EXTRACTING RURAL CRASH INJURY AND FATALITY PATTERNS DUE TO CHANGING CLIMATES IN RITI COMMUNITIES BASED ON ENHANCED DATA ANALYSIS AND VISUALIZATION TOOLS (PHASE II)

FINAL PROJECT REPORT

by

Guohui Zhang, Ph.D., Hanyi Yang, Ph.D., Hao Yu, Ph.D., Zhenning Li, Ph.D., Rong Zou, Ph.D., Runze Yuan, Panos D. Prevedouros, Ph.D., and Tianwei Ma, Ph.D.

Department of Civil and Environmental Engineering
University of Hawaii at Manoa

for

Center for Safety Equity in Transportation (CSET)
USDOT Tier 1 University Transportation Center
University of Alaska Fairbanks
ELIF Suite 240, 1764 Tanana Drive
Fairbanks, AK 99775-5910

In cooperation with U.S. Department of Transportation,
Research and Innovative Technology Administration (RITA)

CSET
rural • isolated • tribal • indigenous

DISCLAIMER

TECHNICAL REPORT DOCUMENTATION PAGE		
1. Report No.	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Extracting Rural Crash Injury and Fatality Patterns Due to Changing Climates in RITI Communities Based on Enhanced Data Analysis and Visualization Tools (Phase II)		**5. Report Date** April 30, 2022 **6. Performing Organization Code**
7. Author(s) and Affiliations Guohui Zhang, Ph.D., P.E., Hanyi Yang, Ph.D., Hao Yu, Ph.D., Zhenning Li, Ph.D., Rong Zou, Ph.D., Runze Yuan, Panos D. Prevedouros, Ph.D., P.E., and Tianwei Ma, Ph.D. Department of Civil and Environmental Engineering University of Hawaii at Manoa		**8. Performing Organization Report No.**
9. Performing Organization Name and Address Center for Safety Equity in Transportation ELIF Building Room 240, 1760 Tanana Drive Fairbanks, AK 99775-5910		**10. Work Unit No. (TRAIS)** **11. Contract or Grant No.**
12. Sponsoring Organization Name and Address United States Department of Transportation Research and Innovative Technology Administration 1200 New Jersey Avenue, SE Washington, DC 20590		**13. Type of Report and Period Covered** Research Report **14. Sponsoring Agency Code**
15. Supplementary Notes Report uploaded to:		

16. Abstract

This report documents the research activities to investigate the traffic crashes in Rural, Isolated, Tribal, or Indigenous (RITI) communities involving considerable incapacitating injuries and fatalities. The traffic crashes occurring in RITI communities, are different from urban traffic crashes, and are related more to the features like speeding, low application of safety devices (for instance, seatbelt), adverse weather conditions and lacking maintenance and repairs for road conditions, and inferior lighting conditions. Thus, it is necessary to study the properties and attributes of traffic crashes at the RITI area using data analysis methods, such as statistical methods, and data-driven methods. This project is trying to analyze the rural crash injury and fatality patterns caused by changing climates in RITI communities based on enhanced data analysis using latest mathematical method. The mixed logit model to examine the risk factors in determining driver injury severity in four crash configurations in two-vehicle rear-end crashes on state roads based on seven-years of data from the Washington State Department of Transportation. The differences between the MLM and the LCM are investigated for exploring the relationships between driver injury severity in the rain-related rural single-vehicle crash and its corresponding risk factors. Moreover, this project develops a latent class mixed logit model with temporal indicators to investigate highway single-vehicle crashes and the effects of significant contributing factors to driver injury severity. The results of this research will be beneficial to transportation agencies to propose effective methods to improve rural crash severities under special climate and weather conditions and minimize the rural crash risks and severities.

17. Key Words		**18. Distribution Statement**	
Traffic safety, Fatality pattern, injury severity, Changing climate, Rural area			
19. Security Classification (of this report)	**20. Security Classification (of this page)**	**21. No. of Pages**	**22. Price**
Unclassified.	Unclassified.	75	N/A

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized.**

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS				
Symbol	**When You Know**	**Multiply By**	**To Find**	**Symbol**
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS				
Symbol	**When You Know**	**Multiply By**	**To Find**	**Symbol**
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380. (Revised March 2003)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

This report documents the research activities completed to investigate the traffic crashes in Rural, Isolated, Tribal, or Indigenous (RITI) communities involving considerable incapacitating injuries and losses. The traffic crashes occurring in RITI communities are related more to the features like speeding, low safety devices application (for instance, seatbelt), adverse climate and weather conditions and lacking maintenance and repairs for road conditions, and inferior lighting conditions than urban crashes. Thus, it is necessary to study the properties and attributes of traffic crashes in the RITI areas using statistical methods and data-driven methods. Unfortunately, there exists not only the unobserved heterogeneities but also the temporal instability in traditional crash data analysis.

To solve this problem, the project analyzed the rural crash injury and fatality patterns caused by changing climates in RITI communities based on enhanced data analysis using latest mathematical methods. The mixed logit model (MLM) used to examine the risk factors in determining driver injury severity in four crash configurations in two-vehicle rear-end crashes on state roads was based on seven-years of data from the Washington State Department of Transportation. The dataset only includes collisions with passenger cars and pickup trucks involved. These vehicles are the most common in these crashes, and the two types typically have different heights and masses. Four crash configurations are examined concerning the type of the vehicles and their relative position in a crash. Four models for these configurations and a model for the overall dataset are estimated. In addition, this project developed a latent class mixed logit model (LCM) with temporal indicators to investigate highway single-vehicle crashes and the effects of significant contributing factors to driver injury severity. The differences between the MLM and the LCM are investigated for exploring the relationships between driver injury severity in the rain-related rural single-vehicle crash and its corresponding risk factors. The results of this research will be beneficial to transportation agencies to propose effective methods to improve rural crash severities under special climate and weather conditions and minimize the rural crash risks and severities.

Due to limited visibility and low skid resistance on the road surface, single-vehicle crashes in rain, especially those that occurred in rural areas, are more likely to result in driver incapacitating injuries and fatalities. A three-year crash dataset including all rural single-vehicle crashes under rainy conditions from 2012 to 2014 in four South Central states, i.e., Texas, Arkansas, Oklahoma, and Louisiana, was selected in this paper to analyze the impact factors on driver injury severity. The MLM and LCM are developed on the same dataset. Several parsimony indices, e.g., the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), and as well as McFadden pseudo r-squared, are calculated for all the models to evaluate their respective performance. Results show that choosing the uniform distribution as the prior for random parameters improves the goodness-of-fit of the MLM more than using normal and lognormal distributions. In addition, the two-class LCM also shows superiority when compared to three- and four-class LCMs. Finally, a careful comparison between these two models is conducted, and the results indicate that the LCM has a slightly better performance in analyzing the aforementioned dataset in this study.

CHAPTER 1. INTRODUCTION

1.1. Problem Statement

Statistical data indicates that more than 50% of fatalities in crashes occurred on rural roadways, resulting in more than 20,000 people losing their lives annually. Rural crashes also cause considerable incapacitating injuries and losses in rural, isolated, tribal, or indigenous (RITI) communities. The fatality rate in rural areas is double the rate in urban areas in the US Department of Transportation (USDOT) 2013 National Highway Traffic Safety Administration (NHTSA) report. The Hawaii Department of Transportation (HDOT) also reported that the rural area fatality rate is 195% larger than that in the urban areas in 2014.

The traffic crashes occurring in RITI communities are different from urban traffic crashes. Many significant impact factors, including speeding, low safety device use (for instance, seatbelt), adverse weather conditions and lacking maintenance and repairs for road conditions, and inferior lighting conditions, contribute to more severe crashes in RITI communities. Thus, it is necessary to study the properties and attributes of traffic crashes in the RITI areas using data analysis methods, such as statistical methods, and data-driven methods. Unfortunately, few studies have been conducted to address unobserved heterogeneity and temporal instability issues in traditional crash analyses.

To address the research gap, the project aimed to: 1) Employ the mixed logit model (MLM) to examine the risk factors in determining driver injury severity in four crash configurations in two-vehicle rear-end crashes on state roads based on seven-years of data from the Washington State Department of Transportation; 2) Develop a latent class mixed logit model (LCM) with temporal indicators to investigate highway single-vehicle crashes and the effects of significant contributing factors to driver injury severity; and 3) Investigate the differences between the MLM and the LCM for exploring the relationships between driver injury severity in the rain-related rural single-vehicle crash and its corresponding risk factors.

The research enables effective traffic safety program management at all levels in RITI communities by aiding in the design and implementation of appropriate countermeasures to mitigate rural crash severities and risks. Weather conditions are identified as a significant impact factor which indirectly reflects climate change impacts on crash outcomes and injury severity. The updated crash data platform will provide more interesting functions, and the Bayesian approach and finite mixture random parameter models make fundamental contributions to the crash data analysis in RITI communities.

The analytical results of the rural crash data records will greatly facilitate active countermeasure development to minimize crash risks and severities in RITI communities. To our best knowledge, based on a thorough literature search, there is no existing literature focusing on investigating the driver injury severity patterns in low-visibility-related crashes considering finite mixture random effects, and on interpreting with missing values, which motivated us to conduct fundamental methodological analysis for rural crash characteristics in RITI communities.

1.2. Research Objectives

This project aimed at developing a variety of statistical models to analyze the specific types of traffic crashes, extracting the critical factors causing them, and proposing a novel finite mixture random parameter model for driver injury severity analysis in RITI communities. Towards this goal, the research objectives were as follows:

- Finding a statistical model to examine the risk factors in determining driver injury severity in four crash configurations in two-vehicle rear-end crashes on state roads based on seven-years of data from the Washington State Department of Transportation.
- Investigating highway single-vehicle crashes and the effects of significant contributing factors to driver injury severity.
- Analyzing the differences between the MLM and the LCM in exploring the relationships between driver injury severity in the rain-related rural single-vehicle crash and its corresponding risk factors.

1.3. Report Organization

The remainder of this report is organized in the following manner. Chapter 2 presents a comprehensive review of previous studies that are relevant to this study, including studies focusing on crash modeling, characteristics in crash modeling, and other critical issues, such as temporal instability and missing data. Chapter 3 employs the mixed logit model to examine the risk factors in determining driver injury severity in four crash configurations in two-vehicle rear-end crashes on state roads based on seven-years of data from the Washington State Department of Transportation. Chapter 4 develops a latent class mixed logit model with temporal indicators to investigate highway single-vehicle crashes and the effects of significant contributing factors to driver injury severity. Chapter 5 investigates the difference between the MLM and the LCM for exploring the relationships between driver injury severity in the rain-related rural single-vehicle crash and its corresponding risk factors. Finally, Chapter 6 presents the conclusion of this research and the recommendations for future research.

CHAPTER 2. LITERATURE REVIEW

2.1. Generic Crash Severity Modeling and Analysis

Many studies have pointed out the distinct mechanisms between single-vehicle and multi-vehicle crashes and the corresponding different relevant factors in determining severity level (Ivan et al., 1999; Geedipally & Lord, 2010; Yu et al., 2013a; Yu & Abdel-Aty, 2013; Wu et al., 2014; Dong et al., 2018; Chen & Chen, 2011). For example, by comparing the impacts of single and multi-vehicle crashes, Dong et al. (2018) found that relevant factors for injury in multi-vehicles crashes include being a weekend, speed limits, traffic volume, number of trucks involved, wet road surface, and even the month of the year, but visibility is the only factor for injury for single-vehicle crashes. Furthermore, among all the multi-vehicle crashes, two-vehicle (2V) crashes are the most common (NHTSA, 2019). Through a detailed analysis, Kitali et al. (2021) verified the necessity to model 2V and more than two-vehicle crashes separately because of their different contributing factors. Their results suggest that disaggregating 2V and other multi-vehicle crashes while allowing correlation best describes their data. In a rear-end collision, where a neck injury is easily caused by the sudden acceleration of the body with respect to the brain, the driver in the leading vehicle is found to be more severely injured in 2V crashes, whereas the driver in the middle car is more severely injured in three-vehicle-involved crashes (Khattak, 2001; Yasmin, 2014). The critical factors in 2V crashes have been determined for different circumstances. For example, the driver's age, vehicle type, alcohol use, intersection, and lighting conditions were found to be key risk factors at signalized intersections in Taipei City (Chiou et al., 2013); female drivers, older (≥65 years) drivers, unbuckled drivers, speeding drivers and drivers in lighter and older vehicles suffer higher injury risks in 2V crashes (Gong et al., 2021). Generally, vehicle types play an important role in influencing the severity of injuries. The size of the vehicle can control the visibility of the following cars, and the vehicle's mass changes the degree of the vehicle's sudden acceleration, causing injury (Abdel-Aty, 2003; Erbulut, 2014). Motorcycles or trucks lead to more severe driver injuries and fatalities in multi-vehicle crashes (Wu et al., 2014). Vehicles with larger mass, such as vans, pickup trucks, and station wagon cars, tend to have less severe injuries in a rear-end crash (Khattak, 2001). Abdel-Aty & Abdelwahab (2003) considers crashes involving only light truck vehicles and passenger cars and the four configurations in the rear-end crashes based on the types of leading and following vehicles. They analyze the key factors to determine the type of configuration. For example, their results indicate that driver distraction and limited sight distance predicted the crash was caused by a regular passenger car following a light truck. Their results reveal that each configuration is associated with a different mechanism. This motivates our research to further study the critical factors in determining the crash severity level of the 2V rear-end crashes in these four configurations, respectively, with unobservable heterogeneity also considered. Rear-end crashes are the second most common crash type, right behind head-on crashes, and passenger cars and light trucks are involved in more than 50% of rear-end crashes (NHTSA, 2019).

For the severity analysis, the multinomial logit (MNL) method and multinomial probit method are widely used (Train, 2009; McFadden,1981; Lee & Mannering, 2002; Shankar & Mannering,1996; Mannering & Bhat, 2014) as well as ordered logit and ordered probit (Chiou et al., 2013; Yamamoto & Shankar, 2014; Chen et al., 2016; Greene, 2012; McCullagh,1980). To address the heterogeneity of the factors to be analyzed, the mixed MNL method, Latent Class MNL method are developed (Haleem & Gan, 2013; Wu et al., 2014; Ye & Lord, 2011; Mannering & Bhat, 2014). As a general statistical model, Mixed MNL can approximate any random utility model with the heterogeneity among individuals, time, and correction between alternatives (Train, 2009; McFadden & Train, 2000). Risk factors in specific circumstances are usually identified by comparing those results to results from the other circumstance. For example, the risk factors unique to large trucks were determined by MNL models developed for truck- and non-truck-involved accidents, respectively (Chang & Mannering, 1999). Wu et al. (2016) used the MNL model to explore the characteristic differences between teenage and adult drivers in intersection-related crashes. They found relevant factors of alcohol, driving technique, and seat belt usage have a significant difference between these two groups. Yu et al. (2020) built mixed MNL models for snow and other weather conditions to compare the different impacts and found that gender and sobriety have significantly higher pseudo-elasticity effects under snow weather than the other weather conditions in rural single-vehicle crashes. Behnood et al. (2014) adopted the Latent class MNL model and examined the differences in driver-injury severity between drivers impaired and not-alcohol-impaired and found substantial differences across age/gender groups in the absence/presence of alcohol. Behnood et al. (2017) used the MNL model with heterogeneity. They found that the different combinations of driver and passenger's age and gender affect the severity levels in single-vehicle crashes. In Ulfarsson & Mannering (2004), separate multivariate multinomial logit models of injury severity were estimated for male and female drivers.

