MODELING MALAYSIAN ROAD ACCIDENTS: THE STRUCTURAL TIME SERIES APPROACH

NOOR WAHIDA BINTI MD JUNUS

UNIVERSITI SAINS MALAYSIA

2018

MODELING MALAYSIAN ROAD ACCIDENTS: THE STRUCTURAL TIME SERIES APPROACH

by

NOOR WAHIDA BINTI MD JUNUS

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

January 2018

ACKNOWLEDGEMENT

First and foremost praise to Allah the Almighty who give knowledge, strength and determination to finally finish my thesis even though the journey was so hard.

This success cannot be achieved without the guidance and assistance from others. Therefore I would like to express my sincere gratitude to my supervisor Assoc. Prof. Dr. Mohd Tahir Ismail for the continuous support of my Ph.D study and related research, for his patience, motivation, and immense knowledge. Besides, I would like to thank my co-supervisor Dr. Zainudin Arsad for his insightful comments and encouragement, but also for the hard question which incented me to widen my research from various perspectives.

A very special gratitude goes out to Ministry of Higher Education as well as Sultan Idris Education University for helping and providing the funding of my study. This special gratitude also goes to Institute of Postgraduate Study and School of Mathematical Sciences that supported several conference fees along my study period.

Finally I am grateful to my family members especially my mother who have provided me through moral and emotional support in my life. Last but by no means least, to all my friends especially everyone in School of Mathematical Sciences postgraduate lab, it was great sharing laboratory with all of you during last four years.

Thanks for all your encouragement.

TABLE OF CONTENTS

ACKNOWLEDGEMENT		ii
TABLE OF CONTENTS		iii
LIST OF TABLES		viii
LIST OF FIGURES		xi
LIST OF SYMBOLS AND ABBREVIATIONS		xiii
ABSTRAK		xxi
ABSTRACT		xxiii
CHAPTER 1- INTRODUCTION		
1.1	Background of the Study	1
1.2	Motivation and Problem Statements	4
1.3	Objective	5
1.4	Contribution of the Study	7
1.5	Scope of data	8
	1.5.1 Road Accidents	9
	1.5.2 Climate Related Variables	10
	1.5.3 Economic Related Variables	13
	1.5.4 Seasonal Related Variables	14
	1.5.5 Road Safety Related Variables	17
1.6	Limitation of the Study	18
1.7	Summary and Thesis Organization	19
CHAPTER 2 - LITERATURE REVIEW		
2.1	Structural Time Series	21
2.2	Advantages of Structural Time Series	23

2.3 Application of the Structural Time Series Model 24
2.3.1 Economics 24
2.3.2 Meteorology, Ecology and Agriculture 26
2.3.3 Other Disciplines 27
2.4 Common Techniques of Road Safety Modeling 28
2.5 Structural Time Series in Road Safety 43
2.6 Comparison of Road Accidents Models between This Study with Previous Studies 47
2.7 Summary 49
CHAPTER 3 - METHODOLOGY
3.1 Properties of Data 50
3.1.1 Descriptive Statistics 50
3.1.2 Time Series Plot 52
3.1.3 Correlation analysis 52
3.1.4 Unit Root Test 53
3.2 Regression Analysis 55
3.2.1 Time Series Regression 55
3.2.2 Parameter Estimation and Hypothesis Testing 56
3.2.3 Diagnostics Checking 59
3.3 Box and Jenkins Analysis 62
3.3.1 Box and Jenkins ARIMA Model 63
3.3.2 Box and Jenkins Model Identification 66
3.3.3 Box and Jenkins Model Estimation and Validation 70
3.4 Structural Time Series 72
3.4.1 Trend Model 73

3.4.2 Seasonal Model 74

3.4.3 Incorporating Explanatory and Intervention Variable 76

3.4.4 State Space Form 77

3.4.5 Kalman Filter Estimation 79

3.5 Evaluation of Structural Time Series Model 82

3.5.1 Model Diagnostic 82

3.5.2 Goodness-of-fit of Structural Time Series 84

3.6 Application on Road Accidents 85

3.7 Summary 87

CHAPTER 4 - PRELIMINARY STUDY

4.1 Descriptive Statistics 89

4.1.1 Road Accidents Series 89

4.1.2 Climate Related Variable 90

4.1.3 Economic Related Variables 92

4.2 Time Series Plot 93

4.2.1 Road Accident Series 93

4.2.2 Climate Related Variables 97

4.2.3 Economic Related Variables 108

4.3 Correlation Analysis 110

4.3.1 Correlation Analysis for Regions 111

4.3.2 Correlation Analysis for Individual State 112

4.3.3 Unit Root Test 113

4.4 Time Series Regression 114

4.4.1 Time Series Regression with Seasonal Dummies 115

4.4.2 Incorporating Explanatory Variables 119

4.5 Box and Jenkins SARIMA 130
4.5.1 Estimating SARIMA Model for Regional Road Accidents 131
4.5.2 Estimating SARIMA Model for Individual States Road Accidents 132
4.6 Summary 134

CHAPTER 5 - MODELING UNIVARIATE ROAD ACCIDENTS MODEL

5.1 Model Estimation 135
5.1.1 Estimating Road Accident Model for Northern Region 136
5.1.2 Estimating Road Accident Model for Other Regions 137
5.2 Understanding Estimated Regional Road Accidents Model. 141
5.2.1 Trend Pattern of Regional Road Accidents 141
5.2.2 Seasonal Pattern for Regional Road Accidents 145
5.3 Estimating Road Accidents Model for Individual State Level 148
5.3.1 Northern State Road Accidents Pattern 150
5.3.2 Road Accidents Pattern for Other States 155
5.4 Special Features of Seasonal Road Accidents Pattern 166
5.5 Prediction and Forecasting Performance of the Structural Time Series 167
5.6 Summary 174

CHAPTER 6 - INCORPORATING EXPLANATORY AND INTERVENTION VARIABLE OF ROAD ACCIDENTS MODEL

6.1 Estimating and Understanding Regional Road Accidents Model 176
6.1.1 Error Estimate 178
6.1.2 Estimation of Trend and Seasonal Component 181
6.1.3 Estimation of Explanatory Variables 187
6.1.4 Observing Outliers and Structural Breaks 188

