0.1375
Seasonal_4	-0.0727	0.0030*	0.0064	0.2504
Seasonal_5	-0.0540	0.0083*	0.0252	0.0000*
Seasonal_6	-0.0646	0.0011*	0.0035	0.2295
Seasonal_7	-0.0091	0.6815	-0.0012	0.7635
Seasonal_8	-0.0202	0.2818	0.0142	0.0006*
Seasonal_9	-0.1018	0.0000*	-0.0022	0.8598
Seasonal_10	-0.0243	0.2922	0.0255	0.0000*
Seasonal_11	-0.0879	0.0000*	-0.0036	0.0921
LRA_1	0.7143	0.0000*	-0.2159	0.0118*
R ²	0.9102		0.5136	
LB	36.306	0.0003*	10.16	0.4266
GQ	0.9288	0.4863	1.197	0.2745
JB	3.4942	0.1743	4.3514	0.1135

*Denotes significance at 10%

While in the structural time series model road accidents were significantly related to with the increasing temperature and during the OPSKP operation. The results of increasing the temperature may increase road accidents by 1% were agree with ³¹ that state, road accidents is more likely to occur during hot temperature. Driving performance during high temperature can possibly worse due psychological and physiology effects of ambient temperature. The increasing of road accidents during OPSKP is expected since the OPSKP is only implemented during the festival seasons, when the traffic volume become much higher as Malaysian take full advantage to go back

Table 5. Prediction Accuracy

	Time Series Regression	Structural Time series
RMSE	0.0467	0.0429
MAE	0.0356	0.0340
MAPE	0.0011	0.0005

to hometown and has contributed higher no of road accidents.

This result shows that the main factor of road accidents is related to the traffic volume and temperature. The traffic volume can be reduced by greater campaigning for car pooling and improving the public transport particularly on bus and train. Greater encouragement of car pooling not only reduces cost of fuel and air pollution, but it also reduced the stress of driving alone that may lead to road accidents. Improvement in public transport such as increasing the frequency and accessibility of local and school buses as well as inter-state train journey, would help to reduce the traffic volume and road congestion. The temperature effect, may not be control but the prevention measure during the adverse temperature should be introduce.

Although the value of R^2 from the structural time series model is smaller than the corresponding value from the time series regression model, the time series regression model show that it does not fulfill the independence assumption for the residual. The better use of structural time series model can also be seen from Figure 5 whereby the fitted values are closer to the observed value as compared to that in Figure 1 for the case in time series regression model. For instance it is also can be observed from Figure 6 of the residual plot of structural time series that

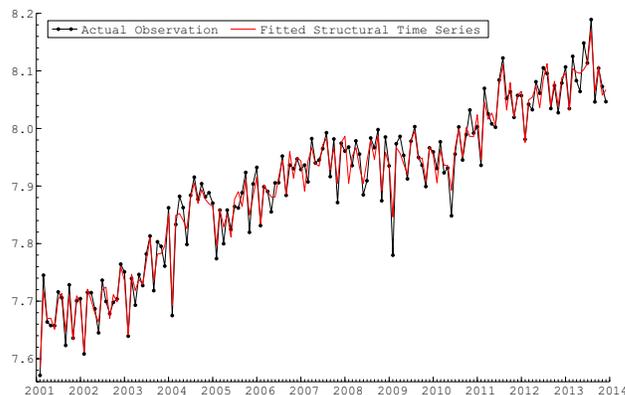


Figure 5. Estimated model of structural time series.

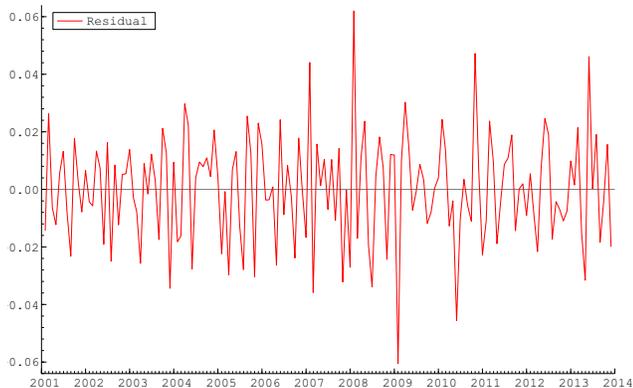


Figure 6. Estimated residual of structural time series model.