Compared with urban areas, rural areas have more fatal traffic crashes and fatalities. NHTSA revealed that rural areas held only 19% of the total U.S. population but induced 48% of fatal traffic crashes and 49% of traffic fatalities in 2015 (NHTSA, 2017). Therefore, rural traffic crashes have drawn worldwide research interest (Islam and Brown, 2017; Rusli et al., 2017; Wang et al., 2017). Similarly, single-vehicle crashes are also found to be more fatality-concentrate than multi-vehicle crashes, as was evidenced by the fact that single-vehicle crashes accounted for 28.9% of all crashes, but 58.1% of all fatal crashes in the U.S. in 2015 (NHTSA, 2017). Given these two fatality-concentrated features of rural crashes and single-vehicle crashes, it is of particular interest to investigate the injury severity patterns in rural single-vehicle crashes, a sub-crash type that belongs to both these two fatality-concentrated crash types. Injury severity in traffic crashes is an indispensable research area in crash data analysis, in addition to crash frequency analysis, and driver injury severity has been widely used as a representative indicator at the individual level (Behnood and Mannering, 2017; Chen et al., 2016b, 2015a; Seraneeprakarn et al., 2017). Numerous studies have already been conducted to investigate the contributing factors and propose effective countermeasures to mitigate driver injury severity in traffic crashes (Chang and Chien, 2013; Chen et al., 2016c, 2015a; Kim et al., 2013; Wu et al., 2014).

Previous studies showed that driving in the rain may be associated with higher crash risk than that in clear weather (Jung et al., 2010). A sizable portion of severe traffic crashes is brought about by these issues and induces significant fatalities and serious injuries. According to the Texas Department of Transportation (TxDOT, 2016), 16,818 rural crashes (159 fatal crashes) occurred under rain conditions in 2015, which is four times as many as those related to all other inclement weather conditions (e.g., blowing sand, sleet, and hail). In addition, crash statistics from Arkansas and Oklahoma (Arkansas Department of Transportation, 2015; Oklahoma Department of Transportation, 2015) showed that single-vehicle crashes under rain conditions, especially those occurring in rural areas, have a probability of drivers being seriously injured approximately twice as high as that for multi-vehicle crashes occurring under the same or similar conditions. However, in most traffic safety studies, weather conditions have been considered as a contributing factor in crash cause-effect analysis, and only a limited number of studies directly focused on crashes under rain conditions. Andrey and Yagar (1993) analyzed the crash risk during and after rain events in urban areas and discovered that the overall crash risk under rain conditions is 70% higher than that in average-day clear conditions. Jung et al. (2010) developed two types of polychotomous response models to analyze rain-related crashes in Wisconsin and concluded that rain-related factors could significantly affect injury severity. However, the safety impacts of rain and other variables in rain-related crashes are found to be unstable among different studies. For instance, a study examining the temporal and spatial distribution of rain-related crashes in Texas suggested that rain is a contributor to fatal crashes only in a few dry counties but has no impacts on crashes in some of the wetter counties (Jackson and Sharif, 2014). Qiu and Nixon (2008) reported that rain is associated with higher injury severity and crash rates. Feng et al. (2016) concluded that severe accidents are about twice more likely to occur on curved roadways on rainy days, although straight and curved roadways have similar impacts in clear days. Shaheed et al. (2016) also reported that gender, seating position, road junction type, and other risk factors have different effects on injury severity in weather-related (rain, snow, blowing sand, etc.) and non-weather-related crashes. Whereas in the article of Lee et al. (2015), estimation results showed that injury severity is relatively lower under rain condition in all crash types since drivers tend to reduce their speeds and be more careful on a wet surface. The sophisticated influences of rain on overall traffic safety indicate that there is a need for detailed analyses regarding external weather conditions and collision types.

In contemporary traffic safety research, unobserved heterogeneity of the police-reported crash dataset has been recognized as a critical issue (Mannering et al., 2016; Mannering and Bhat, 2014). In this study, the dataset was obtained from a representative sample of police-reported motor vehicle crashes with discrete injury severity outcomes, where there are still certain elements related to crashes not recorded and remaining unobserved to researchers, even though many elements (e.g., driver age, gender, number of lines of a roadway, etc.) have been covered. Traditional models that are usually used in traffic crash data analysis, e.g., multinomial logit model (MNL), ordered logit model, etc., cannot adequately address the unobserved heterogeneity within such dataset. Therefore, models that can account for unobserved heterogeneity should be developed. One of the first practices on unobserved heterogeneity model was a study by Kim et al. (2008), where they developed a mixed logit model (MLM) to investigate the effect of driver age on driver injury severity outcome in single-vehicle crashes. A further study on driver injury severity also conducted by Kim et al. (2013) demonstrated that MLMs are superior to traditional discrete choice models in that they are more flexible and can approximate any random utility model. Wu et al. (2014) developed MLMs to analyze driver injury severities in single-vehicle crashes and compared the results with multi-vehicle crashes. Noticeably, elasticity analyses and transferability tests were applied to discuss the models' parameters estimation outcomes, and the results showed that elasticity analysis is a necessary supplement to MLMs. The MLM can account for individual unobserved heterogeneity by allowing parameters to vary across observations and therefore yield more reliable estimations (Kim et al., 2013; Milton et al., 2008; Moore et al., 2011). It should be mentioned that some recent studies tried to explicitly examine the possible heterogeneity in means and/or variance. For instance, Behnood and Mannering (2017) adopted an MLM with heterogeneity in parameter means to explore the differences in driver-injury severities. Seraneeprakarn et al. (2017) developed an MLM of injury severity while allowing for heterogeneity in parameter means and variances. Models with no mean-variance related heterogeneity, and with mean related heterogeneity only, are also developed and compared with the proposed model. The estimation results showed that for their dataset, the proposed model has better performance over the other two, and some variables were found to randomly distributed with significant heterogeneity in both means and variances. Interested readers are referred to these papers and the references cited therein. However, this model also has its own drawbacks. Due to its flexible structure, the MLM requires appropriate distribution assumptions for potential random parameters; otherwise, these random effects may remain undetected. This restriction is released in another widely used method, the latent class model (LCM), where specific distributions for parameters of interest are not required. Instead, determining a proper number of classes becomes a critical step when using this approach. The unobserved heterogeneity is then identified by these different classes with homogeneous characteristics of the within-class observations (Gelman and Hill, 2007; Ma et al., 2016; Mannering et al., 2016). Xie et al. (2012) developed the LCM to deal with the single-vehicle crashes, and concluded that the LCM has the potential to overcome the problems associated with the irrelevant alternatives (IIA) property that commonly exists in multinomial logit models (Abdel-Aty, 2003). Both MLMs and LCMs were developed on a pedestrian-injury dataset to ensure reliable estimation, and the results showed that both models are appropriate to capture unobserved heterogeneity (Behnood and Mannering, 2016). However, there are only a few references that have directly compared the MLM and the LCM for driver injury severity analysis, and the comparison is not always comprehensive. For instance, Cerwick et al. (2014) compared the two models by their model fit, inferences, and predicted crash severity outcome probabilities by a large sample of crash data on multiple vehicle crashes. Behnood and Mannering (2016) developed both the MLM and the LCM to study the risk factors on the pedestrian crash dataset from Chicago city. However, most of the previous studies did not explicitly conclude which model is superior to the other.

Another issue in traffic safety analysis is that analysts always aggregated the crash data over a specified time period to gather sufficient observations for analyzing. However, some recent research suggests that the impact of factors affecting injury severity may not be temporally stable (Behnood and Mannering, 2015; Mannering, 2018). In our dataset, the proportion of different types of crashes each year is not constant. The occurrence of rain-related crashes may have relationships with the weather or even climate change, both of which are virtually impossible to measure with existing data sources. In addition, another potential problem is the fact that driver involved in rain-related crashes may be a non-random sample since safer drivers may choose to take other modes of travel due to compromised road friction and visibility. Ignoring possible temporal effects may adversely affect the inferences drawn from model estimations as well as their ability to be used to forecast and evaluate the effects of safety countermeasures (Mannering, 2018). However, both MLM and LCM are not able to explicitly distinguish that the unobserved heterogeneity revealed by these models are entirely due to temporal instability, or a combination of temporal shifts and other traditional sources of unobserved heterogeneity. An article by Behnood and Mannering (2016) provided some insightful technique details to determine whether there exists temporary instability or not in the estimates of unobserved heterogeneity models. A series of likelihood ratio tests were conducted to compare models developed for two time periods and examine if the parameter estimates are stable between these periods. This technique is also adopted in this study for the aim of temporal stability testing and model comparison of the MLM and the LCM.

2.2. Random Parameters Models in Traffic Crash Analyses

Single-vehicle crashes pose increasing challenges in traffic safety. For instance, from 2015 to 2016, fatalities in single-vehicle crashes increased by 1180, a 5.9% national-wide increase (National Highway Traffic Safety Administration, 2017). In addition, in 2017, there were 11109 fatalities caused by single-vehicle crashes national-wide, accounting for 46.5% of all crash fatalities according to Insurance Institute for Highway Safety statistics (2017). Factors affecting the severity of single-vehicle crashes, such as crash exposures, road geometries, and driver features, have been explored extensively in previous studies (Behnood and Mannering, 2015; Gong and Fan, 2017; Kim et al., 2013; Lee and Mannering, 2002; Lee and Li, 2014; Li et al., 2018b; Savolainen and Mannering, 2007; Shaheed and Gkritza, 2014; Wu et al., 2016b; Xie et al., 2012; Yu and Abdel-Aty, 2013). For example, Lee and Mannering (2002) modelled the severity of run-off-road crashes using a nested-logit model considering a combination of temporal indicators, driver status, environmental characteristics, and roadway conditions. Xie et al. (2012) investigated the impact factors for rural single-vehicle crashes via a latent-class logit model. Compared to Lee and Mannering (2002), additional information such as crash types, lighting conditions, and in-vehicle protections, were involved in Xie's model (2012). Kim et al. (2013) proposed a random parameter logit model to analyze unobserved heterogeneous effects of drivers' age and gender on injury severities in single vehicle crashes.

In the past few years, unobserved heterogeneity received growing concerns. Note that a great part of factors affecting crash severity are not available in post-crash observation, such as the mental status of deceased drivers, or not included in crash record, such as dynamic traffic flow conditions. Unobserved factors are correlated with both the crash outcome and observed factors. These factors thus lead to potential variations in the impacts of observed ones on crash severity, which constitute unobserved heterogeneity (Mannering et al., 2016). Random parameters approaches and their variants, such as random parameters model (Kim et al., 2013; Li et al., 2019b; Wu et al., 2016b), random parameters ordered probability models (Eluru et al., 2008; Fountas et al., 2018), latent class models with random parameters within classes (Li et al., 2018b; Liu and Sharma, 2018), and mixed logit models with heterogeneity in means and variances (Alnawmasi and Mannering, 2019), have been the most frequently used methods in coping with the unobserved heterogeneity in crash severity analysis (see Mannering et al. (2016)).

Moreover, it is commonly understood that researchers employed crash records collected in multiple years in their studies in order to obtain a large enough sample size (Elvik, 2008; Washington et al., 2010; Yu et al., 2014). Statistical analysis was then applied on the multi-year dataset to investigate the impacts of various factors on crash severities. In most of these studies, it was implicitly assumed that the effects of the statistically identified factors are constant over time, i.e., temporally stable (Lord and Mannering, 2010; Mannering and Bhat, 2014; Mannering, 2018). However, several studies found that the impacts of factors on injury severity of highway crashes may vary over time. For instance, Wu et al. (2016b) tested the temporal transferability of their model using crash data collected in different years. The test results indicated that the transferability of estimated parameters was rejected at a significance level of 0.001, indicating there existed a temporal instability issue in the dataset. In a Markov switching model proposed by Xiong et al. (2014), evidence supporting temporal instability was found via allowing varying random parameters to be time dependent. Similar findings were also supported by works of Wu et al. (2016a) and Venkataraman et al. (2016). With respect to individual-specified crash records in a sufficiently long period, not only the observed explanatory variables associated with crash severity may change, but also the unobserved factors, such as individuals' cognitive biases, attitudes and behavior patterns, can evolve over time (Mannering, 2018). Ignoring possible temporal effects in crash severity analysis could adversely affect the inferences drawn from model estimation. Interested readers are referred to the article of Mannering (2018) and references cited therein for a comprehensive discussion of temporal considerations. Besides the numerical evidences provided in studies mentioned above, temporal instability has also been addressed in many crash studies by accommodating temporal correlations in different forms, such as temporal indicator in linear explanatory variables (Holdridge et al., 2005; Lee and Mannering, 2002; Sze and Wong, 2007; Cerwick et al., 2014), autoregressive correlation structure (Huang et al., 2009; Wang et al., 2006), time-varying intercepts (Cheng et al., 2017) and temporal structures in random effects (Li et al., 2019a; Liu and Sharma, 2018; Zeng et al., 2018). However, when it comes to the driver injury severity analysis domain, limited efforts have been conducted to investigate temporal instability of various impact factors, i.e., crash injury severity analysis allowing time-varying interactions among variables.

2.3. Summary

Recent studies on crash modeling, impact factor analyses on crash injury severity, and other critical issues in the crash analysis were reviewed in this section. In this study, the project team will employ the MLM to examine the risk factors in determining driver injury severity in four crash configurations in two-vehicle rear-end crashes on state roads based on seven-years of data from the Washington State Department of Transportation (WSDOT); develop an LCM with temporal indicators to investigate highway single-vehicle crashes and the effects of significant contributing factors to driver injury severity, and investigate the differences between the MLM and the LCM for exploring the relationships between driver injury severity in the rain-related rural single-vehicle crash and its corresponding risk factors. The proposed research enabled effective traffic safety program management at all levels in RITI communities to design and implement appropriate countermeasures to mitigate rural crash severities and risks.

CHAPTER 3. SEVERITY ANALYSIS OF TWO-VEHICLE REAR-END CRASHES OF DIFFERENT CONFIGURATIONS BY MIXED LOGIT MODELS

This chapter of the report employs the mixed logit model (MLM) to examine the risk factors in determining driver injury severity in four crash configurations in two-vehicle rear-end crashes on state roads based on seven-years of data from the WSDOT. The dataset only includes collisions that involved passenger cars and pickup trucks. These vehicles are the most common in these crashes, and the two types typically have different heights and masses. Following the configuration analysis by Abdel-Aty & Abdelwahab (2003), four MLMs are constructed for four configurations of 2V crashes, crash of two passenger cars (PP crash), crash of two pickup trucks (TT crash), a crash of a passenger car followed by a pickup truck (PT crash), a crash of a pickup truck followed by a passenger car (TP crash), and for comparison purposes, a model of the overall data is also constructed. The impacts of risk factors are compared among those five models, and their differences and similarity are addressed in-depth with the assistance of the elasticity analysis.