6.2 Estimating and Understanding Individual State Road Accidents Model 191

6.2.1 Error Estimate 193

6.2.1 States Level Road Accidents Pattern with Explanatory Variables 197

6.2.2 Estimation of Explanatory Variables 210

6.2.3 Observing Possible Outliers and Structural Breaks 213

6.3 Prediction Performance of STS with Explanatory Variables 214

6.4 Summary 222

CHAPTER 7 - CONCLUSION AND RECOMMENDATION

7.1 Concluding Remarks 224

7.2 Implications of the Study 227

7.3 Suggestion for Future Research 229

REFERENCES 230

APPENDICES

LIST OF PUBLICATIONS

LIST OF TABLES

		Page
Table 1.1:	List of Variables and unit of measurements	9
Table 1.2:	Aggregated regions and their corresponding states	10
Table 1.3:	Location of stations that record climate related variables	12
Table 1.4:	Computation of Aggregation Samples	12
Table 1.5:	An example of BLKG coding	16
Table 1.6:	An example of SAFE coding	18
Table 2.1:	Recent studies on application of structural time series	27
Table 2.2:	Recent methods/ models used in road safety study	38
Table 2.3:	Recent studies on the application of structural time series on road safety	45
Table 2.4:	Comparisons between model used in this thesis with previous models	49
Table 3.1:	Value range of coefficient of correlation	53
Table 3.2:	Durbin-Watson test decision	60
Table 3.3:	Model identification	69
Table 3.4:	Structural time series specification model (trend+seasonal)	75
Table 4.1:	Descriptive statistics of road accidents series	90
Table 4.2:	Descriptive statistics of climate related variables	91
Table 4.3:	Descriptive statistics of the monthly economic related variables	93
Table 4.4:	Correlation coefficient between number of road accident for each region with selected dependent variables	111
Table 4.5:	Correlation coefficient between the number of road accidents and selected dependent variables for each state	113

Table 4.6:	Estimated regional road accidents model	116
Table 4.7:	Road accidents model for individual states	117
Table 4.8:	Durbin-Watson test of autocorrelation	119
Table 4.9:	Variance of inflation factor for regions	120
Table 4.10:	Variance of inflation factor for individual states	120
Table 4.11:	List of possible outlier observations	121
Table 4.12:	Estimated regional road accidents model by incorporating explanatory variables	123
Table 4.13:	Estimated individual states' road accidents model by incorporating explanatory variables	127
Table 4.14:	Estimated regional road accidents model based on Box and Jenkins SARIMA models	131
Table 4.15:	Estimated road accidents model for individual states based on Box and Jenkins SARIMA models	132
Table 5.1:	Estimated results and performance criteria of STS model for northern region road accidents	137
Table 5.2:	Estimated results and performance criteria of STS model for central region road accidents	138
Table 5.3:	Estimated results and performance criteria of STS model for east coast region road accidents	139
Table 5.4:	Estimated results and performance criteria of STS model for southern region road accidents	140
Table 5.5:	Estimated results and performance criteria of STS model for Borneo region road accidents	140
Table 5.6:	Final estimation results according to regions	142
Table 5.7:	Best road accident model specification for each individual state	149
Table 5.8:	Final estimation result of northern state road accident model	151
Table 5.9:	Final estimation result of central and southern state road accident model	155

Table 5.10:	Final estimation results of east coast and Borneo states road accident models	161
Table 5.11:	Special feature of seasonality of road accidents pattern	166
Table 5.12:	Error values for prediction road accidents models	173
Table 5.13:	Error values for forecasting road accidents models	174
Table 6.1:	Estimation of regional road accidents after adding explanatory variables	178
Table 6.2:	Final road accidents model estimates with explanatory variables	182
Table 6.3:	Estimation of state level road accidents with explanatory variables	192
Table 6.4:	Final estimate of road accidents model with the explanatory variables	195
Table 6.5:	Error values for prediction road accidents models with explanatory variables	221

LIST OF FIGURES

		Page
Figure 1.1:	An illustration of process determining g_1 and g_2	16
Figure 3.1:	Box and Jenkins methodology	67
Figure 3.2:	Box and Jenkins model identification process	69
Figure 3.3:	Step by step procedure of designing of road accidents modeling	86
Figure 4.1:	Monthly time series plot of road accidents for all regions.	94
Figure 4.2:	Monthly time series plot of road accidents for individual states	95
Figure 4.3:	Monthly time series plot of amount of rainfall for all regions	98
Figure 4.4:	Monthly time series plot of amount of rainfall for individual states	100
Figure 4.5:	Monthly time series plot of number of rainy days for all regions	101
Figure 4.6:	Monthly time series plot of number of rainy days for individual states	102
Figure 4.7:	Monthly time series plot of maximum temperature for all regions	103
Figure 4.8:	Monthly time series plot of maximum temperature for individual states	105
Figure 4.9:	Regional time series plot for monthly maximum API	106
Figure 4.10:	Time series plot of monthly maximum API for individual states	107
Figure 4.11:	Monthly time series plot for economic effect	109
Figure 5.1:	Seasonal components of northern region road accidents	137
Figure 5.2:	Seasonal components of southern region road accidents	139

Figure 5.3:	Trend components according to regions	143
Figure 5.4:	Seasonal components according to regions	146
Figure 5.5:	Trend components of road accidents for northern states	151
Figure 5.6:	Seasonal component for northern states	154
Figure 5.7:	Trend components of road accidents in Negeri Sembilan, Melaka and Johor	156
Figure 5.8:	Seasonal component for southern states	158
Figure 5.9:	Trend components of road accidents in for central states	159
Figure 5.10:	Seasonal components of road accidents for central states	160
Figure 5.11:	Trend components of road accidents for east coast states	162
Figure 5.12:	Seasonal components of road accidents in east coast states	163
Figure 5.13:	Trend components of road accidents for Borneo states	164
Figure 5.14:	Seasonal components of road accidents for Borneo states	165
Figure 5.15:	Real and estimated states road accidents produced by TSR, SARIMA and STS model	168
Figure 5.16:	Real and estimated states road accidents produced by TSR, SARIMA and STS model	169
Figure 6.1:	Auxiliary residual of regional road accidents model	180
Figure 6.2:	Trend components with explanatory and intervention variable according to regions	183
Figure 6.3:	Seasonal components according to regions	185
Figure 6.4:	Trend pattern of state level road accidents	198
Figure 6.5:	Seasonal pattern road accidents model for individual states	204
Figure 6.6:	Real and estimated regional road accidents produced by TSR and STS models	215
Figure 6.7:	Real and estimated states road accidents values produced by TSR and STS models	217

LIST OF SYMBOLS AND ABBREVIATIONS

Y_t	Dependent / response variable
μ_t	Trend component
γ_t	Seasonal component
ε_t	Irregular component/observation error / disturbance
β	Regression coefficient
η_t	Level error/disturbance
ς_t	Slope error/disturbance
ν_t	Slope component
ω	Seasonal error/ disturbance
t	Time of t
α	state component
c	constant
$\bar{Y}$	Mean of Y observation
θ	Moving average parameter
ϕ	Autoregressive parameter
d	Order of differencing
r	Correlation coefficient