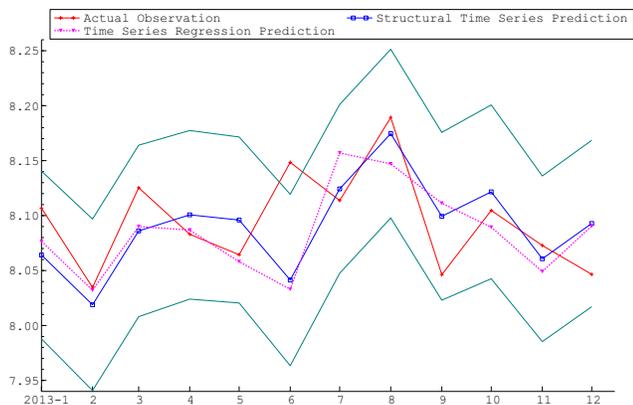


Figure 7. Comparison of model prediction.

are much closer to independent random values compared in Figure 2. It is proved from independent test of the diagnostic residual as in Table 4.

3.3 Prediction Accuracy

Recall that in this study two statistical model were used to model Penang road accidents. In that case to ensure the best fitted model between this two preferred model, loss measurement such as RMSE, MAE and MAPE were use. Twelve in sample data that is from January 2013 until December 2013 were used as a benchmark to test the prediction accuracy. The result were tabulated as in Table 5. The result indicate that, both model have no much difference but the structural time series model prediction is superior compared to time series regression. It it can be seen through the both loss function that show structural time series give less error than time series regression. The illustration of this prediction were illustrated in Figure 7. This figure clearly show that the

estimated of structural time series were closed to the actual value.

4. Conclusion

This study compares two time series methodology which is the time series regression and structural time series in modeling the number of road accidents in Penang. In addition, this study aims to investigate the factors that affect the number of road accidents. The study found that the two models are different in terms of the significant factor affecting the number of road accidents. The structural time series model is found to not only better fit the number of road accidents but give better prediction with all the residual assumption were satisfied.

Not surprisingly the structural time series is a better model since it allows the level of the number of road accidents to vary overtime. The structural time series model shows the temperature, and operation of Ops Sikap are significant factors influencing the number of road accidents in Penang. The enforcement of rear seat belt law also helps in reducing the number of road accidents.

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. Besides, the model may not accurately represent the number of road accidents as the data only representing reported cases of road accidents while there might be more unreported cases of accidents. Further investigation may include other relevant variable such as the traffic volume and economic factor.

5. Acknowledgement

The authors acknowledge the financial support from the School of Mathematical Sciences, Universiti Sains Malaysia in funding this study and indebted to the staff of the Unit Traffic of Royal Malaysia Police, Bukit Aman, Department of Environment and Meteorology Department for providing the data for this study.

6. References

1. Manan AMM, Varhelyi A. Motorcycle fatalities in Malaysia. *IATSS Res. International Association of Traffic and Safety Sciences*; 2012 Jul; 36(1):30–9.
2. Sarani R, Allyana SSM, Jamilah M, Wong S. Predicting Malaysian Road Fatalities for Year 2020 Predicting