3.1. Data

The study is based on a dataset of rear-end 2V crashes by forward-moving passenger cars and pickup trucks on urban divided two-way roads extracted from traffic crash records over seven years from 2010 to 2016 from the Washington State Department of Transportation. In the dataset, the severity is classified into three types: no injury (N), injury (I), and fatality (F), where serious injuries and deaths are both included in the fatality (F) to maintain a statistically meaningful sample size (Yu et al. 2020). The total dataset is divided into four sub-datasets: PP crash dataset, PT crash dataset, TP crash dataset, and TT crash dataset.

Information on the crashes included in the dataset can be grouped into four categories, including: (1) crash information (e.g., driver injury severity category, temporal information, and crash location); (2) environmental information (e.g., weather, light conditions, road surface condition, weather condition, road characteristics, and indicators for work zone); (3) driver information (e.g., gender, age, seat belt usage, license status, insurance, and sobriety conditions); (4) vehicle information (e.g., airbag condition and vehicle's position in a crash). Efforts are devoted to pre-processing the selected crash datasets to define dummy variables of binary values, i.e., Yes (1) or No (0), which includes carefully combining similar variables and decomposing variables with continuous values into several variables. For instance, for driver age, three variables are determined, i.e., Age under 24 (including 24), Age above 65 (including 65), and Others (Age between 25 and 64) based on practical experience (Chen et al., 2016; Yu et al. 2020).

Erroneous or fragmentary data were removed. Finally, there are 23099 2V crash cases, including 8264 PP crashes, 4148 TT crashes, 3925 PT crashes, and 6762 TP crashes. The PP crashes constitute more than 1/3 of the total crashes, and TP crashes are slightly less than 1/3, while TT and PT crashes include the remaining approximately 1/3 of the total crashes. A detailed statistical description of the selected dataset is presented in Table 3-1.

Table 3-1 Variable Definitions and Descriptive Statistics of 2V crashes

Variables	Driver's severity						Total	%
	(N)	%	(I)	%	(F)	%		
PP crashes	14227	86.08%	2470	14.94%	11	0.07%	16528	35.78%
TT crashes	7159	86.29%	1136	13.69%	1	0.01%	8296	17.96%
PT crashes	6640	84.59%	1205	15.35%	5	0.06%	7850	17.00%
TP crashes	11795	87.22%	1723	12.74%	6	0.04%	13524	29.27%
Week								
Weekday	32543	85.69%	5418	14.27%	15	0.04%	37976	82.20%
Weekend	7098	86.33%	1116	13.57%	8	0.10%	8222	17.80%
Season								
Spring	9158	85.93%	1492	14.00%	7	0.07%	10657	23.07%
Summer	10038	85.26%	1727	14.67%	8	0.07%	11773	25.48%
Fall	11347	85.69%	1890	14.27%	5	0.04%	13242	28.66%
Winter	9098	86.43%	1425	13.54%	3	0.03%	10526	22.78%
Weather								
Clear	23449	85.70%	3901	14.26%	13	0.05%	27363	59.23%
Unclear	9360	86.65%	1437	13.30%	5	0.05%	10802	23.38%
Others	6832	85.05%	1196	14.89%	5	0.06%	8033	17.39%
Surface Condition								
Dry	27438	85.51%	4633	14.44%	15	0.05%	32086	69.45%
Wet	11736	86.57%	1814	13.38%	7	0.05%	13557	29.35%
Snow and Ice	289	85.25%	50	14.75%	0	0.00%	339	0.73%
Others	178	82.41%	37	17.13%	1	0.46%	216	0.47%
Lightning Condition								
Daylight	28983	86.15%	4646	13.81%	13	0.04%	33642	72.82%
Dawn and Dust	1966	87.49%	280	12.46%	1	0.04%	2247	4.86%
Dark with Light	6762	84.84%	1202	15.08%	6	0.08%	7970	17.25%
Dark without Light	1841	82.59%	385	17.27%	3	0.13%	2229	4.82%
Others	89	80.91%	21	19.09%	0	0.00%	110	0.24%
Roadway Characteristics								
Straight and Level	25856	86.37%	4067	13.58%	15	0.05%	29938	64.80%
Straight but not Level	9824	84.24%	1833	15.72%	5	0.04%	11662	25.24%
Curve and Level	1432	86.21%	229	13.79%	0	0.00%	1661	3.60%
Curve but not Level	1652	84.76%	294	15.08%	3	0.15%	1949	4.22%
Others	877	88.77%	111	11.23%	0	0.00%	988	2.14%
Work Zone								
Yes	1415	85.29%	243	14.65%	1	0.06%	1659	3.59%
Gender								
Male	23606	88.68%	3003	11.28%	9	0.03%	26618	57.62%
Female	15793	81.82%	3495	18.11%	14	0.07%	19302	41.78%
Others	242	87.05%	36	12.95%	0	0.00%	278	0.60%
Age								
(0,24]	9776	89.76%	1109	10.18%	6	0.06%	10891	23.57%
[25,64]	27591	84.64%	4993	15.32%	15	0.05%	32599	70.56%
[65,100)	2230	83.87%	427	16.06%	2	0.08%	2659	5.76%
Others	44	89.80%	5	10.20%	0	0.00%	49	0.11%
Belt Use								
Yes	38900	85.83%	6404	14.13%	19	0.04%	45323	98.11%
No	138	83.13%	27	16.27%	1	0.60%	166	0.36%

Variables	Driver's severity						Total	%
	(N)	%	(I)	%	(F)	%		
Others	0		0		0		0	0.00%
Sobriety								
Not Drinking	38577	85.83%	6350	14.13%	20	0.04%	44947	97.29%
HBD Impaired	399	84.71%	71	15.07%	1	0.21%	471	1.02%
HBD Not Impaired	80	78.43%	22	21.57%	0	0.00%	102	0.22%
Others	585	86.28%	91	13.42%	2	0.29%	678	1.47%
License Possession								
Licensed	39190	85.79%	6469	14.16%	23	0.05%	45682	98.88%
Unlicensed	164	85.86%	27	14.14%	0	0.00%	191	0.41%
Others	287	88.31%	38	11.69%	0	0.00%	325	0.70%
Airbag Status								
Airbag Ejected	8	57.14%	6	42.86%	0	0.00%	14	0.03%
Airbag Not Ejected	39555	85.82%	6511	14.13%	22	0.05%	46088	99.76%
Others	78	81.25%	17	17.71%	1	1.04%	96	0.21%
Vehicle Position								
In Front	17197	79.37%	4460	20.58%	10	0.05%	21667	46.90%
Behind	20527	92.32%	1699	7.64%	8	0.04%	22234	48.13%
Others	1917	83.46%	375	16.33%	5	0.22%	2297	4.97%

3.2. Methodology

3.2.1. MNL model

In the present study, the mixed MNL models are constructed to quantify the effect of impact factors on driver's injury severity. A hybrid MNL model is developed as follows. The Utility function is defined as

$$U_{ij} = \alpha_j + \sum_k \beta_{jk} X_{ik} + \epsilon_{ij} \tag{3-1}$$

Where X_{ik} is the *k*th explanatory variable for alternative *j* and driver *i*, β_{ji} is the coefficients of the explanatory variables to be estimated, and ϵ_{ij} is the error term which includes the random information. Notice that generally, X_{ik} can be dependent on choice *j* but not in this study. The error term is assumed to satisfy the Gumbel and type I extreme value distribution (Gumbel, 1958), and its probability density function is

$$g(\epsilon_{ij}) = e^{-\epsilon_{ij}} e^{-e^{-\epsilon_{ij}}}. \tag{3-2}$$

In our case, the probability of alternative *j* for individual *i* is the severity level *j* of *i*th driver, where *j*=1,2,3 corresponds to no injury (N), injury (I), and fatality (F), respectively. This probability is determined by the Utility Maximization Principle (Train, 2009), i.e.

$$P_{ij} = P(U_{ij} > U_{ik}), \ \forall k \neq j. \tag{3-3}$$

For the basic MNL model, the coefficients α_j, β_{jk} is assumed constant, and the probability is solved as

$$P_{ij} = \frac{e^{\alpha_j + \sum_k \beta_{jk} X_{ik}}}{\sum_j e^{\alpha_j + \sum_k \beta_{jk} X_{ik}}}. \tag{3-4}$$

This paper adopts the mixed MNL model, taking that the coefficients $\alpha_j,\ \beta_{jk}$ are random parameters instead of describing the driver's unobservable heterogeneities. Therefore, the possibility becomes

$$P_{ij} = \int \frac{e^{\alpha_j + \sum_k \beta_{jk} X_{ik}}}{\sum_j e^{\alpha_j + \sum_k \beta_{jk} X_{ik}}} f(\alpha_{ij}) \prod_k f(\beta_{jk})\, dV, \tag{3-5}$$

where $f(\beta_{jk})$ is the density function of β_{jk}.The coefficients are $\alpha_j = a_j + b_j\gamma_j$, $\beta_{jk} = a_{jk} + b_{jk}\gamma_{jk}$, and the integral is taken over the domain of $\alpha_j,\ \beta_{jk}$. In (5) a_j, b_j, a_{jk}, b_{jk} are constant, and γ_j, γ_{jk} are random parameters with the density function chosen from four different distributions: Normal distribution, Uniform distribution, Triangular distribution, and Exponential distribution. This paper aims to determine the best distribution concerning the lowest Akaike information criterion (AIC) value (Akira, 1974). The coefficients α_j, β_{jk} are estimated by using the maximum likelihood method. In this paper, the mixed MNL model is assessed utilizing NLOGIT (v5.0) (Hensher et al., 2005), and simulation with 1000 Halton draws has been adopted, which is verified to provide a tradeoff between model goodness-of-fit and computing efficiency (Bhat, 2003; Train, 2009; Yu et al., 2019).

3.2.2. Elasticity analysis

As is noted that the estimated coefficients of models can only interpret the risk factors in a qualitative manner (Greene, 2012; Kim et al., 2010). Accordingly, elasticity analysis as a posterior estimation is used to evaluate the sensitivity of the probability change concerning the change of the risk variables discovered in the mixed MNL models on driver's injury severity. The elasticity is defined to be (Washington et al., 2010)

$$E_{X_{ijk}}^{P_{ij}} = \frac{\partial \log P_{ij}}{\partial \log X_{ijk}} = \frac{X_{ijk}}{P_{ij}} \frac{\partial P_{ij}}{\partial X_{ijk}}. \tag{3-6}$$

For a basic MNL model, by manipulating (4), it becomes

$$E_{X_{ik}}^{P_{ij}} = X_{ik} \sum_m (\delta_{jm} - P_{ij}) \beta_{mk}. \tag{3-7}$$

where δ_{jm} is the Kronecker delta, and the summation is since the value of the observatory variables, in this case, are driver-specific, i.e. X_{ik} does not depend on severity level *j*. For the mixed MNL model (3-5), the integration is needed

$$E_{X_{ik}}^{P_{ij}} = \int X_{ik} \sum_m (\delta_{jm} - P_{ij}) \beta_{mk} f(\alpha_{ij}) \prod_k f(\beta_{jk})\ dV. \tag{3-8}$$

3.3. Result

The model estimation is conducted using NLOGIT 5 software. There are, in total, five models constructed concerning the PP crashes, TT crashes, PT crashes, TP crashes and overall crashes. Generally, a lower AIC value indicates a better model fit on the studied dataset. The elasticity result is also obtained for each model.

3.3.1. **Mixed MNL models**

The results for the five models are shown in Table 3-2, respectively. The observable variables with a significance level of more than 10% are removed from each list due to their impacts at the level of *p=0.10*. The random variables' distributions are found by trial and error for all parameters. The candidate distributions are the normal, lognormal, triangular, and uniform distributions, and the final selection is determined by the lowest AIC value, except for the PT case. In this case, the Normal distribution is finally chosen rather than the triangular distribution because its AIC value is slightly higher, and the Normal distribution is more commonly used. The random parameters indicating heterogenicity are summarized in Table 3-3. For PP crashes, the random parameter is only for Licensed (I) with the triangular distribution; for TT crashes, the teenage drivers show heterogenicity in their driver skills, the coefficient of Age under 24 follows the Exponential distribution; in the PT crashes, the coefficients of Age under 24 (I) and Daylight (I) are normally distributed; for the TP crashes, the intercept and the coefficients of Winter (I), Male (I), Surface Wet (I) are distributed uniformly; for the overall dataset, there are five random coefficients for intercept, Surface wet(I), Straight but not level(I) and Airbag not ejected (I) (F). This shows that general crash variables, environment-specific, driver-specific, and vehicle-specific variables may be heterogeneous but different configurations. Moreover, key factors also differ across these five models. For example, as shown in Table 3-2, the parameter Age above 65 is significant for fatality (F) in PP crashes, while in the PT, TP and overall impacts, it is a critical factor for injury (I). In TT crashes, it is not essential. This is under the result of (Abdel-Aty & Abdelwahab, 2003), which conclude that the senior drivers tend to be a key factor for the PP and PT crashes. Daylight makes it safer for all the cases except for the TT crashes. Being the front vehicle is a risk factor of fatality (F) only in PT crashes. Surface conditions influence TP crashes the most. There are also similar critical factors for all the models: alcohol, whether impaired or not impaired, increases the risk of injury and fatality; male drivers reduce the probability of injury in all cases; straight but not level contributes to unsafe driving; the vehicle in front directly relates to injury (I). The coefficients can only provide us with the qualitative influence of the risk factors. Comparisons will be made using the elasticity results to estimate the sensitivity of the probability change concerning the variable evolution.