R^2	Coefficient of determination
$\mathbf{Z}_t, \mathbf{T}_t, \mathbf{R}_t, \mathbf{H}_t, \mathbf{Q}_t$	System matrices
$\mathbf{F}_t$	Variance of 1-step ahead prediction error
$\mathbf{v}_t$	1-step ahead prediction error
I	Dummy / intervention variable
W	Non seasonal different function
Z	Seasonal different function
s	Seasonal periodic
σ^2	Variance
g	Number of hoilday
m	Total number of holiday
P	Order of seasonal autoregressive
D	Order of seasonal differencing
Q	Order of seasonal moving average
n	Number of sample size
t_r	Test of significance correlation coefficient
t_β	Test of significance regression coefficient
X	Explanatory variable

F_0	F-statistics
d_L	Lower limit of Durbin-Watson statistic
d_U	Upper limit of Durbin-Watson statistics
b_i	Standardize residuals
k	Number of lag
κ	Number of estimated parameter
$E(\ )$	Expected value
$Var(\ )$	variance
$\mathrm{cov}(\ )$	covariance
Δ	Differencing process
W_t	Non-seasonal differencing function
Z_t	Seasonal differencing function
I_t	Intervention variable
λ	Intervention coefficient
ρ_k	Autocorrelation function
τ	State component error matrix
$\mathbf{v}_t$	Prediction error matrix
F_t	Variance of prediction error matrix

e	exponent
$\simeq$	Approximate
H_t	Variance of measurement error matrix
Q_t	Variance of state component error matrix
AADK	National Anti-Drugs Agency
ACF	Autocorrelation function
ADF	Augmented Dickey Fuller
AIC	Akaike information criterion
ANN	Artificial neural network
API	Air pollution index
AR	Autoregressive
ARIMA	Autoregressive integrated moving average
ARMA	Autoregressive moving average
ASEAN	Association of Southeast Asian Nations
BIC	Bayesian information criterion
BLKG	Balik Kampung
BSM	Basic structural model
CNY	Chinese new year
CO2	Carbon dioxide
CPI	Consumer price index for transportation

CUSUM	cumulative sum control chart
DL	Deterministic level
DLDS	Deterministic linear with deterministic seasonal
DLSS	Deterministic level with stochastic seasonal
DOS	Department of Statistics
DTDS	Deterministic trend with deterministic seasonal
DTSS	Deterministic trend with stochastic seasonal
dw	Durbin watson
EM	Expectation-maximization
FENB	Fixed effect negative binomial
FEP	Fixed effect Poisson
GDP	Gross domestic product
GLM	Generalized Linear Model
GQ	Goldfeld-Quandt test
I	integrated
INAR	Integer autoregressive
JB	Jarque Bera test
JPJ	Road Transport Department
KILL	Killed
KSI	killed and seriously injured
LB	Ljung-Box test

LDDS	Local level drift with deterministic seasonal
LDSS	Local level drift with stochastic seasonal
LL	Local level
LLDS	Local level deterministic seasonal
LLSS	Local level stochastic seasonal
LRT	Latent risk time series
LTDS	Linear trend deterministic seasonal
LTSS	Linear trend with stochastics seasonal
MA	Moving average
MAAP	Microcomputer Accident Analysis Package
MAPE	Mean absolute percentage error
Max	Maximum
Min	Minimum
MSE	Mean square error
MSP	Motorcycle Safety Programme
NA	Not applicable
NB	Negative binomial
NO2	Nitrogen dioxide
O3	Ozone
OECD	Organisation for Economic Co-operation and Development
OILP	Crude oil price

OLS	Ordinary least square regression
p	Order of autoregressive
PACF	Partial autocorrelation function
PCR	Principal component regression
PCR	Principal component regression
PM10	Particulate matter less that 10 microns
q	Order of moving average
RAIND	Number of rainy day
RAINF	Monthly average of rainfall amount
RENB	random effect negative binomial
RMP	Royal Malaysia Police
RMSE	Root mean square error
SAFE	Road safety operation (OPS sikap/OPS selamat)
SAR	Seasonal autoregressive
SARIMA	Seasonal autoregrssive integrated moving average
SARMA	Seasonal autoregressive moving average
SD	Standard deviation
SMA	Seasonal moving average
SO2	Sulphur dioxide
SPAD	Land Public Transport Comission
STDS	Smooth trend with deterministic seasonal

STS	Structural time series
STSS	Smooth trend with stochastic seasonal
SUTSE	Seemingly Unrelated Time Series Equations
SWOV	Dutch Foundation of Road Safety Research
TEMP	temperature
TSR	Time series regression
UPM	Universiti Putra Malaysia
US	United States of America
USM	Universiti Sains Malaysia
VIF	Variance inflation factor
WHO	World Health Organization
WN	White noise

PERMODELAN KEMALANGAN JALAN RAYA DI MALAYSIA: PENDEKATAN SIRI MASA BERSTRUKTUR

ABSTRAK

Permodelan bilangan kemalangan jalan raya telah menjadi topik umum sejak kebelakangan ini. Beberapa kajian berkaitan telah dijalankan dengan tujuan untuk mendapatkan model terbaik yang dapat meramal kemalangan jalan raya dengan lebih tepat. Walau bagaimanapun corak atau pola arah aliran dan kebermusiman bagi kemalangan jalan raya jarang dititikberatkan. Dengan menganggarkan corak arah aliran dan kebermusiman, secara tidak langsung sistem peramalan menjadi lebih baik. Secara tradisinya, penganggaran corak arah aliran dan kebermusiman menggunakan kaedah penguraian. Namun kaedah ini menghasilkan peramalan yang kurang tepat dan tidak dapat menggambarkan keadaan sebenar. Oleh yang demikian pendekatan siri masa berstruktur (STS) dicadangkan untuk memodelkan corak arah aliran dan kebermusiman kemalangan jalan raya. Hal ini kerana pendekatan STS membolehkan interpretasi secara terus dan menawarkan komponen siri masa berubah-ubah mengikut masa. Dalam kajian ini, model kemalangan jalan raya dibangunkan dengan menggunakan pendekatan STS. Melalui kaedah ini, corak arah aliran dan kebermusiman kemalangan jalan raya dapat diperhatikan. Kajian ini dijalankan ke atas 5 rantau utama dan semua 14 buah negeri di Malaysia. Kajian ini juga menyiasat pengaruh terhadap kemalangan jalan raya dengan menggunakan pembolehubah penerang yang bersesuaian. Lapan pembolehubah penerang telah dipilih termasuk empat pembolehubah iklim, dua pembolehubah ekonomi, pembolehubah bermusim, dan pembolehubah berkaitan keselamatan jalan raya. Keberkesanan model untuk menjangkakan dan meramal kemalangan masa depan dibandingkan dengan model sedia ada seperti model siri masa regresi (TSR) dan model autoregresi bersepadu purata bergerak bermusim (SARIMA). Kajian mendapati, corak arah aliran dan kebermusiman kejadian kemalangan jalan raya berbeza mengikut lokasi. Bilangan kemalangan jalan raya dianggarkan meningkat pada musim perayaan terutamanya di negeri-negeri yang kurang membangun. Di samping itu ciri-ciri khas perilaku stokastik bagi kemalangan jalan raya dapat diperhatikan. Dalam tempoh kajian, corak kemalangan jalan raya berfluktuasi turun dan naik. Pada masa yang sama pengaruh terhadap kemalangan jalan raya juga berbeza mengikut lokasi. Dari segi prestasi peramalan, STS menunjukan peramalan yang boleh percaya berbanding dengan TSR dan SARIMA.