- Malaysian Road Fatalities for Year 2020. Malaysian Institute of Road Safety Research. Kuala Lumpur; 2012.
3. Smeed R. Variations in the pattern of accident rates in different countries and their causes. *Traffic Eng Control*. 1968; 10(7):364–71.
 4. Andreassen D. Linking deaths with vehicles and population. *Traffic Eng Control*. 1985; (11):547–9.
 5. Valli PP. Road Accident Models For Large Metropolitan Cities of India. *IATSS Res*. 2005; 29(1):57–65.
 6. Nasaruddin N, Wah YB, Mara U, Voon WS. Fatality Prediction Model for Motorcycle Accidents in Malaysia. 1999; (Revision).
 7. Zlatoper TJ. Regression Analysis of Time Series Data on Motor Vehicle Death in the United States. *Journal Transp Econ Policy*. 1984; 18(3):263–73.
 8. Desai MM, Patel A. Road Accidents Study Based On Regression Model: A Case Study of Ahmedabad City. National Conference on recent trends in Engineering and technology. 2011; 1–8.
 9. Ali GA, Bakheit CS. Comparative Analysis and Prediction of Traffic Accidents in Sudan Using Artificial Neural Network and Statistical Method. Proceedings of the 30th Southern African Transport Conference (SATC 2011). 2011. p. 202–14.
 10. Umar RR, Murray GM, Brian LH. Modelling of Conspicuity- Related Motorcycle Accidents in Seremban and Shah Alam, Malaysia. *Accid Anal Prev*. 1996; 28(3):325–32.
 11. Sarani R, Hashim HH, Yaacob WWF, Mohamed N, Radin Sohadi RU. The Effect of Rear Seatbelt Advocacy and Law Enforcement in Reducing Injuries among Passenger Vehicle. *Int J Public Heal Res*. 2013; 3(1):267–75.
 12. Greibe P. Accident prediction models for urban roads. *Accid Anal Prev*. 2003; 35(May 2001):273–85.
 13. Harnen S, Wong SV, Hashim WIW. Motorcycle accident prediction model for junctions on urban roads in Malaysia. *Adv Transp Stud an Int J*. 2006; 8:31–40.
 14. Abusini S. The effect of road characteristics on motorcycle accident in Batu east Java Indonesia. 2013; 241:241–6. Available from: <http://link.aip.org/link/APCPCS/v1557/i1/p241/s1&Agg=doi>.
 15. Manan AMM, Jonsson T, Várhelyi A. Development of a safety performance function for motorcycle accident fatalities on Malaysian primary roads. *Saf Sci*. Elsevier Ltd; 2013 Dec; 60:13–20. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0925753513001343>.
 16. Scott PP. Modelling Time Series of British Road Accidents Data. *Accid Anal Prev*. 1986; 18(2):109–17.
 17. Law TH, Umar RSR, Zulkaurnain S, Kulanthayan S. Impact of the effect of economic crisis and the targeted motorcycle safety programme on motorcycle-related accidents, injuries and fatalities in Malaysia. *Int J Inj Contr Saf Promot*. 2005 Mar; 12(1):9–21. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15814371>
 18. Yaacob WWF, Wan Husin WZ, Abd Aziz N, Nordin NI. An Intervention Model of Road Accidents: The Case of OPS Sikap. *J Appl Sci*. 2011; 11(7):1105–12.
 19. Noland RB. Traffic fatalities and injuries: the effect of changes in infrastructure and other trends. *Accid Anal Prev*. 2003 Jul; 35(4):599–611. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12729823>
 20. Brijs T, Karlis D, Wets G. Studying the effect of weather conditions on daily crash counts using a discrete time-series model. *Accid Anal Prev*. 2008 May; 40(3):1180–90. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18460387>
 21. Yaacob WWF, Lazim MA, Wah YB. Evaluating Spatial and Temporal Effects of Accidents Likelihood using Random Effects Panel Count Model. International Conference on Science and Social Research (CSSR 2010). 2010. p. 960–4.
 22. Wan Yaacob WF, Lazim MA, Bee Wah Y. Modeling Road Accidents using Fixed Effects Model: Conditional versus Unconditional Model. Proceeding of the World Congress on Engineering 2008. 2012.
 23. Harvey A, Durbin J. The Effects of Seat Belt Legislation on British Road Casualties : A Case Study in Structural Time Series Modelling. *J R Stat Soc*. 1986; 149(3):187–227.
 24. Lazim MA. Introductory Business Forecasting a Practical Approach. Shah Alam: Pusat Penerbitan Universiti(UPENA) UiTM; 2005.
 25. Montgomery DC, Peck EA, Vining GG. Introduction to Linear Regression Analysis. 3rd ed. Canada: John Wiley & Sons; 2001.
 26. Durbin J, Koopman SJS. Time Series Analysis by State Space Method. 2nd ed. London: Oxford University Press; 2012.
 27. Royal Malaysian Police. Laporan Perangkaan Kemalangan Jalan Raya , Malaysia. Bukit Aman; 2011.
 28. Key K, Simmonds I. Road accidents and rainfall in a large Australian city. *Accid Anal Prev*. 2006; 38(Jan 2005):445–54.
 29. Yaacob WWF, Lazim MA, Bee Wah Y. Applying Fixed Effect Panel Count Model to Examine Road Accident Occurrence. *J Appl Sci*. 2011; 11(7):1185–91.
 30. Majed MJ. Apabila musim cuti sekolah, kemalangan jalan raya pun meningkat. *Utusan Malaysia*. 2012.
 31. Jawi ZM, Sarani R, Voon WS, Farhan A, Sadullah M. Weather as a Road Safety Hazard in Malaysia - An Overview. 2009.