Table 3-2 Estimation Result of Mixed MNL Model for five Models

	PP			TT			PT			TP			Overall		
Variable	**Coef.**	**Std.**	**p**	**Coef.**	**Std.**	**p**	**Coef.**	**Std.**	**p**	**Coef.**	**Std.**	**p**	**Coef.**	**Std.**	**p**
Constant(P)	2.41***	0.29	0.00	3.71***	0.11	0.00	2.78***	0.16	1.00	11.33***	2.27	0.00	11.33***	2.27	0.00
Std. Dev										8.77***	2.44	0.00	1.52***	0.26	0.00
General variables															
Weekday（I）				0.82***	0.10	0.00									
Weekday（F）										-2.12***	0.74	0.00	-0.91**	0.42	0.03
Fall（F）													-1.25**	-2.16	0.03
Winter(I)	-0.39***	0.14	0.01							-2.72*	1.64	0.10	-0.23***	0.07	0.00
Std. Dev										9.87**	4.04	0.01			
Winter(F)													-1.55**	0.66	0.02
Environment-related variables															
Clear(I)				0.33***	0.08	0.00									
Surface wet(I)										-1.92***	0.87	0.03	-0.20***	0.06	0.00
Std. Dev										7.60***	2.68	0.00	7.60***	2.68	0.00
Daylight(I)	-0.39***	0.14	0.01	0.45***	0.09	0.00	-0.52***	0.20	0.01	-0.14	0.26	0.58	-0.35***	0.08	0.00
Std. Dev							1.26***	0.35	0.00						
Daylight(F)													-1.28***	0.44	0.00
Dawn and dust(I)							-0.30*	0.18	0.10				-0.48***	0.13	0.00
Dark without light (I)	0.48**	0.23	0.04	0.72***	0.18	0.00				1.36**	0.58	0.02	0.31***	0.13	0.01
Curve and level(I)				0.37**	0.18	0.04									
Straight but not level(I)	0.28**	0.12	0.02	0.34***	0.08	0.00	0.19**	0.09	0.03	0.85***	0.28	0.00	0.24*	0.13	0.06
Std. Dev													1.43*	0.83	0.09

	PP			TT			PT			TP			Overall		
Variable	**Coef.**	**Std.**	**p**	**Coef.**	**Std.**	**p**	**Coef.**	**Std.**	**p**	**Coef.**	**Std.**	**p**	**Coef.**	**Std.**	**p**
Curve and no level(I)													0.20*	0 .12	0.09
Curve and no level(F)							2.77***	0.91	0.00				1.26**	0.63	0.05
Driver-specific variables															
Male(I)	-0.96***	0.20	0.00	-.43***	0.07	0.00	-0.56***	0.09	0.00	-3.78***	1.27	0.00	-0.94***	0.12	0.00
Std. Dev										8.13***	2.71	0.00			
Male(F)				-4.71***	1.00	0.00							-1.12***	0.43	0.01
Age under 24(I)	-0.64***	0.16	0.00	-1.68**	0.76	0.03	-0.64*	0.34	0.06	-0.95***	0.29	0.00	-0.42***	0.07	0.00
Std. Dev				2.22***	0.69	0.00	1.46***	0.52	0.01						
Age above 65(I)							0.38**	0.16	0.02	1.38***	0.52	0.01	0.23**	0.11	0.03
Age above 65(F)	1.61**	0.80	0.04												
Belt usage															
No(I)	1.23*	0.66	0.06												
No(F)													2.33**	1.08	0.03
Sobriety Level															
HBD not impaired (I)	1.98**	0.95	0.04	2.25***	0.56	0.00	1.70**	0.71	0.02				1.56***	0.52	0.00
HBD impaired (I)	1.49***	0.49	0.00							4.20**	1.63	0.01	0.92***	0.24	0.00
HBD impaired (F)	2.21**	1.06	0.04												
Licensed(I)	-2.04**	0.80	0.01												
Std. Dev	8.69***	2.08	0.00												
Licensed(F)				-2.60**	1.23	0.03									
Vehicle-specific variables															
Airbag not ejected(I)										4.57***	0.69	0.00	-2.47***	0.68	0.00
Std. Dev													5.20***	1.10	0.00
Airbag not ejected(F)	-5.10***	0.47	0.00				-4.36***	1.31	0.00				-6.40***	1.00	0.00
Std. Dev													5.20***	1.10	0.00
Vehicle Position															
In front(I)	2.06***	0.40	0.00	1.20***	0.07	0.00	2.14***	0.17	0.00	2.09***	0.54	0.00	1.65***	0.19	0.00
In front(F)							2.76**	1.34	0.04						

Coef. Stands for coefficient.

Std. Stands for standard error.

Std. Dev stands for standard deviation.

p. stands for p-value.

***, **, * are significance at 1%, 5%, 10% level.

Table 3-3 Random Parameters for All Cases

Crash type	**Random Distribution**	**Significant Random Parameters**
PP	Triangular	Licensed(I)
TT	Exponential	Age under 24(I)
PT	Normal	Age under 24(I)
		Daylight(I)
TP	Uniform	Constant
		Winter(I)
		Male(I)
		Surface wet(I)
Overall	Uniform	Constant

Crash type	Random Distribution	Significant Random Parameters
		Surface wet(I) Straight but not level(I) Airbag not ejected (I) (F)

3.3.2. Elasticity results

Table 3-4 presents the results of the elasticity estimation averaged overall drivers for 2V crashes for the four crash configurations PP, TT, PT, TP and the comprehensive dataset. It is a posterior estimation based on the mixed MNL model estimated. The variables that significantly relate to severity level with a large elasticity are discussed.

Effects of Weekday

For the TT crashes, the weekday contributes negatively and increases the injury (I) probability, and in TP crashes, the fatality is decreased on a weekday. The increase of TT crashes might result from the business role of pickup trucks on weekdays.

Effects of Season

Winter is a crucial parameter in PP, TP, and overall crashes. However, it only increases the probability of injury for TP crashes, while it appears to be a positive factor for PP crashes and the comprehensive data, especially for fatality prevention. The negative effect may result from the bad weather in winter, while the positive impact can be because of the driver’s cautiousness in a freezing weather condition. Winter also shows heterogeneity in TP crashes, which might be influenced partially by the surface condition in winter discussed afterward.

Effects of Surface condition

Surface wetness also shows a distinct effect in TP crashes (negative) and overall data (positive), which, like the condition of winter, can result from the unfavorable driving condition and the extra carefulness of the drivers, respectively. Therefore, the driver should pay extra attention under bad surface conditions when a passenger car follows a pickup truck (or another type of vehicle with a larger mass).

Effects of Light conditions

Daylight provides visibility for safe driving and may cause the drivers to drive at high speed and cause crash injury. Daylight is shown as an impact factor in all the cases, but it does not always function the same. For PP crashes, it reduces the possibility of injury with an elasticity 0.11, while for the TT crashes and PT crashes, it significantly increases the possibility of damage. Although in TP crashes, the wound is reduced, the fatality is raised. It turns out that the pickup truck involved in crashes are inclined to get drivers injured while recalling that in the PT crash, the coefficient of Daylight is normally distributed with a mean of -0.52 and standard deviation of 1.26 as shown in Table 3-3, which also underlies the heterogeneity of the daylight in PT crashes. Generally, in the overall data model, daylight tends to decrease the possibility of both injury and fatality. In three of the four configurations, the dark without light significantly increases the likelihood of injury and is believed to influence the driver's judgment directly. This result emphasizes that the lighting condition at night is crucial for safe driving. One high-impact research area is to push for vision-enhancing and night-vision devices (Khattak, 2001).

Effects of Roadway characteristics

Roadway characteristics also contribute significantly to injury in a crash. Among all types of the roadway, the straight but not level increases the possibility of damage for PP, TT, PT, TP and overall crashes with elasticities 0.0292, 0.0682, 0.0533, 0.0532, 0.0556, respectively. On the other hand, curve and not level are shown significantly increasing the possibility of fatality for the PT crashes and the overall crashes, with the elasticity of 0.1099 and 0.0529, respectively. This indicates that the slope cause injury in most hits, and the fatality is increased if the road is not level and even curved, especially in the PT crashes. The result warns that a road sign indicating slope is essential on the road with hill, sag, and grade. The negative effect of slope and curve can also be found in other works (Li et al., 2019; Yu et al., 2020).

Effects of driver specific variables

Among all the driver-specific variables, the driver's gender, age, sobriety, and license significantly impact the driver injury severity (Ulfarsson & Mannering, 2004; Haleem & Gan, 2013; Wu et al., 2014). In general, the male driver has a positive action in avoiding injury and fatality coinciding with other works (Chen & Chen, 2001; Wu et al., 2014; Li et al., 2019). In the PP, TT and PT crashes, the male driver contributes positively, especially in the TT crashes male driver reduces the possibility of fatality with an elasticity -2.9484. In TP crashes, the elasticity result shows male drivers tend to increase the probability of fatality, which may result from the fact that the Male coefficient has firm heterogeneity, as shown in Table 2. Driver's age is also a significantly related factor to the injury severity. In the PP crash, the Age above 65 raises the possibility of fatality with an elasticity of 0.0818. In TT crashes, the Age above 65 is not a significant factor. In both the PT and TP crash, the age above 65 is related to injury (I), but in the TP crash, the related fatal possibility reversely decreases with an elasticity of 0.021. This may be explained by the fact that the next car has minor damage in a crash. In TT and PT crashes, the probability of injury is increased in younger groups of age under 24. In these two situations, the coefficients of Age under 24 have an exponential and normal distribution, respectively. This implies variation in the driving skill and expertise among individual young people, especially with the pickup truck. Another crucial factor is the Sobriety level. The result shows that if the driver drinks, whether the behavior is impaired or not, the possibility of injury increases, warning of the danger of driving under the influence of alcohol. In addition, whether a driver is licensed or not is a significant factor. In PT crash, the possibility of fatality is decreased with elasticity -2.5664. In PP crashes, the chance of injury increases, but Table 3-2 also shows that the parameter of licensed is random with a mean of -2.04 and a standard deviation of 8.69, which is consistent with the empirical recognition of the heterogeneity of driving behavior among authorized individuals, especially for passenger car drivers.

Effects of Airbag not ejected

The elasticity result shows that Airbag has not ejected decreases the probability of fatality in PP, PT, TP, and overall crashes. Only for the TT crashes, it is not a key factor. It is also essential to observe from table 3-2 that the coefficient of Airbag not ejected in overall data shows to be normally distributed. As shown, for injury (I), the mean is -2.47, and the standard derivation is 5.20, which indicates a 68.3% possibility for it to be negative. For the fatality (F), the mean is -6.40, and the standard variation is 5.20.

Thus, the coefficient is negative almost 90% of the time. These facts demonstrate the protection of airbags.

Effects of Crash position

When two vehicles crash, the dependence of the driver's injury severity on the vehicle's crash position relative to the other car is of interest for a 2V crash examination (Yasmin 2014). The result shows that, in general, the vehicle in front significantly relates to the injury in all configurations. An interesting phenomenon is that in the PT crash, the passenger car's possibility of fatality (F) is raised with an elasticity of 1.1398. In contrast, in the TP crash, the pickup truck's possibility of fatality (F) is reduced with an elasticity -0.2635. This indicates that, in general, the passenger car has more disadvantages than the pickup truck in terms of driver safety, and extra caution should be paid when a passenger car is followed by a car.

Table 3-4 Elasticity Estimation

	PP			TT			PT			TP			All		
Variables	**(N)**	**(I)**	**(F)**	**(N)**	**(I)**	**(F)**	**(N)**	**(I)**	**(F)**	**(N)**	**(I)**	**(F)**	**(N)**	**(I)**	**(F)**
Weekday				-9.14	53.54	-9.14				0.05	0.06	-174.26			
Winter	0.60	-3.99	0.60							-1.03	8.34	1.52	0.41	-2.90	-34.9
Clear				-2.78	15.95	-2.78									
Surface wet										-0.70	6.50	3.71	0.45	-3.24	0.45
Daylight	1.74	-11.28	1.74	-4.48	26.77	-4.48	-0.96	20.01	-0.96	0.40	-2.89	2.65	2.04	-13.92	-90.74
Dawn and Dust							0.19	-1.23	0.19				0.17	-1.38	0.17
Dark without light	-0.22	1.02	-0.22	-0.47	2.46	-0.47				-0.23	1.17	-1.11	-0.14	0.73	-0.14
Curve and Level				-0.19	1.00	-0.19									
Straight but not level	-0.53	2.92	-0.53	-1.29	6.82	-1.29	-0.65	3.33	-0.65	-0.90	5.32	-5.49	-0.82	5.56	-0.82
Curve and not level							-0.07	-0.06	10.99				-0.08	0.45	5.29
Male	3.03	-23.41	3.03	2.84	-21.95	-294.84	3.41	-26.46	3.41	0.06	2.21	11.54	3.81	-32.87	-60.74
Belt not use	-0.06	0.19	-0.06										-0.01	0.00	0.8
HBD impaired	-0.18	0.79	2.90							-0.14	0.43	-0.77	-0.08	0.49	-0.08
HBD not impaired	-0.06	0.19	-0.06	-0.17	0.29	-0.17	-0.08	0.25	-0.08				-0.04	0.14	-0.04
Licensed	-8.56	76.48	-8.56				0.18	0.14	-256.64						
Age under 24	1.00	-8.84	1.00	-0.46	6.32	-0.46	-0.38	6.71	-0.38	0.70	-7.48	4.54	0.64	-6.24	0.64
Age above 65	-0.02	-0.01	8.18				-0.37	1.71	-0.37	-0.36	1.62	-2.10	-0.12	0.66	-0.12
Airbag not ejected	0.34	0.15	-508.22				0.30	0.23	-434.49	-16.79	118.25	-103.94	-4.51	49.49	-223.17
In front	-8.53	30.74	-853	-10.52	41.04	-10.52	-22.88	64.20	113.98	-4.37	21.76	-26.35	-8.17	32.20	-8.17

All of the values are real values multiplied by 100. (N), (I), and (F) stand for no injury, injury, and fatality, respectively.

3.4. Summary

We studied the rear-end 2V crashes of passenger cars and pickup trucks on divided two-way roads in the State of Washington for seven years, from 2010 to 2016. The risk factors for the severity in four configurations (PP, TT, PT and TP crashes) are examined. Four mixed MNL models for each crash configuration and one model for the overall data are constructed. The climate season-specific (Winter), environment-specific (Daylight, Surface wet, straight but not level), driver-specific (Male, Age under 24), and vehicle-specific (In front, Airbag not ejected) variables show heterogeneity on the injury but only in specific groups. Each model's elasticity analysis is conducted to determine the sensitivity of the possibility of severity to the change if the key factors are estimated in these five mixed MNL models. The similarities across all the models include drinking alcohol (whether impaired or not impaired) raises the risk of injury and even fatality in all cases; male drivers reduce the probability of injury in all circumstances; straight but not level contributes to unsafe driving; the vehicle in front significantly relates to injury; effects of airbag not ejecting and dark without light are related to the severity of injury. Besides similarity, each configuration has specific characteristics. For example, daylight driving is safer for all the cases except the TT crashes; In front and Curve and no level are impact factors of fatality only in PT crashes. Age above 65 is a risk factor of fatality only for PP crashes; surface condition influences TP crashes the most. The differences in the key elements and those with heterogeneity imply that specific strategies should be adopted in each configuration accordingly. It is worth digging deeper into each crash configuration to understand their risk factors better. More research efforts could be made, including studies of more potential risk factors such as the vehicle velocities, passengers in the vehicle, and so on. Time and space heterogeneity could also be a further consideration.