MODELING MALAYSIAN ROAD ACCIDENTS: THE STRUCTURAL TIME SERIES APPROACH

ABSTRACT

Modeling the number of road accidents occurrence is a quite common topic in recent years. A number of studies have been developed with the aim to find the best model that gives better prediction. However, statistical patterns such as trend and seasonality of road accidents is rarely observed. Estimating the pattern of trend and seasonal will indirectly provide a better impact on prediction system. Traditionally, estimation of trend and seasonal patterns are made based on decomposition method. Yet, this type of estimation shows intangible predictions as the estimation are based on deterministic form. Therefore, structural time series (STS) approach is proposed to model the trend and seasonal pattern of road accidents occurrence. The STS approach offered a direct interpretation and allowed the time series component including trend and seasonal to vary over time. In this thesis the road accidents model is developed using the STS approach with the aim to observe the pattern of trend and seasonality of road accidents occurrence. This thesis was done on all 5 main regions and 14 states in Malaysia. The study further enhance investigation on road accidents influences at different locations with appropriate explanatory variables. There are 8 explanatory variables considered in this study, which includes four climate variables, two economic variables, seasonal related variable and safety related variable. Effectiveness of the model is measured by comparing their prediction and forecasting performance with time series regression (TSR) and seasonal autoregressive integrated moving average (SARIMA) models. The study found that the trend and seasonal patterns of road accidents occurrence vary in different locations. The number of accidents was estimated to be higher

during festival seasons especially in non-developing states. Besides, the special features of the stochastic behavior of road accidents pattern is also observed. During the study period, the pattern of road accidents is fluctuate between increasing and decreasing. Similarly, the influence of road accidents in different locations also varies. In terms of the prediction and forecasting performance, STS gave more reliable prediction and forecasting compared to TSR and SARIMA models.

CHAPTER 1

INTRODUCTION

This chapter begins with the background of the study followed by the motivation of the thesis and proceeds with the objective, contribution of the study to the knowledge and society as well as the scope and limitation of the study. The summary which discusses the structure of the thesis will be presented at the end of this chapter.

1.1 Background of the Study

One of the aim of a developed country is to enhance the survival rate of its population by improving the community's healthcare and quality of life. In order to determine this, it is important to know the exact number and causes of mortality as components of the population's health status. Besides, the figures are also important for social economic planning and monitoring in which at the same time it can be used as a good evidence for policy making and implementation.

Across all countries, one of the leading causes of mortality is attributed to road accidents. Aderamo (2012a) revealed that road accidents in developing countries contributed 85 percent of world's mortality. Meanwhile, World Health Organization (WHO) in 2014 reported that the ninth leading cause of mortality with 1.3 million deaths is caused by road accidents, and in 2013, it is also the fourth leading cause of death in the United States. In Malaysia, for year 2013, Malaysian

Department of Statistics (DOS) reported that transport accidents have become the fifth causes of mortality among Malaysian populations and second cause of mortality among Malaysian male population.

Death from road crash or also known as road fatalities have a big impact to economic growth and at the same time affects the victims families emotionally. In 2004, WHO reported that in Bangladesh, over 70% of households state that their households income, food consumption and food production had decreased after a road death occured to one of their family members.

Therefore, a safe road traffic network system is very important to facilitate the movement of goods apart of improving the community health care by reducing the road death. The important key here is to reduce traffic accident that is main contributor to road fatalities. There are various factors which contribute to road accidents. It can be categorized into driver factor, vehicle factor and roadway factor (Bun, 2012).

Driver factor includes all factors related to the drivers and other road users. It includes the driver behavior, visual, clarity or clearness of hearing and reaction speed. The vehicle factor includes vehicle design, safety maintenance and safety feature that may reduce accidents occurrence. On the other hand, meteorological or climate condition such as temperature, precipitation, wind speed and fog are also important contributing factor to road accidents as they reduce visibility and cause the loss of vehicle control.

Various efforts have been done in order to reduce the number of road accidents. Specifically in Malaysia starting from early 1970, the first motorcycle lane was built along federal highway with the aim to reduce motorcycle accidents. Study by Radin Umar *et al.* (1996) found that this intervention has successfully reduced motorcycle accidents by 34%. In 1989, the Road Commissions Safety Cabinet was formed that is responsible to formulate 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, Malaysian government established a 5 years National Road Safety Target. The target is to reduce the number of accident death by 30% by year 2000. Various initiatives were carried out to achieve the target. In 1997, the road safety research centre which is under Universiti Putra Malaysia (UPM) was mandated to conduct research on motorcycle safety as one of its initiatives. In 2000 the reported accidents death was 6035, which is 5% lower than predicted death by Radin Umar, (1998) that is 6389.

In the Malaysian road safety plan (2006-2010) the government target to reduce 52.4% of road death by 2010. Among the initiatives to achieve the target was enforcement of Ops Sikap since 2001. This operation was conducted to ensure safety on all roads in Malaysia during festive seasons. It is followed by introducing rear seat belt legislation in 2009. However, in 2010, the index of road death stood at 3.4 per 10000 vehicles which are higher than expected that are 2.0 per 10000 vehicles (Sarani *et al.*, 2012). This is a relatively poor performance and it puts Malaysia as one of the developing countries that contributed the highest number of road fatalities per 100000 population among the ASEAN countries (Abdul Manan & Várhelyi, 2012).

1.2 Motivation and Problem Statements

As discussed before, Malaysia need a strong road safety analysis. Therefore, over the past few years, a number of studies on road safety have been developed. The aim of the studies is to investigate factors that contribute to road accidents as well as to identify the most accurate methods to predict road accidents. Numerical modeling is a common tool for estimating number of road accidents. The model can be either deterministic or probabilistic (stochastic). However, some of the study gives a poor prediction results especially in term of error structure. Sometimes, the studies produced models which either gave accurate prediction without explaining the phenomenon or could describe the phenomenon without being able to explain or predict it (Hakim, 1991).

The models which describe the main features of the series may give a better prediction model. These features can be examined from the pattern of the trend and the seasonal behaviour of the series. Trend and seasonal analysis are best carried out by means of unobserved components or structural time series (Harvey, 2006b) Unfortunately, road safety study which is the focus of this feature is very rare and limited especially in Malaysia. The studies usually focus on cross sectional studies and effectiveness of the intervention procedure. Therefore, the better model which can describe these valuable features and at the same time investigate the effectiveness of the intervention procedure may give a great impact in improving the road safety.

On the other hand, the scope of the variables used in the road safety study may not suitable especially the dummy variable which involved time series analysis. For example, the study by Radin Umar *et al.*, (1996) that incorporated the moving holiday effect describing festival holiday. They applied dummy variable to represent this event and name the variable as *Balik Kampung* (BLKG). It is coded as "0" to represent not BLKG season and "1" to represent BLKG season. In this case, this variable is quite relevant since the study use weekly data. However if the study involves a monthly, quarterly or annual series the dummy variable "0" and "1" is not suitable as the event only occurred partially during the unit data.

Recently, studies done on road safety either focus on regional of population specific aspects. It was found that road safety behaviour in larger population is more risky than smaller populations (Houston, 2007). Yet, these kinds of studies that compared between states or regions are very limited. Up to our knowledge, in Malaysia, only Wan Yaacob *et al.* (2012) made the comparison on the number of road accidents between each state. However their study was based on the panel data analysis. This method somehow resricted on the limited number of observation.

1.3 Objective

The main objective of this thesis is to model the number of road accidents occurrence in Malaysia using the structural time series approach. Indirectly, the model developments of this model allows to observance of stochastic behavior or pattern of road accidents. This study will observe and compare the variation of trends and patterns of road accidents during the study period that is between January 2001 to December 2013.

To obtain a better understanding of the trends and seasonal patterns the model is applied to aggregate datasets that includes five main regions and 14 states of Malaysia. The five main regions consist of the northern, southern, central, east coast and Borneo regions. The aim is to allow the investigation of pattern changes at different locations of regions and states.

After the trends and seasonal patterns have been observed, it is important to investigate the main contributors to these changes. In order to do that the explanatory variables which may explain the changes are incorporated in the model. The variables include climate related variables, economic related variables, rules and regulations enforced during the study period as well as seasonality related variables. Scott (1986) found that, besides the controllable explanatory variables can identified, incorporating the explanatory variables indirectly creates greater understanding of what “drive” the series, produce fluctuation and provides a basis against which to evaluate further impose changes on safety enforcement implementation.

Modeling and predicting road accidents occurrence has been commonly practiced by many researchers in the recent years. Many models have been introduced to predict road accidents occurrence. One of the most famous approaches from seventies is Box and Jenkins SARIMA model. Thus, the study will compare the forecasting performance of the univariate structural time series with Box and Jenkins SARIMA model. At the same time, as the starting point of structural time series is a regression model in which the explanatory variables are function of times (Harvey, 1989), the predicting and forecasting performance between two methodology are also compared for both models with and without the explanatory variables. After all the objective of this study can be summarized as follows:

i. To propose alternative road accidents model for each state in Malaysia by using the structural time series approach.

ii. To observe the deterministic and stochastic behaviour or pattern of road accidents for different regions and states.
iii. To investigate and to understand the influence of road accidents for different regions and states using the right explanatory variables.
iv. To compare the performance of the structural time series with time series regression and seasonal autoregressive integrated moving average model.

1.4 Contribution of the Study

Road safety study is not a new area of interest. This field has been studied by different researchers since a long time ago. The most common approach used is cross sectional model. However, the cross sectional data and their appropriate analysis provide a frozen snapshot on the road safety situation at a fixed point in time (Stipdonk, 2008). The changes and risk exposure over time cannot be observed. Therefore the most suitable approach is by considering time series data and their appropriate analysis. Time series method allows the investigation of changes in exposure, risk, of road safety overtime. In other words, it may provide the estimate of road safety which can help policy makers in developing realistic quantitative safety target.

There are various time series techniques that can be used to model road accidents occurrence. The Box and Jenkins model is among the common models preferred by researchers. However, in this study, the structural time series model is introduced in developing a road accidents model for Malaysia as it is offered a lot advantages. This is the first study that applied this approach for the Malaysian case. Kalman filter estimation technique is used in estimating the model parameters.

Through this model, time series components such as trends and seasonal components are extracted and modeled. Thus, the stochastic and deterministic behaviour of trends and seasonal patterns are observed and interpreted. On the other hand, the estimated unobserved component found in the model is important in giving a clear indication of the future long term movement of the series. Indirectly, the model may strengthen the system of road safety modeling in the future.

The best model with relevant explanatory variables may give a better understanding of the road accidents occurrence. In this study, the appropriate way of incorporating the festive seasons and safety operation enforcements are introduced into the model. This approach replaces the common procedure of incorporating those variables that are based on dummy variables of “0” and “1”. This approach is more sensible to the situation and expected to improve the time series of road safety modeling.

First time applied to model Malaysian road accidents, this study is expected to be beneficial to the society as well as the relevant parties. The road accidents model is developed according to regions and individual states instead of only small relative number of countries is covered as in existing study. Therefore, the proposed model may help the society and responsible parties in monitoring the road on a smaller scale, that focused on regions and individual states.

1.5 Scope of data

The main restriction in developing road accidents model is the suitability and availability of data. Some of the data may not be available during the study period and some of them may include missing values. The data are handled with extra care and the handling procedure is explained in details in the appropriate subsections. The variables considered in this thesis include the number of road accidents as the dependent or output variable, and the independent variables consist of climate related variable, economic related variables, seasonal related variables, and rules and regulation that have been enforced during the period of the study. As a summary the list of variables used in this study are tabulated in Table 1.1.

Table 1.1: List of variables and unit of measurement.

Variables	Description	Unit of Measurements
RA	Monthly number of road accidents	Log of RA
RAINF	Monthly Amount of rainfall	Milimeter (mm)
RAIND	Monthly number of rainy day	Day
TEMP	Monthly average of maximum temperature	Degree celcius (°C)
API	Monthly average of maximum air pollution index	Index
CPI	Consumer price index for transportation	Index
OILP	Crude oil price	Ringgit Malaysia (RM)
BLKG	Balik Kampung culture	Weight variable
SAFE	Operation of Ops Sikap dan Ops Selamat	Weight variable

1.5.1 Road Accidents

Majority studies made on road safety research employed number of injuries, number of casualties and frequency of road accidents as their variables of interests. In this study, monthly frequency or monthly number of road accident occurrences in all states is considered as the dependent variable. The number of road accidents was obtained from Royal Malaysia Police (RMP). RMP has defined road accidents as follows:

“The occurrence of accidents on public or private roads due to negligence or omission by any party concerned (on the aspects of road users conduct, maintenance of vehicle and road condition) or due to environmental factors (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 one moving vehicle, structure or animal and is recorded by the police”

The number of road accidents recorded include all 14 states in Malaysia. In this study, the number of road accidents is further aggregated into five main regions. The aggregated regions and corresponding states are defined as in Table 1.2. Throughout the study, each variable included are also aggregated into region and analysis were performed based on respective regions and states.