CHAPTER 4. A LATENT CLASS APPROACH FOR DRIVER INJURY SEVERITY ANALYSIS

Temporal instability has been recognized as one of the major sources of unobserved heterogeneity in traffic safety research that has not been completely addressed. Overlooking temporal instability may result to biased estimates of effects of impact factors. In this chapter, we develop a latent class mixed logit model with temporal indicators to investigate highway single-vehicle crashes and the effects of significant contributing factors to driver injury severity. Crash data from 2010 to 2016 in the State Washington are collected with a total of 31115 single-vehicle crashes. The developed model is able to interpret both within- and across- class unobserved heterogeneity and temporal instability. After a careful comparison, a two-class model is selected as the final model. Estimation results show that: two temporal indicators show significant influence on latent class membership; urban indicator and principal type are found to be random parameters and have significant heterogeneity in the mean as a function of male indicator and driver's age indicator. Variables with fixed effects, including animal collision, overturn collision, off-road collision, winter, minor arterial, interstate, wet, snow, ice, speed limit [5, 30), vehicle age [8, 12), [16, 70), turning movement, out control movement, lane-change movement, no airbag, deployed airbag, partial and totally ejection, seatbelt, and no liability, show significant influences on different levels of injury severity in each class. This study provided an insightful understanding of the time-varying effects of the significant factors on driver injury severity using pseudo elasticity analysis. The rest of the chapter is organized as follows: Section 4.1 provides the explicit description of the dataset. The details of model development are illustrated in Section 4.2. In Section 4.3, the model analysis results are comprehensively presented and discussed regarding the implication of the proposed model and the impacts of different risk factors. Finally, the entire research effort is concluded in Section 4.4.

4.1. Data

This study was conducted based on a highway single vehicle crash dataset extracted from traffic crash records acquired from the Washington Department of Transportation (WSDOT). The dataset was drawn from a 7-year period from 2010 to 2016. The dataset can be classified into four categories: (a) general crash information, (b) environmental information, (c) vehicle information, and (d) driver and passenger information. More specifically, general crash information includes the crash severity in terms of five accident-severity categories (i.e., no injury, possible injury, evident injury, serious injury and fatality), collision type, temporal information, and county name. Environmental information involved information regarding weather, surface condition, lighting condition, speed limits, roadway characteristics, and indicators for work zone. Vehicle information contains the vehicle type, vehicle age, airbag condition, and ejection status. Driver and passenger information included drivers' age, gender, seat belt usage, license status, insurance, and passengers restrain and sobriety conditions.

The present study focused on single-vehicle crashes. Effort was made to pre-process the selected datasets. Variables that have similar definitions or similar impacts on driver injury outcomes were carefully examined and combined, which is consistent with the existing literature in roadway safety analysis (Gong and Fan, 2017; Li et al., 2018b; Chen et al., 2016b). For instance, three variables, including the roadway type, the state function class, and the federal function class, were all related to the road segment categories. The three variables were fused into one variable, and the new variable classified the roadway segments into four categories: interstate, principal arterial, minor arterial, and collector. Records that were fragmentary and erroneous, such as records with 'unknown' information, were removed from the dataset. Some continuous variables, such as driver' age and vehicle's age, were categorized based on previous traffic safety research experience (Chen et al., 2016a, 2016b; Li et al., 2018a, 2019b; Wu et al., 2016b). Finally, a total of 31115 single vehicle crash records were extracted, involving 327 fatality crashes, 1,121 serious injury crashes, 4,891 evident injury crashes, 6,606 possible injury crashes and 18,170 no injury crashes. In order to maintain a meaningful sample size, fatality crashes and serious injury crashes were merged into one injury severity level, i.e., serious and fatality injury (S/F). Consequently, four levels of injury severity were considered, including no injury (N), possible injury (P), evident injury (E), and serious and fatality injury (S/F). Similar simplified classification methods have also been applied in Lee and Mannering (2002), Gong and Fan (2017), and Li et al. (2018b). The descriptive statistics of the dataset were illustrated in Table 4-1.

Table 4-1. Variable Definitions and Descriptive Statistics.

Variable	Value	Driver Injury Severity								Total
		N	%	P	%	E	%	S/F	%	
GENERAL										
Collision	Fixed	13996	78.3%	2205	12.3%	1282	7.2%	400	2.2%	17883
	Animal	5765	94.8%	185	3.0%	108	1.8%	21	0.3%	6079
	Overturn	1937	57.0%	609	17.9%	697	20.5%	155	4.6%	3398
	Off-road	2793	74.4%	486	12.9%	387	10.3%	89	2.4%	3755
Season	Spring	4743	76.5%	738	11.9%	557	9.0%	162	2.6%	6200
	Summer	6060	76.7%	869	11.0%	741	9.4%	228	2.9%	7898
	Fall	6611	79.3%	941	11.3%	611	7.3%	170	2.0%	8333
	Winter	7077	81.5%	937	10.8%	565	6.5%	105	1.2%	8684
ENVIRONMENT										
Route Type	Urban[a]	10556	78.9%	1670	12.5%	907	6.8%	253	1.9%	13386
	Rural	13935	78.6%	1815	10.2%	1567	8.8%	412	2.3%	17729
Function	Collector	14054	78.6%	1827	10.2%	1577	8.8%	418	2.3%	17876
	Minor arterial	2420	77.9%	371	11.9%	225	7.2%	90	2.9%	3106
	Principle	2692	79.3%	428	12.6%	226	6.7%	50	1.5%	3396
	Interstate	5325	79.0%	859	12.8%	446	6.6%	107	1.6%	6737
Surface	Dry	13509	77.4%	1881	10.8%	1570	9.0%	484	2.8%	17444
	Wet	6507	79.4%	993	12.1%	560	6.8%	137	1.7%	8197
	Snow	2560	84.8%	288	9.5%	155	5.1%	17	0.6%	3020
	Ice	1915	78.0%	323	13.2%	189	7.7%	27	1.1%	2454
Lighting	Daylight	11886	77.4%	1812	11.8%	1350	8.8%	312	2.0%	15360
	Twilight	1392	77.9%	202	11.3%	159	8.9%	33	1.8%	1786

Variable	Value	Driver Injury Severity								Total
		N	%	P	%	E	%	S/F	%	
	Dark with light	3889	77.8%	644	12.9%	350	7.0%	117	2.3%	5000
	Dark	7324	81.7%	827	9.2%	615	6.9%	203	2.3%	8969
Speed Limit	[5,30)	2912	80.1%	425	11.7%	240	6.6%	60	1.6%	3637
	[30,60)	7708	77.6%	1073	10.8%	867	8.7%	280	2.8%	9928
	[60,80)	13871	79.0%	1987	11.3%	1367	7.8%	325	1.9%	17550
Road Feature	Level	14043	78.6%	2011	11.3%	1421	8.0%	387	2.2%	17862
	Grade	9757	78.7%	1392	11.2%	984	7.9%	259	2.1%	12392
	Hill and sag	691	80.3%	82	9.5%	69	8.0%	19	2.2%	861
Work zone	No	24153	78.7%	3438	11.2%	2449	8.0%	655	2.1%	30695
	Yes	338	80.5%	47	11.2%	25	6.0%	10	2.4%	420
VEHICLE										
Type	Passenger car	12068	78.7%	1820	11.9%	1139	7.4%	305	2.0%	15332
	Pickup	11648	79.9%	1515	10.4%	1120	7.7%	295	2.0%	14578
	Truck	351	80.9%	46	10.6%	30	6.9%	7	1.6%	434
	Other	424	55.0%	104	13.5%	185	24.0%	58	7.5%	771
Age	[0,4)	3234	84.7%	330	8.6%	206	5.4%	47	1.2%	3817
	[4,8)	4799	82.4%	582	10.0%	352	6.0%	92	1.6%	5825
	[8,12)	6398	79.6%	837	10.4%	653	8.1%	149	1.9%	8037
	[12,16)	5193	76.9%	812	12.0%	605	9.0%	144	2.1%	6754
	[16,70)	4867	72.8%	924	13.8%	658	9.8%	233	3.5%	6682
Movement	Moving	21907	79.0%	3024	10.9%	2203	7.9%	594	2.1%	27728
	Turning	1082	80.7%	153	11.4%	82	6.1%	23	1.7%	1340
	Parking	166	62.9%	43	16.3%	34	12.9%	21	8.0%	264
	Backing	23	92.0%	0	0.0%	1	4.0%	1	4.0%	25
	Merging	57	83.8%	6	8.8%	5	7.4%	0	0.0%	68
	Out control	506	80.1%	72	11.4%	48	7.6%	6	0.9%	632
	Lane-change	750	70.9%	187	17.7%	101	9.5%	20	1.9%	1058
Airbag	No airbag	6039	77.9%	803	10.4%	661	8.5%	246	3.2%	7749
	Not deployed	15243	84.5%	1569	8.7%	1046	5.8%	173	1.0%	18031

Variable	Value	Driver Injury Severity								Total
		N	%	P	%	E	%	S/F	%	
	Deployed	3209	60.1%	1113	20.9%	767	14.4%	246	4.6%	5335
Ejection	No	24092	80.1%	3355	11.1%	2220	7.4%	424	1.4%	30091
	Partial	385	53.7%	90	12.6%	161	22.5%	81	11.3%	717
	Totally	14	4.6%	40	13.0%	93	30.3%	160	52.1%	307
DRIVER										
Gender	Female	11784	76.7%	1970	12.8%	1275	8.3%	327	2.1%	15356
	Male	12707	80.6%	1515	9.6%	1199	7.6%	338	2.1%	15759
Age	(0,24]	14794	80.6%	1893	10.3%	1356	7.4%	321	1.7%	18364
	(25,45]	5430	76.3%	887	12.5%	617	8.7%	180	2.5%	7114
	(45,65]	3059	75.4%	503	12.4%	371	9.1%	122	3.0%	4055
	Above 65	1208	76.4%	202	12.8%	130	8.2%	42	2.7%	1582
Belt	Unused	1631	54.4%	458	15.3%	554	18.5%	357	11.9%	3000
	Used	19899	80.2%	2818	11.4%	1802	7.3%	285	1.1%	24804
	Child seat[b]	2961	89.4%	209	6.3%	118	3.6%	23	0.7%	3311
License	No	270	68.2%	62	15.7%	59	14.9%	5	1.3%	396
	Yes	24221	78.8%	3423	11.1%	2415	7.9%	660	2.1	30719
Liability	No	3782	67.6%	855	15.3%	691	12.4%	265	4.7%	5593
	Yes	20709	81.1%	2630	10.3%	1783	7.0%	400	1.6%	25522
Sobriety	Other	946	67.9%	183	13.1%	170	12.2%	94	6.7%	1393
	No drink	22017	80.9%	2886	10.6%	1949	7.2%	362	1.3%	27214
	Not impaired[c]	300	69.9%	69	16.1%	50	11.7%	10	2.3%	429
	Impaired	1228	59.1%	347	16.7%	305	14.7%	199	9.6%	2079

[a]indicates the highway section located in urban area
[b]indicates both seat belt and child seat are used
[c]driving ability not impaired based on tox test

4.2. Methodology

4.2.1. Model development

The single-vehicle crash injury severities were generally investigated using a latent class mixed logit model with temporal indicators. To begin, a finite number *Q* is pre-determined to group the highway single vehicle crash dataset into *Q* classes. Note that there is no strict rule on the selection of *Q* value (Xie et al., 2012). In this study, proposed models with different *Q* values were compared. Optimal class number is yielded in terms of the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), in addition to the log-likelihoods and McFadden's pseudo R^2 at convergence (McFadden et al., 1973; Yamaoka et al., 1978; Joo et al., 2010). It is not necessary to assign the crash records into any specific class. The classification will be done during model fitting which maximizes the heterogeneity among different classes. The function determining the probability, of which the *i*th crash record belongs to latent class q ($q \in Q$) and injury severity level j, is defined as follows

$$U_{ji|q} = \beta_{ji|q}\boldsymbol{x}_{ki} + \varepsilon_{ji|q} \tag{4-1}$$

where x_i is the vector of explanatory variables.

In our notation, the same explanatory vector x_i is used for all the injury types; however, this is not a restriction, since it would be possible to replace appropriate row of $\beta_{j|q}$ with the value of 0 if a specific variable does not have significant impact on *j*th injury severity in class q. $\varepsilon_{ji|q}$ is the unobserved heterogeneity for *i*th driver with *j*th injury severity in class q, which is designed as an independent and identically distributed random variable. $\beta_{j|q}$ accounts for class-specified impacts of variable x_i on *j*th injury severity. Accordingly, conditional on class $\hat{q}$, the probability of the *i*th crash getting involved in $\hat{j}$ the injury severity is then given by

$$f_{\hat{j}i|\hat{q}}(\beta) = \frac{\exp\left(\beta_{\hat{j}|\hat{q}}^{T} x_i\right)}{\sum_{j=1}^{J} \exp\left(\beta_{j|\hat{q}}^{T} x_i\right)} \tag{4-2}$$

The class membership, e.g., the probability for *i*th crash belonging to class q, is given as follows (Li et al., 2018b; Gong and Fan, 2017; Shaheed and Gkritza, 2014; Xie et al., 2012)

$$g_{i|\hat{q}}(\theta) = \frac{\exp\left(\theta_{\hat{q}}^{T} z_i\right)}{\sum_{q=1}^{Q} \exp\left(\theta_q^{T} z_i\right)} \tag{4-3}$$

where ϑ_q is a vector of class-specific parameters and z_i is a vector of temporal indicators. Note that ϑ_Q is set to zero as reference. In Eq. (4-3), z_i demonstrates the interaction among temporal instability and various characteristics involved in the model. It is possible that the probabilities $P\left(q_i = \hat{q}\right)$ is determined by a set of fixed constants ϑ_qs if no characteristic z_i is observed. Then the unconditional probability of *i*th crash getting involved in level $\hat{j}$ injury severity is given as follows

$$F_{ji}(\beta,\theta) = \sum_{q=1}^{Q} g_{i|q}(\theta) f_{ji|q}(\beta) \tag{4-4}$$

In addition to the across-class heterogeneity identified by the class-specific parameters, the unobserved heterogeneity within-class is accounted for by letting $\beta_{j|q}$ be a vector of identified parameters that varies across individual crashes, shown as follows.

$$\beta_{j|q} = \beta_q + \delta_{j|q} x_i + \sigma_q \upsilon_{j|q} \tag{4-5}$$

where β_q is the population mean. $\delta_{j|q}$ is a vector of estimable parameters corresponding to the vector of variables x_i, which influence the mean of $\beta_{j|q}$. $\upsilon_{j|q}$ is a randomly distributed term that captures unobserved heterogeneity across crashes. In this study, we simply assume that $\upsilon_{j|q}$ follows a standard normal distribution with a mean of 0 and the standard deviation of 1 for two reasons. Firstly, previous studies have found that density functions, such as, lognormal, gamma, Weibull, and so on, were not statistically superior to the normal distribution (Moore et al., 2011; Shaheed and Gkritza, 2014; Li et al., 2018b); and by differentiating the within and across-class heterogeneity, the required distribution assumption for the random parameters becomes less important since parameters can vary across crashes in a more flexible way. Hence, σ_q is the indicator for random parameters. By substituting Eq. (4-5) into Eq. (4-2), the conditional probability of the ith crash getting involved in level $\hat{j}$ injury severity within class $\hat{q}$ is given by

$$f_{\hat{j}i|\hat{q}}(\beta)=\frac{\exp((\beta_{\hat{q}}+\delta_{\hat{j}|\hat{q}}x_i+\sigma_{\hat{q}}\upsilon_{\hat{j}|\hat{q}})^T x_i)}{\sum_{j=1}^{J}\exp((\beta_{\hat{q}}+\delta_{j|\hat{q}}x_i+\sigma_{\hat{q}}\upsilon_{j|\hat{q}})^T x_i)} \tag{4-6}$$