Table 1.2: Aggregated regions and their corresponding states

Region	States
Northern	Penang, Perlis, Kedah, and Perak
Southern	Negeri Sembilan, Melaka and Johor
Central	Kuala Lumpur and Selangor
East Coast	Kelantan, Terengganu, Pahang
Borneo	Sabah and Sarawak

1.5.2 Climate Related Variables

Weather variations have some influence on road conditions and road users. Hot day with high temperature may affect the mood of drivers. Heavy rain and hazy day might influence the vision of drivers. Heavy rain also made the road wet and slippery. These conditions, may contribute to road safety. In this case, climate variables would be the best factor to consider as one of the factors that caused road accidents.

Climate factors that are considered in this study include monthly average of rainfall amount (in millilitre) (RAINF), number of rainy days (RAIND), monthly maximum temperature (in degrees Celsius) (TEMP), and air pollution index (API).

Majority of the data were based on the Monthly Statistical Bulletin and Compendium of Environmental Statistics, which are published by the DOS, while other data were obtained from Department of Meteorology, the main body that is responsible for compiling the environmental data in Malaysia.

Daily rainfall was considered if the amount of rainfall recorded is equal or exceeds 0.1mm. API was calculated based on the average concentration of each air pollutant, namely SO2, NO2, CO2, O3, and PM10 and air pollutant with the highest concentration will determine the API. Typically, concentration of a fine particulate matter (PM10) is the highest compared to other pollutants, and this determines the API. The API can be categorized 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. However, API data are quite limited for the states of Selangor and Perlis. The data only covers the period of January 2004 to December 2013 for both states. The details of climate related variables incorporated in this study are tabulated in Table 1.3 together with the stations that collected the data. Besides, as in this study the series are aggregated into a regions, the climate related variable for regions are computed as fin Table 1.4

The similar variable such as amount of rainfall, number of rainy day and temperature were used in road safety modeling literature such as Scott (1986), Keay and Simmonds (2006), Wan Yaacob *et al.* (2011a, 2012) and Brijs et al. (2008). It was found that these factors have some influence on road accident occurrence. In 2012 Dutch Foundation of Road Safety Research (SWOV), stated that visibility can be reduced to 50 meters during heavy rain as well as during snow and thick fog. On the other hand, extreme temperature tends to cause harmful effects on driver's performance, road infrastructure, and vehicle components.

Table 1.3: Location of stations that record climate related variables

State	Station Location		
	RAINF & RAIND	**TEMP**	**API**
Penang	Bayan Lepas/ Butterworth	Bayan Lepas	Prai, USM
Perlis	Chuping	Chuping/ Kangar	Kangar
Kedah	Alor Setar, Langkawi	Alor Setar	Alor Star
Perak	Ipoh, K. Kangsar, Sitiawan	Ipoh/ Sitiawan	Tanjong Malim, Ipoh
Negeri Sembilan	Seremban	Seremban	Seremban
Melaka	Bandaraya Melaka	Bandaraya Melaka	Bandaraya Melaka
Johor	Batu Pahat, Senai, Kluang, Mersing	Mersing	Johor Bahru
Kuala Lumpur	Parlimen	Kuala Lumpur	Batu Muda
Selangor	Sepang, Petaling Jaya, Subang	Sepang, Petaling Jaya, Subang	Shah Alam
Kelantan	K. Bharu, K. Krai	Kota Bharu	Kota Bharu
Terengganu	K. Terengganu	Kuala Terengganu	Kuala Terengganu
Pahang	Jerantut, Cameron Highland, Muadzam Shah, Temerloh	Kuantan	Kuantan
Sabah	Kota Kinabalu	Kota Kinabalu	Kota Kinabalu
Sarawak	Kuching	Kuching	Kuching

Table 1.4: Computation of Aggregation Samples

Climate Related Variables	**Computation**
RAINF	The total amount of rainfall for each states under the regions
RAIND	The average number of rainyday for each states under the region
TEMP	The average of maximum temperature for each states under the region
API	The average of maximum air pollution index for each states under the region

Unfortunately, some of the climate related variables may involve missing values problem due to technical error. The missing values are observed in amount of rainfall, temperature and air pollution index for selected states. In order to handle these missing values, this study used linear interpolation method as suggested by Law *et al.* (2008). Interpolations were only done for short period of time by averaging the observations over preceding and posterior periods. However, because the missing values in this study involve a long period time, it is handle by interchanging the dataset into annual data. The preceding and posterior values are based on annual values. For example, if the missing value is for January 2005, the preceding value will be January 2004 and the posterior value will be January 2006.

1.5.3 Economic Related Variables

Numerous economic related variables could be incorporated in the study, however, their influence on accidents data may be indirect in changing the characteristics of traffic and road environments (Scott, 1986). The economic related variables that are considered in this study include crude oil price (in Malaysian Ringgit per Barrel) (OILP) and Consumer Price Index for transport (CPI). OILP is accessed from the World Bank website. It is calculated based on the simple average of three spot prices which are Dated Brent, West Texas Intermediate and Dubai Fateh.

CPI is computed based on number of vehicles purchased, operation of personal transport equipment (including spare parts, accessories or lubricant) and transport services. The data for this variable are gathered from monthly statistical bulletin provided by DOS. Both economic related variables above have been used as explanatory variables in this study to test whether they really influence road accidents frequency.

1.5.4 Seasonal Related Variables

Festival celebrations are usually caused more road accidents to occur. This is because the traffic suddenly becomes heavier because citizens return to their hometown (known as Balik Kampung) to visit their relative during the festivals. Such festivals include Chinese New Year, Eid-ul-Fitr, and Deepavali are determined based on the lunar calendar. The dates of these celebrations are not fixed every year and they change on yearly basis. Radin Umar *et al.* (1996) incorporated similar variables in measuring the effect of festival celebrations on motorcycle accidents. They applied dummy variable to represent this event and name the variable as Balik Kampung (BLKG) . It is coded "0" to represent not BLKG season and "1" to represent BLKG season. The study is sensible as it involved weekly data.

However, the BLKG which represents festival holidays are not absorbed by monthly dummies. Therefore this study applied one weight variable for moving holidays as in Shuja *et al.* (2007). From a survey made on 350 respondents, it is found that the number of off days that is usually taken for Eid- ul-Fitr was 7 days (2 days before festival and 5 days during and after the festival), 8 days for Chinese New Year (2 days before festival and 6 days during and after the festival) and 4 days for Deepavali (1 day before festival and 3 days during and after the festival). In this study, the variable to represent BLKG events were coded as in the expression below and example of the coding for this variable will be as in Table 1.5. In this study, BLKG variable only considered three main festivals that is Chinese New Year, Eid-ul-Fitr and Deepavali.

Case1: If the date of the festival falls in the beginning of the month (1st-15th), the weight value is define as follows

$$BLKG1 = \begin{cases} \dfrac{g_1}{m} & \text{in the respective festive month} \\ \dfrac{g_2}{m} & \text{before the respective month} \\ 0 & \text{otherwise} \end{cases}$$

where g_1 is the number of holidays that fall in the respective month, g_2 is the number of holidays before the respective month and m is the total of holiday ($m = 7$ for Eid-ul-Fitr, $m = 8$ for Chinese New Year and $m = 4$ for Deepavali).

Case2: If the date of the festival falls at the end of the month (16th-31st), the weight value is defined as follows

$$BLKG2 = \begin{cases} \dfrac{g_1}{m} & \text{in the respective festive month} \\ \dfrac{g_2}{m} & \text{after the respective month} \\ 0 & \text{otherwise} \end{cases}$$

where g_1 is the number of holidays that fall in respective month, g_2 is the number of holidays after the respective month and m is total of holiday ($m = 7$ for Eid-ul-Fitr, $m = 8$ for Chinese New Year and $m = 4$ for Deepavali).

Table 1.5: An example of BLKG coding

Year	Month	Festival	Date of festival	Ratio	BLKG
2004	1	Chinese New Year	22 -Jan	1	1.00
2004	2				0.00
2004	10				0.00
2004	11	Deepavalli Eid -ul -Fitr	12-Nov 14-Nov	1 1	2.00
2005	1				0.00
2005	2	Chinese New Year	9 –Feb	1	1.00
2005	9				0.00
2005	10			1/4	0.25
2005	11	Deepavalli Eid -ul -Fitr	1 Nov 4 Nov	3/4 1	1.75
2006	1	Chinese New Year	29 Jan	5/8	0.63
2006	2			3/8	0.37
2006	3				0.00
2006	10	Deepavalli Eid-ul-Fitr	21 Oct 24 Oct	1 1	2.00

For example, in 2006 Chinese New Year falls on 29 Jan, $g_1 = 5$ and $g_2 = 3$. Given in Figure 1.1 is an illustration of how to determine g_1 and g_2 as suggested by Shuja *et al.* (2007).

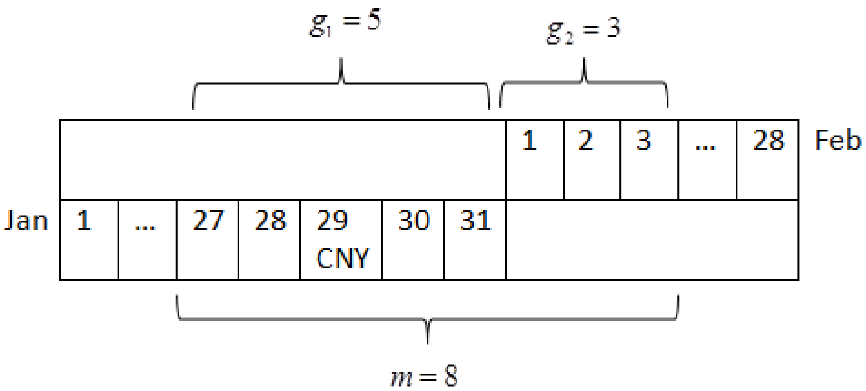

Figure 1.1: An illustration of process determining g_1 and g_2

1.5.5 Road Safety Related Variables

Other data that were also considered include the road safety related variable which is enforcement of road safety ,Ops Sikap (SAFE). Ops Sikap or Attitude Ops is a traffic safety operation carried out by Royal Malaysia Police to nurture peoples' safety awareness on all roads in Malaysia during festive seasons such as Eid-ul-Fitr, Deepavali, Christmas and Chinese New Year. This operation began in 2001 which involves the collaboration of Malaysian Road Transport Department (JPJ), Land Public Transport Comission (SPAD) and The National Anti-Drugs Agency (AADK).

Ops Sikap variable has been used by Wan Yaacob *et al.* (2011b) in examining its effect on road accidents in Malaysia. The study implement dummy variable "0" to represent no SAFE and "1" to represent SAFE operation. However, it is found that this notation will be quite not irrelevant if its date involves two consecutive months. In such cases this study suggests to use weight variable for SAFE where the representation of the Ops Sikap variable are based on the rate number of day the operation is carried out. The total of operation day for the enforcement of Ops Sikap for both Chinese New Year and Eid-ul-Fitr is 15 days. If SAFE involved two consecutive months, the total number of days of the operation on those months were divided by 15. While other months were coded as "0" to represent no Ops Sikap. Table 1.4 illustrates this case.

Table 1.6: 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

1.6 Limitation of the Study

The study fails to take into account the influence of some other important or relevant variables since these variables are either not available in monthly unit or there are not available in state by state basis. For example the data on gross domestic product (GDP) only available in quarterly, while the data for volume of traffic not collected in state by state basis.

As state in earlier section, the period of the study is from January 2001 up to Disember 2013. However, the variable of air pollution index (API) for Perlis and Selangor only can be retrieved from 2004 onwards. Therefore, the model of road accidents for these both states are developed based on data from year 2004 until 2013.

The study also, only cover univariate analysis with and without explanatory variables and no multivariate analysis has been developed. Besides, the prediction and forecasting of road accidents model only applicable for univariate time series model without explanatory variables as the lack of information of other explanatory variables for year above 2013. Furthermore, this study does not include mathematical proving since all the equations used are mostly taken from published literature.

1.7 Summary and Thesis Organization

This thesis is divided into seven chapters which include this introductory chapter, followed by literature review in Chapter 2, methodology in Chapter 3, the analysis and discussion of the result in Chapter 4 to Chapter 6 and conclusion of the thesis is in Chapter 7.

Chapter 1, the introductory chapter, presents the background of the research including the research problem followed by the objectives and significance of the study. Besides, the scope of the study which describes the variables used in this thesis is also presented in this chapter.

In Chapter 2, the background definitions of structural time series approach is given and the advantages of this technique is reviewed. Furthermore, previous literature on the application of common techniques to model road safety study especially road accidents occurrence is discussed. Chapter 2 is important for the understanding of some related idea in developing road accidents model in this thesis.