Hence, we have

$$F_{ji}(\beta,\theta)=\sum_{q=1}^{Q}\frac{\exp\left(\theta_{\hat{q}}^T z_i\right)}{\sum_{q=1}^{Q}\exp\left(\theta_q^T z_i\right)}\times\frac{\exp((\beta_{\hat{q}}+\delta_{\hat{j}|\hat{q}}x_i+\sigma_{\hat{q}}\upsilon_{\hat{j}|\hat{q}})^T x_i)}{\sum_{j=1}^{J}\exp((\beta_{\hat{q}}+\delta_{j|\hat{q}}x_i+\sigma_{\hat{q}}\upsilon_{j|\hat{q}})^T x_i)} \tag{4-7}$$

Model estimation is undertaken using maximum simulated likelihood (MSL) estimation (McFadden and Train, 2000; Train, 2009). In this study, the MSL evaluates the aforementioned parameters in the likelihood expression. The contribution of the ith crash to the total simulated likelihood, i.e., the simulated probability, is:

$$\overline{F}_{ji}(\beta,\theta)=\frac{1}{R}\sum_{r=1}^{R}\sum_{q=1}^{Q}\mathrm{g}_{i|q}(\theta)f_{ji|q}(\beta^r) \tag{4-8}$$

where R is the number of simulated draws; β^r denotes the value of all the random vectors draw from the pre-specified distribution in the rth draw; $\overline{F}_{ji}(\beta,\theta)$ is an unbiased estimator of $F_{ji}(\beta,\theta)$ by construction. Its variance decreases as R increases. Collecting all terms, the simulated log likelihood (LL) is obtained as follows:

$$LL=\sum_{j=1}^{J}\sum_{i=1}^{N}d_{ji}\ln\overline{F}_{ji}(\beta,\theta) \tag{4-9}$$

where N is the total number of crash records; d_{ji} indicates the injury severity of ith crash record.

The number of simulated draws, i.e., R, ranges in different studies (Blackburn and Gaston, 2001; Cappellari and Jenkins, 2003; Seraneeprakarn et al., 2017; Li et al., 2018a). Generally, with the increasing number of parameters to be identified within the model, the required number of random draws which can provide reasonable model convergence performance increases largely. Therefore, Halton draw was applied instead of random draw. Halton draw is able to produce the same level of performance with a much smaller draw counts (Train, 2009; Bhat, 2003). A total of 1000 Halton draws are conducted with the MSL estimation in this study.

4.2.2. Elasticity analysis

The elasticity analysis is conducted to evaluate the impacts of variables on the likelihood of crash severity. The elasticity of a continuous independent variable x_{jik} is given as follows (Washington et al., 2010).

$$E_{x_{jik}}^{F_{ji}} = \frac{\partial F_{ji}}{\partial x_{jik}} \frac{x_{jik}}{F_{ji}} \tag{4-10}$$

where $E_{x_{jik}}^{F_{ji}}$ is the elasticity outcome for ith crash; x_{jik} is the value of kth variable for the ith crash in the identified model of jth injury severity. Note that Eq. (4-10) is not applicable for indicator variables. The pseudo-elasticity, termed as $E_{x_{jik}}^{F_{ji}}$, for measuring the influence of the indicator variables x_{jik} on jth driver injury severity is expressed as follows (Kim et al., 2007).

$$E_{x_{jik}}^{F_{ji}} = \frac{\overline{F}_{ji}\left[\text{given } x_{jik} = 1\right] - \overline{F}_{ji}\left[\text{given } x_{jik} = 0\right]}{\overline{F}_{ji}\left[\text{given } x_{jik} = 0\right]} \tag{4-11}$$

The elasticity in Eq. (4-10) and the pseudo-elasticity in Eq (4-11) are both different for each individual crash i and each alternative j. In order to measure variable influence, the average pseudo-elasticity is calculated based on all the data. In addition, due to the random structures in the impact functions, the estimated distributions were adopted to generate the parameters of the corresponding variables instead of applying the fixed means.

4.3. Model Estimation Results and Discussions

4.3.1. Model estimation

The model estimation is conducted using the NLOGIT 5 software. In order to reduce the model estimation bias caused by the multi collinearity between explanatory variables, the correlation coefficient between each pair of variables is estimated before model estimation. The estimated coefficients are shown in Figure 4.1. If two variables are found significantly correlated, they would be inputted into the model formulation one by one while monitoring the overall model fit, e.g., AIC and BIC, and the significance of the variables. The proposed model is fitted with the variables summarized in Table 4-1. As discussed above, no rigid rule is applicable in determining the optimal latent class number. The comparison results for models with different number of latent classes are presented in Table 4-2.

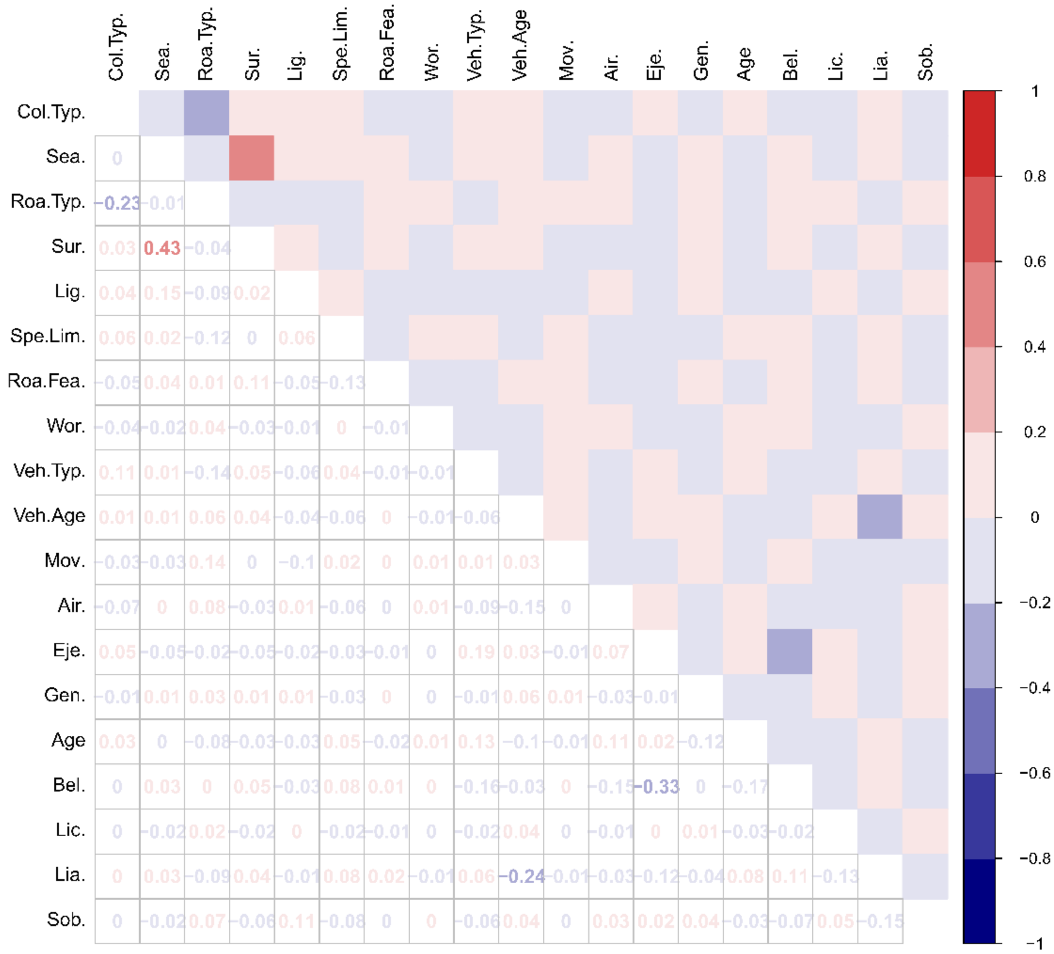

Figure 4.1. Correlation analysis results.

Note that no temporal indicator is found to have significant impacts on latent class membership when the class number is greater than five and the class membership probabilities are determined by a set of fixed constants, which is, however, beyond the scope of the study. As shown in Table 4-2, the AIC and BIC values increase with the class number, indicating that increasing class number should not be recommended in this study. Moreover, when the class number is four, one of the four latent classes has no significant temporal indicators or significant constant term at 95% level of confidence, indicating that the significant difference of the specific class is rejected, while in the five-class model, two of the five latent classes are rejected at 95% level of confidence. Therefore, two-class model is selected as the final model in this study.

Table 4-2. Comparison Results for Models with Different Numbers of Latent Classes

Number of Class	Class #	Prob.[a]	Temporal Indicator[b]	*LL*[c]	AIC	BIC
2	1	0.654*[d]	Year2015, Year2016	-19140	38558	39717
	2	0.346	n/a			
3	1	0.592*	-	-19126	38571	39898

Number of Class	Class #	Prob.[a]	Temporal Indicator[b]	*LL*[c]	AIC	BIC
	2	0.139*	Year2016			
	3	0.268	n/a			
4	1	0.695*	Year2015, Year2016	-19122	38588	40024
	2	0.119	Year2016			
	3	0.095	-			
	4	0.091	n/a			
5	1	0.548*	-	-19114	38631	40317
	2	0.155*	Year2015, Year2016			
	3	0.066	-			
	4	0.144	-			
	5	0.117	n/a			

[a]estimated average class probability
[b]significant indicators at 95% level of confidence influencing the class membership
[c]Log Likelihood value
[d]indicates that the constant term is significant at 95% level of confidence

The detailed estimation results of the two-class model are listed in Table 4-3 and Table 4-4, respectively, including the coefficients, standard errors, and the *p*-values. The entire dataset is classified into two classes, and the average class probabilities are 65.4% and 34.6%, respectively. The McFadden's pseudo R^2 for the final model is 0.556, which means that model shows reasonably performance compared to the intercept-only model. In summary, explanatory variables can be classified into four types, including variables with standard fixed parameter estimates, variables influencing class membership, variables that produces statistically significant random parameters, and variables contributing to heterogeneity in the means. The number of variables falling in the four groups varies across different injury severity levels in the two classes, indicating that the proposed model is appropriate for analyzing the given dataset by capturing both within and across class heterogeneity (Li et al., 2018b). For instance, it is found that off-road collision has a positive sign on possible injury (P) in Class 1, but a negative sign in Class 2. Ice surface condition shows significant effects on evident injury (E) in Class 1 but is rejected in Class 2. Year 2015 and Year 2016 shows significant influence in latent class membership, indicating that the effects of the explanatory variables on injury severity varies significantly in 2015 and 2016. And also, in Table 4-3 and Table 4-4, urban indicator and principal indicator are identified to be random parameters and have significant heterogeneity in the means. More specifically, urban indicator is found to have significant heterogeneity in the mean as a function of age indicator for (45, 65] and above 65, while principal indicator has significant heterogeneity in the mean as a function of male and age indicator for above 65.

Table 4-3. Results of Model Estimation for Class 1

Variable	Value	P			E			S/F		
		Coef.	*s.d.*[a]	*p*[b]	Coef.	*s.d.*	*p*	Coef.	*s.d.*	*p*
Variable influence class membership[c]										
Year	2015	0.219	0.093	0.018						
Year	2016	0.470	0.101	0.000						
Variable with random parameter										
Route Type	Urban	1.196	0.400	0.003						
Function	Principle	2.030	0.441	0.000						
Heterogeneity in mean										
Gender-Male: Function-Principle		-1.704	0.365	0.000						
Age-(45,65]: Route Type-Urban		1.090	0.184	0.000						
Age-above 65: Route Type-Urban		1.377	0.230	0.000						
Age-above 65: Function-Principle		1.451	0.665	0.029						
Variable with standard fixed parameter										
Collision	Animal	-1.580	0.186	0.000	-1.080	0.517	0.037	-2.111	0.959	0.028
Collision	Overturn	0.508	0.150	0.001	2.394	0.291	0.000	0.235	0.134	0.078
Collision	Off-road	0.704	0.133	0.000	0.083	0.053	0.118			
Season	Winter				-0.696	0.334	0.037	-0.484	0.246	0.049
Route Type	Urban				0.111	0.070	0.113	-1.735	0.645	0.007
Function	Minor arterial	1.478	0.421	0.000				-2.443	0.809	0.003
Function	Principle							-2.677	0.822	0.001
Function	Interstate	1.514	0.409	0.000				-2.188	0.722	0.002
Surface	Wet				-0.806	0.299	0.007	0.403	0.170	0.018
Surface	Snow	0.348	0.161	0.031	-1.361	0.778	0.080	-1.580	0.487	0.001

Variable	Value	P			E			S/F		
		Coef.	*s.d.*[a]	*p*[b]	Coef.	*s.d.*	*p*	Coef.	*s.d.*	*p*
Surface	Ice				-0.658	0.452	0.146	-0.698	0.259	0.007
Speed Limit	[5,30)							-0.700	0.369	0.058
Vehicle Age	[8,12)				-0.067	0.022	0.003			
Vehicle Age	[16,70)	0.170	0.063	0.007	0.179	0.078	0.022	0.329	0.105	0.002
Movement	Turning				-0.588	0.248	0.017	-1.750	0.216	0.000
Movement	Out control				-1.174	0.373	0.002			
Movement	Lane change	0.178	0.051	0.001						
Airbag	No airbag	-0.079	0.046	0.084	0.669	0.219	0.002	1.168	0.362	0.001
Airbag	Deployed	1.171	0.112	0.000	0.968	0.246	0.000	0.096	0.049	0.048
Ejection	Partial	-1.993	0.585	0.001				0.173	0.085	0.041
Ejection	Totally	2.523	0.787	0.001	3.106	0.673	0.000	5.990	0.871	0.000
Belt	Unused	0.237	0.062	0.000	0.770	0.243	0.002	2.203	0.392	0.000
Belt	Child seat	-0.223	0.096	0.020	-1.606	0.823	0.051	-1.018	0.534	0.057
Liability	No				0.315	0.185	0.089	0.071	0.045	0.113
Sobriety	Impaired	0.162	0.087	0.064	0.361	0.135	0.008	2.029	0.366	0.000

[a]stands for standard deviation
[b]stands for *p*-value
[c]variables influencing class membership are fixed across different injury severity level