Chapter 3 is concerned with the statistical analysis or theoretical technique used in this thesis which includes descriptive statistics and correlation analysis. Moreover, this chapter discusses all common methods used in developing road accidents models as well as introducing the structural time series method in modeling road accident. This chapter also includes step by step procedure of developing road accidents model which is applied in this thesis.

Chapter 4 describes the properties of data collected based on descriptive statistics, time series plot and correlation analysis. Descriptive statistics is important in describing the basic feature of the data, while time series plot is useful in observing the basic pattern of the series such as trends and seasonality. The correlation analysis measures the strength of relationship among the variable. In addition, common time series methodology such as time series regression (TSR) and seasonal autoregressive integrated moving average (SARIMA) analysis are applied. This chapter is important as an early stage of the study before it is applied to the other analysis. In addition, common time series analysis used in this chapter will be compared with the other methods, which will be employed in the next two chapters.

Chapter 5 estimates the model for the number of road accidents using structural time series approach. The chapter begins with the model identification, followed by estimating the model for the number of road accidents model for five regions as well as for individual states in Malaysia. The statistical trend and seasonal pattern of each series is also observed as one of the objectives in this chapter. Next, the estimated road accidents models for the regions as well as the individual states are then compared with TSR and SARIMA model to measure their performance.

Next, the number of road accidents models is refitted in Chapter 6. However, the estimated model incorporates explanatory variables to investigate their influence to road accidents. The estimation of explanatory variables as well as their discussion will be thoroughly described. Besides, the stochastic trends and seasonal patterns after incorporating the explanatory variables and considering the outliers will be observed. The performance of the estimation model between STS and TSR will be discussed at this chapter.

The last chapter summarises the conclusion of this thesis from both theoretical and applied points of view. It also contains suggestion of further research related to the idea of this thesis.

CHAPTER 2

LITERATURE REVIEW

There are numerous statistical and mathematical methods that are introduced to model and predict the road safety. Some of the models are less sophisticated which could not describe the phenomenon or give a poor prediction. This chapter provides a historical perspective of the structural time series approach and the developments in road safety research empirically and methodologically.

2.1 Structural Time Series

In the beginning, structural model is developed as a traditional decomposition of time series component as a sum of trend, seasonal and irregular components (Harvey and Durbin, 1986).

$$Y_t = \mu_t + \gamma_t + \varepsilon_t \qquad t = 1,\ 2,\ \ldots,\ n \tag{2.1}$$

where Y_t denotes the $\boldsymbol{t}$ -th observation possibly after the logarithmic transformation and μ_t, γ_t, and ε_t are the trend, seasonal and irregular components. The trend component is simply deterministic linear model written as $\mu_t = c + v_t$ and the seasonal component, γ_t is the seasonal periodic function such as the number of month, quarter or week. Its limited application can be enhanced based on this form as many series have a better fit if its structures evolve overtime.

The fundamental thought of how this can be accomplished originated from Muth (1960) who considered the situation where there is no seasonality and trend occurred without slope but the level, μ_t varied over time in random walk giving the model.

$$Y_t = \mu_t + \varepsilon_t, \qquad \mu_t = \mu_{t-1} + \eta_t \tag{2.2}$$

where ε_t and η_t are independent white noise terms. Later, Theil and Wage (1964) and Nerlove and Wage (1964) extended the model by including a trend with slope that yielded local linear trend model. The model made both level, μ_t and slope, v_t components evolved overtime which gives the model below

$$\mu_t = \mu_{t-1} + v_{t-1} + \eta_t, \qquad v_t = v_{t-1} + \varsigma_t \tag{2.3}$$

where ς_t is a slope disturbance term that independent of ε_t and η_t.

In 1965, Schweppe (1965) showed that a likelihood function could be used to evaluate both models by using the Kalman filter via prediction error decomposition. However, a constraint in the computation technology in the 1960s made the results cannot be exploited properly. During that time, Box and Jenkins technique is the most influential time series methods. Box and Jenkins (1976) have observed that the first difference of Equation (2.2) and second difference of Equation (2.3) yield first order moving average process and second order moving averages process respectively. This has led to formulation of the class of ARIMA (Autoregressive integrated Moving Average) model class and the development of model selection strategy.

Although the ARIMA approach has dominated the time series literature in 1970s and 1980s, the structural approach was more prevalent in control engineering (Harvey, 2006a). It is largely due to the familiarity of the Kalman filter approach in control engineering area since the appearance of Kalman (1960). Kalman filter is a set of mathematical equations that recursively estimate the state parameters by minimizing mean square error (Welch and Bishop, 2006). Another advantage of the Kalman filter approach is that it can be used to construct complex models. In 1970s, an early example of application of Kalman filter approach in economic and statistical research can be found in Rosenberg (1973) on time varying parameters, and in 1980s in Young (1984); Harvey(1989); West and Harrison (1986) and Kitagawa and Gersch (1996).

In order to handle seasonal component in structural time series Harrison and Stevens (1971) suggest two general techniques which employ trigonometric model and time varying seasonal dummy. Besides, structural time series could also be extended by including explanatory variables and intervention variable which will be briefly explained in the next chapter.

2.2 Advantages of Structural Time Series

Structural time series has a direct interpretation in time series modeling and explanatory variable can be added in direct manner. The model can be put in state space form, and estimated by Kalman filter estimation technique (Harvey and Durbin, 1986). In addition, structural time series also has good performance in forecasting annual, quarterly and monthly data especially for long forecasting horizons and seasonal data. The forecasting results is quite reliable and accurate compared to others forecasting methods (Andrews, 1994).

Besides, structural time series make it easy to handle missing values and outliers once it is in state space form (Harvey, 1989). The missing values were estimated using Kalman filter approach while outliers were handled by including intervention variable. On the other hands, structural time series will model the seasonal and trend components compared to ARIMA model which eliminate both components using differencing of the original series (Jalles, 2009). This condition indicated that structural time series does not easily remove the important information from original series.

However, structural time series also have its flaws. Referring to Karlis and Hermans (2012), the structural model are usually more complicated and less interpretable compared to standard time series model. Besides, extra computational effort is needed and there is still a lack of statistical software that implement this approach.

2.3 Application of the Structural Time Series Model

Recent contribution on the application of structural time series can be seen in various applications such as economics, sociology, management science, operational research, geography meteorology and engineering (Harvey, 1989). This section will review some application of the structural time series approach in several disciplines.

2.3.1 Economics

In economics, the application of structural time series (STS) can be found in Thury and Witt (1998) which generates monthly forecast of Austrian and German industrial production. The specification of the STS model used in this study is basic