Table 4-4. Results of Model Estimation for Class 2

Variable	Value	P			E			S/F		
		Coef.	*s.d.*[a]	*p*[b]	Coef.	*s.d.*	*p*	Coef.	*s.d.*	*p*
Variable influence class membership[c]										
Year	2015	0.000	(fixed parameter)[c]							

Variable	Value	P			E			S/F		
		Coef.	*s.d.*[a]	*p*[b]	Coef.	*s.d.*	*p*	Coef.	*s.d.*	*p*
Year	2016	0.000	(fixed parameter)[c]							
Variable with random parameter										
Route Type	Urban	0.896	0.233	0.000						
Function	Principle	0.840	0.346	0.015						
Heterogeneity in mean										
Gender-Male: Function-Principle		-0.706	0.343	0.040						
Age-(45,65]: Route Type-Urban		-1.924	0.587	0.001						
Age-above 65: Route Type-Urban		-1.536	0.684	0.247						
Age-above 65: Function-Principle		1.530	0.774	0.048						
Variable with standard fixed parameter										
Collision	Animal	-1.796	0.221	0.000	-2.273	0.259	0.000	-2.455	0.434	0.000
Collision	Overturn	1.302	0.222	0.000	1.117	0.235	0.000	1.508	0.262	0.000
Collision	Off-road	-0.816	0.324	0.012	0.263	0.130	0.043			
Season	Winter				0.045	0.030	0.133	1.508	0.262	0.000
Route Type	Urban				1.111	0.131	0.000	-0.168	0.101	0.094
Function	Minor arterial	0.141	0.096	0.143				-1.109	0.266	0.000
Function	Principle							-1.767	0.322	0.000
Function	Interstate	0.092	0.063	0.144				-1.956	0.346	0.000
Surface	Wet				-0.222	0.110	0.044	-2.034	0.307	0.000
Surface	Snow	-2.799	0.990	0.005	-1.165	0.178	0.000	-0.901	0.249	0.000
Surface	Ice				-0.341	0.162	0.035	-1.793	0.423	0.000
Speed Limit	[5,30)							-0.887	0.382	0.020

Variable	Value	P			E			S/F		
		Coef.	*s.d.*[a]	*p*[b]	Coef.	*s.d.*	*p*	Coef.	*s.d.*	*p*
Vehicle Age	[8,12)				0.243	0.099	0.014			
Vehicle Age	[16,70)	0.616	0.155	0.000	0.551	0.130	0.000	0.408	0.202	0.044
Movement	Turning				-0.709	0.235	0.003	-0.999	0.383	0.009
Movement	Out control				-0.134	0.078	0.084			
Movement	Lane change	0.702	0.259	0.007						
Airbag	No airbag	0.706	0.162	0.000	0.436	0.126	0.001	0.907	0.251	0.000
Airbag	Deployed	1.578	0.209	0.000	1.578	0.157	0.000	2.527	0.281	0.000
Ejection	Partial	0.732	0.329	0.026				1.553	0.322	0.000
Ejection	Totally	2.903	1.278	0.023	3.329	1.217	0.006	4.440	1.339	0.001
Belt	Unused	1.100	0.248	0.000	1.260	0.208	0.000	1.809	0.284	0.000
Belt	Child seat	-1.317	0.300	0.000	-0.884	0.176	0.000	-0.973	0.408	0.017
Liability	No				0.318	0.105	0.003	0.458	0.171	0.007
Sobriety	Impaired	0.742	0.274	0.007	0.842	0.217	0.000	1.345	0.285	0.000

[a]stands for standard deviation
[b]stands for *p*-value
[c]parameters for class memberships are fixed to zero in class 2

4.3.2. Elasticity analyses

It is easy to understand that the sign and the value of an estimated coefficient does not always represent the overall impact effects of variables (Washington et al., 2010). Therefore, the pseudo elasticity estimation, defined in Eq. (4-11), is applied to cope with this issue. Table 4-5 presents the average results of the elasticity analysis. The variables that have significant impacts in estimation results in Tables 4-3 and 4-4 are discussed in depth.

Table 4-5. Results of Pseudo Elasticity Estimation

Variable	Value	Base Value	N	P	E	S/F
Collision	Animal	Fixed	259.51%	-26.91%*[a]	-32.58%*	-59.97%*
	Overturn	Fixed	-53.95%	-4.43%*	147.28%*	8.80%*
	Off-road	Fixed	-16.08%	4.59%*	28.56%*	-9.02%

Variable	Value	Base Value	N	P	E	S/F
Season	Winter	Spring	-0.99%	-0.22%	-29.23%[*]	73.61%[*]
Route Type	Urban	Rural	-44.74%	129.12%	-73.85%[*]	-42.73%[*]
Function	Minor Arterial	Collector	-29.36%	54.24%[*]	-12.94%	-79.65%[*]
	Principle	Collector	-19.54%	54.47%	-12.59%	-88.36%[*]
	Interstate	Collector	-25.64%	66.03%[*]	-4.39%	-87.44%[*]
Surface	Wet	Dry	12.83%	9.69%	-15.00%[*]	-27.23%[*]
	Snow	Dry	71.94%	-0.80%[*]	-24.69%[*]	-27.46%[*]
	Ice	Dry	20.80%	14.90%	-17.15%[*]	-58.79%[*]
Speed Limit	[5,30)	[30,60)	9.89%	6.88%	11.65%	-47.93%[*]
Vehicle Age	[8,12)	[0,4)	-1.50%	-0.75%	10.97%[*]	-2.13%
	[16,70)	[0,4)	-22.23%	5.14%[*]	8.34%[*]	8.20%[*]
Movement	Turning	Moving	27.22%	17.98%	-33.78%[*]	-63.80%[*]
	Out control	Moving	6.19%	5.42%	-31.35%[*]	4.56%
	Lane Change	Moving	-15.97%	13.86%[*]	-20.72%	-16.95%
Airbag	No airbag	Not Deployed	-24.92%	-9.37%[*]	15.63%[*]	102.91%[*]
	Deployed	Not Deployed	-64.63%	22.04%[*]	10.34%[*]	46.89%[*]
Ejection	Partial	No	70.24%	-43.67%[*]	13.15%	147.17%[*]
	Totally	No	-95.85%	-37.30%[*]	12.98%[*]	708.80%[*]
Belt	Unused	Used	-51.99%	-16.41%[*]	20.19%[*]	236.13%[*]
	Child seat	Used	69.41%	-2.51%[*]	-31.38%[*]	-32.11%[*]
Liability	No	Yes	-7.27%	-5.40%	23.27%[*]	17.79%[*]
Sobriety	Impaired	No Drink	-42.66%	-14.82%[*]	3.29%[*]	211.01%[*]
Gen	Male[b]	Female	12.74%	-4.47%	9.35%	10.24%
Age	[45,65)[b]	[24,44)	-6.14%	0.71%	26.94%	0.39%
	Above 65 [b]	[24,44)	-18.94%	10.01%	10.28%	0.45%

[a]variables have significant impacts on injury severity at 95% level of confidence
[b]variables have significant impacts on heterogeneity in means of random parameters

Effects of general factors

In the group of general factors, different collision objects play significant roles contributing to the crash injury severity. As shown in Table 4-5, the indicator "Animal" reduces the probabilities of injury severity, possible injury (P), evident injury (E), and serious and fatality injury (S/F), by 26.91%, 32.58%, and 59.97%, respectively. This finding indicates that hits an animal producing a low severity injury, which is consistent with previous studies (Haghighi et al., 2018). The indicator "Overturn" and "Off-road" represented that the single vehicle crash occurs due to overturning and running off-road. It shows that these two indicators reduce the probability of no injury crash (N), while increase the probability of evident injury (E) by 147.28% and 28.56%, respectively. Moreover, the “Overturn” increases the probability of serious and fatality injury (S/F) by 8.80%. Similar results have also been obtained in Shaheed and Gkritza (2014).

As for the season characteristics, “Summer” and “Fall” are not significant at 95% level of confidence. It is found that, the “Winter” indicator significantly increases the probability of serious and fatality injury (S/F), while reduces the probability of the rests. This result may be caused by two reasons. On one hand, in winter, reduced free flow speed has been found in many previous studies (Hanbali, 1994; Qiu and Nixon, 2009), which produces positive effects on crash severity. on the other hand, the “Winter” indicator may capture unobserved heterogeneity resulting from the cold weather and winter disasters, and it increases the probability of serious and fatality injury (S/F) by 73.61%.

Effects of environmental factors

Environmental indicators including urban, minor arterial, principle, interstate, various road surface conditions, and low speed limits, have significant impacts on injury severities of single vehicle crashes. The results of urban indicators are consistent with previous studies that, compared to rural area, evident injury (E) and serious injury and fatality (S/F) crashes are less likely to occur in an urban area (Xie et al., 2012). Compared to “Collector” segments, the road function indicators, i.e., minor arterial, principle, and interstate, reduce the probability of evident injury (E) by 12.94%, 12.59% and 4.39% and serious and fatality injury (S/F) by 79.65%, 88.36% and 87.44%, respectively. The results reveal the fact that single vehicle crash occurring on collector segments have higher probabilities of evident injury, serious injury and fatality than those occurring on other segments (Gong and Fan, 2017). It is interesting to find that those adverse road surface conditions, such as wet, snow, and ice surface, reduce probabilities of evident injury (E), serious injury and fatality (S/F). These counter-intuitive results might be explained as risk compensation, i.e., driver’s behavior will be more conservative with adverse surface conditions (Mannering and Bhat, 2014). Cautious driving behaviors, such as slowing down and concentration, tend to reduce the probability of being severely or fatally injured in single vehicle crashes, which is supported by Lu et al. (2010) and Gong (2017). Analysis results for low-speed limits shown in Table 4-5 are consistent with the general understanding that increasing speed encourages higher injury severity level.

Effects of vehicle factors

Among the vehicle information collected in the single-vehicle crash dataset, factors, including vehicle age, vehicle movement, airbag condition, and ejection condition, have statistically significant effects on driver injury severity. The estimated elasticity results show that the probability of more serious injury, i.e., evident injury (E) and serious and fatality injury (S/F), grow with increasing vehicle age. For instance, compared to vehicles aged from 0 to 4, vehicles aged older than 16 years have a higher probability of evident injury (E) and serious and fatality injury (S/F) by 8.34% and 8.20%, respectively. As for the vehicle movement condition, it is possible to explain the different injury resulting via the drivers’ concentration in conducting those movements. More specifically, turning and lane-change movements are usually associated with much higher attention than normal moving. As a result, turning movement reduces the probabilities of evident injury (E) and serious and fatality injury (S/F) by 33.78% and 63.80%, while the lane-change movement reduces the probabilities of the two levels of injury severity by 20.72% and 16.95%, respectively. The different effects between the two movements may be caused by the speed requirement in conducting the movement, i.e., turning movement is usually associated with slowing down, while lane-change, in most cases, requires speeding up. Note that in airbag information, both “No airbag” and “Deployed airbag” are associated with increasing the probability of higher injury severities, which have different explanations. As noted in previous literature, drivers owning the vehicle with the safety features, e.g., airbag, tend to be a safer driver (Mannering and Bhat, 2014). On one hand, the airbag itself is an important in-vehicle safety protection device. On the other hand, people, who choose vehicles with no airbag, tend to be unsafe drivers (Levitt and Porter, 2001). In this case, “No airbag” increases the injury severity level due to not only lacking airbag, but also low safety consciousness of the driver. However, “Deployed airbag” is another story. Usually, “Deployed airbag” is strongly related to heavy impact, which naturally indicates a serious crash. A similar explanation fits the ejection conditions. Although ejection might do some help in relieving the power of hitting, the ejection condition captures the uncontrolled impact severities. Accordingly, the heaviest impact throws the driver out of the vehicle, i.e., “Ejection totally”, and the driver is likely seriously injured, e.g., increasing the probability of serious and fatality injury (S/F) by 708.80%; and less heavy impact makes the driver partially out of the vehicle, i.e., “Ejection partially”, and increases the probability of evident injury (E) and serious and fatality injury (S/F) by 13.15% and 147.17%.

Effects of driver factors

As a most widely applied safety measure, seatbelt has been supported in many studies for saving life in crashes. There is no doubt that “Unused” seatbelt increases the evident injury (E) and serious injury and fatality (S/F) by 20.19% and 236.13%. And since the usage of seatbelt is restricted by law in Washington State, “Unused” seatbelt also captures the unobserved heterogeneity in drivers’ safety consciousness (Levitt and Porter, 2001, Kim et al., 2013, Gong and Fan, 2017). Driving without liability insurance, i.e., “No” liability indicator, is almost the same case. The use of “Child seat” indicates the fact that child is on board, which results in more conservative driving behaviors. Accordingly, as shown in Table 4-5, the “Child seat” indicator reduces the probabilities of possible injury (P), evident injury (E) and serious and fatality injury (S/F) by 2.51%, 31.38%, and 32.11%, respectively. When it comes to the sobriety condition, it is obviously that “Impaired” condition increases the probability of serious and fatality injury (S/F) dramatically, i.e., 211.01%. Gender and age of drivers are identified to capture heterogeneity across crash records in the means of random parameters. And the elasticity test results show that the impacts of the gender and age characteristics are in consistent with common knowledge. More specifically, male drivers feature aggressive driving behaviors when compared to female drivers. As a result, the probabilities of evident injury (E) and serious and fatality injury (S/F) are increased; aged drivers are easily injured due to declined physical function, which is reflected by the increasing probability of different injury severities, which is also observed in past research (Kim et al., 2013).

Temporal instability analysis

In addition to the average elasticity analysis discussed above, time-varying pseudo elasticity analysis is separately conducted for 2010-2014, 2015, and 2016 in accordance with the significant temporal indicators. Figure 4.2 illustrates the pseudo elasticity estimations for variables with significant impacts. The pseudo elasticity estimation mainly focused on injury severities, and the estimation results for non-injury are omitted.

It is found that the temporal indicators in class membership capture the instability of effects of the explanatory variables across the three time periods. As shown in Figure 4.2, a large proportion of significant variables show unification trends from 2010 to 2016. And the rest variables show almost the same effects on the probabilities of some specific levels of injury severity, such as animal collision for serious and fatality injury (S/F), low speed limit for serious and fatality injury (S/F), and so on. By comparing the elasticity among the three time periods, Figure 4.2 shows that half of the variables have the reducing elasticity value, which may indicate the achievements of the safety countermeasures applied in Washington state and the safety improvement in vehicle techniques from 2010 to 2016. However, there still exist a set of factors, whose elasticity value increased from 2010 to 2016, which indicate the potential directions for future improvement. For instance, the pseudo elasticity value of vehicle age [16, 70) on serious and fatal injury (S/F) crashes increased from 7.91%, during 2010-2014, to 9.31% in 2016. The possible explanation may be that new safety techniques are less likely to be applied in old vehicles (Hoye, 2019). As a result, as time goes by, the difference in safety performance between new vehicles and old vehicles, especially for those older than 16 years, is enlarged. New safety-related techniques to fit old vehicles and new police encouraging the elimination of old cars are of great interests.

4.4. Summary

A seven-year crash dataset from 2010 to 2016 is utilized to investigate highway single-vehicle crashes and the effects of contributing factors on driver injury severities in Washington States. A latent class mixed logit model with temporal indicators is developed for analyzing this dataset. The proposed model is able to interpret both within- and across- class unobserved heterogeneity and temporal instability. Model goodness-of-fit measurements, including AIC and BIC, were conducted to compare the models with different numbers of latent classes. The two-class model outperformed the other models with higher number of classes in terms of lower AIC and BIC.

The temporal indicators, including Year 2015 and Year 2016, show significant influence on latent class membership, indicating that the effects of the explanatory variables on injury severity varies significantly in 2015 and 2016. Urban indicator and principal indicator are identified to be random parameters and have significant heterogeneity in the means as different functions of male, driver's age indicator for [45, 65) and driver's age for above 65. The model also includes a wide variety of factors relating general crash characteristics (collision object type and seasonal indicator), environment characteristics (road function, surface type, and speed limit), vehicle characteristics (vehicle age, airbag condition, and ejection condition), and driver characteristics (driver age, belt usage, liability condition, and sobriety condition). The effects of the significant factors on driver injury severities are analyzed using pseudo elasticity estimations. Our results are generally in line with past studies that investigated the factors affecting single-vehicle crash severity. Based on the temporal elasticity analysis results, it is found that elasticity estimates of some significant variables (such as overturn collision, off-road collision, winter, snow, turning movement, deployed airbag, child seat, no liability and old driver indicator on serious injury and fatality) reduce during the studying periods, while the others (such as wet surface, ice surface, old vehicle, lane-change movement, no-airbag, partial and total ejection, impaired driver, and male driver on serious injury and fatality) increase or maintain the same values. Based on the time-varying effects and previous engineering experience, appropriate countermeasures and police recommendations could be implemented to reduce highway single-vehicle crashes.

There exist some limitations that may affect result estimation and interpretations in this study. Although a large range of impact factors have been considered in this study, the spatial spillover effect, i.e., some observed characteristics at one crash location not only influence injury severity at this location, but also effect on the probability of injury severity at neighboring sites, is overlooked (Mannering and Bhat, 2014). As a result, spatial correlation is not considered in the proposed model. Moreover, the proposed model applied temporal indicators to illustrate the temporal instability of effects of various impact factors across different time periods. However, the proposed model can hardly be used in injury severity prediction. In order to address this issue, models, that allow the macroscopic variables to track temporal heterogeneity, can be developed (Xiong et al., 2014; Li et al., 2019b).

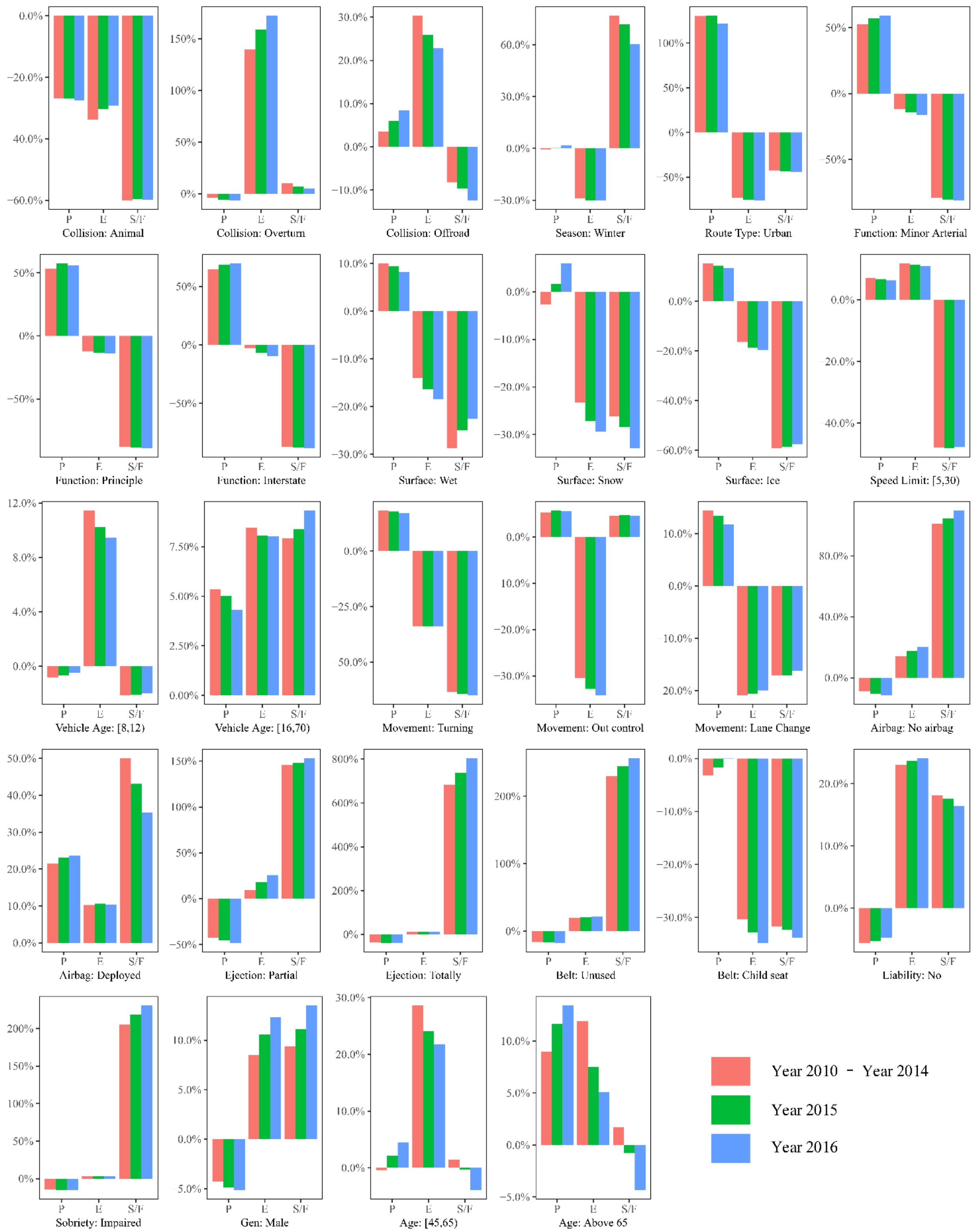

Figure 4.2. Time-varying elasticity estimation results.

CHAPTER 5. MIXED LOGIT AND LATENT CLASS MODEL APPROACHES IN DRIVER INJURY SEVERITY ANALYSIS

In this chapter, we compare the mixed logit model and the latent class model in investigating the driver injury severity of the single vehicle crashes under rain conditions. In this study, mixed logit model (MLM) and latent class model (LCM) are both developed and compared based on a three-year crash dataset in four south central states, i.e., Texas, Arkansas, Oklahoma, and Louisiana. In addition, a pseudo-elasticity analysis approach is proposed to analyze the impact factors on driver injury severity. Also, several parsimony indices, e.g., AIC and BIC, and as well as McFadden pseudo r-squared, are calculated for both the models to evaluate their respective performances. Finally, this study provides insights on casualties and injury prevention. Results show that choosing the uniform distribution as the prior for random parameters could better improve the goodness-of-fit of the MLM than using normal and lognormal distributions. In addition, the two-class LCM also shows superiority when compared to three- and four-class LCMs. Finally, a careful comparison between these two models is conducted, and the results indicate that the LCM has a slightly better performance in analyzing the study dataset in this study. Model estimation results show that *curve*, *on grade*, *signal control*, *multiple lanes*, *pickup*, *straight*, *drug/alcohol impaired*, and *seat belt not used* can significantly increase the probability of incapacitating injuries and fatalities for drivers in the two models. On the other hand, *wet*, *male*, *semi-trailer*, and *young* can significantly decrease the probability of incapacitating injuries and fatalities for drivers. This study provides an insightful understanding of the effects of these attributes on rural single-vehicle crashes under rain conditions and beneficial references for developing effective countermeasures for severe injury prevention. The rest of the chapter is organized as follows: Section 5.1 provides an explicit description of the dataset. The detailed methodology design is described in Section 5.2. The model analysis results and discussions are illustrated in Section 5.3. Finally, the entire research effort is concluded in Section 5.4.

5.1. Data

A three-year crash dataset including all rural single-vehicle crashes under rain conditions in four South Central states from 2012 to 2014 is utilized in this research. This dataset is obtained from the Texas Department of Transportation (TxDOT), Arkansas State Highway and Transportation Department (AHTD), Oklahoma Department of Transportation (OKDOT), and Louisiana Department of Transportation and Development (LADOTD). This new data set includes more explicit crash attributes which can be captured more appropriately by LCMs so that we can fully demonstrate the modeling process and develop the model specifications. The geographical locations of these four states are illustrated in Figure 5.1. These states are concentrated in the south-central United States and have similar climatic characteristics, such as precipitation characteristics (Carter et al., 1974), and annual temperature variations (Aguilar et al., 2005). In addition, the similar demographic features of these states indicate that they can be studied as a whole, as evidenced by numerous peer studies in different fields (Adams et al., 2016; Miller et al., 2013; Munn et al., 2002).

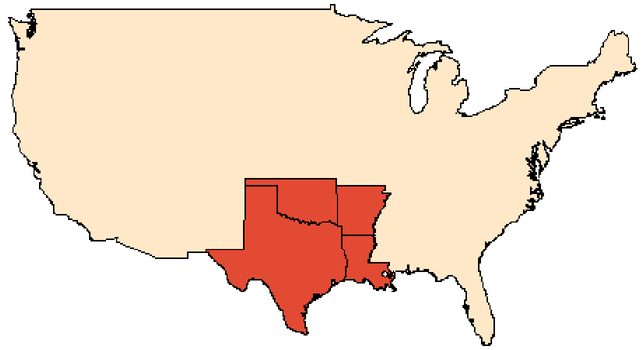

Figure 5.1 Location map of study areas

In view of the differences in the process of recording crash reports in the four states, only the identical variables from the sub-dataset of each of the four states were selected in the study. After careful examination, the incomplete and erroneous records in the original data set were deleted. Finally, 17,929 accurate records were retained in the study for modeling analysis. The final dataset contains detailed information on driver injury severity and potential contributing factors regarding the characteristics of the crash, vehicle, and driver, such as road geometry, vehicle type, and driver demographics. The driver injury severity was classified into five subtypes in the original files, including fatal injury, incapacitating injury, visible injury, complaint of injury, and no apparent injury, respectively. In this study, fatal injury and incapacitating injury are combined together as the most severe injury severity level to maintain a statistically meaningful sample size. The classifications of other injury severities are consistent with the original files. Therefore, driver injury severity is categorized into four subtypes in this study, including O (original category: no apparent injury), C (original category: complaint of injury), B (original category: visible injury), and AK (original category: fatal injury and incapacitating injury), respectively. Some continuous integer variables (including driver age and number of vehicles in the crash) are categorized as discrete variables with a finite number of exclusive values, as a constant coefficient may not fully reveal the various impacts of a continuous variable on driver injury severity when its numerical value falls into different ranges (Chen et al., 2016b). Moreover, based on our previous studies and engineering experience (Chen et al., 2015a, 2016c, 2015b), some multi-categorical variables with an excessive number of original values are simplified to improve modeling efficiency. For example, right turn and making a right turn on red are combined as a variable, right turn, to reduce the number of categorical values in a variable. Furthermore, variables with relatively similar impacts on driver injury severity but not having enough records of presence, such as alcohol-impaired and drug-impaired, are combined as an integrated factor for modeling simplification purpose. The descriptive statistics of the studied dataset are shown in Table 5-1.

Table 5-1 Variable Definition and Description

Variable	Driver injury severity								Total
	O (Mean)		C (Mean)		B (Mean)		AK (Mean)		
Severity	10850	0.61	3855	0.22	2387	0.13	837	0.05	17929
Light Condition									
Dark	3909	0.61	1308	0.2	896	0.14	337	0.05	6450
Dawn	151	0.64	24	0.1	41	0.17	20	0.08	236

Variable	Driver injury severity								Total
	O (Mean)		C (Mean)		B (Mean)		AK (Mean)		
Daylight	6790	0.6	2523	0.23	1450	0.13	462	0.04	10982
Road Character									
No curve	8734	0.61	3302	0.23	1747	0.12	564	0.04	14347
Curve	2116	0.59	553	0.15	640	0.18	273	0.08	3582
Road Grade									
Level	9339	0.61	3345	0.22	2059	0.13	662	0.04	15405
Hillcrest	255	0.63	74	0.18	54	0.13	20	0.05	403
On grade	1256	0.59	436	0.21	274	0.13	155	0.07	2121
Road Surface Condition									
Dry	143	0.52	62	0.23	40	0.15	28	0.1	273
Wet	10629	0.61	3776	0.21	2288	0.13	786	0.05	17258
Snow	78	0.68	17	0.15	14	0.12	5	0.04	114
Traffic Control									
No Control	1489	0.62	540	0.22	334	0.14	44	0.02	2407
Stop-Yield Sign	53	0.65	12	0.15	13	0.16	3	0.04	81
Signal Control	9308	0.6	3303	0.16	2040	0.17	790	0.07	8666
Number of Lanes									
One Lane	589	0.64	165	0.18	135	0.15	29	0.03	918
Two Lanes	8738	0.59	3383	0.23	1923	0.13	670	0.05	14714
Multiple Lanes	1523	0.66	307	0.13	329	0.14	138	0.06	2297
Vehicle Type									
Passenger Car	9481	0.61	3501	0.23	1998	0.12	692	0.04	13715
Pick-up	1214	0.59	342	0.17	374	0.18	141	0.07	2071
Semi	117	0.79	12	0.08	15	0.1	4	0.03	148
Bus	38	1	0	0	0	0	0	0	38
Action									
Straight	7990	0.61	3052	0.23	1443	0.11	560	0.04	13044
Right Turn	1319	0.57	364	0.16	455	0.2	164	0.07	2302
Left Turn	1450	0.59	434	0.18	479	0.19	113	0.05	2476
U-Turn	7	1	0	0	0	0	0	0	7
Slowing	70	0.82	5	0.06	10	0.12	0	0	85
Backing	14	1	0	0	0	0	0	0	14
Seat Belt used									
Used	10288	0.61	3659	0.22	2162	0.13	694	0.04	16803
Not used	562	0.5	196	0.17	225	0.2	143	0.13	1126
Drug/Alcohol Impaired	257	0.39	142	0.21	178	0.27	86	0.13	663
Gender									
Male	6323	0.63	1937	0.19	1289	0.13	487	0.05	10036
Female	4527	0.57	7893	1	1098	0.14	350	0.04	7893
Age									
Young (<25 years)	4468	0.61	1544	0.21	1033	0.14	259	0.04	7304
Middle (25~64 years)	5938	0.6	2141	0.22	1246	0.13	530	0.05	9855
Old (>64 years)	444	0.58	170	0.22	108	0.14	48	0.06	770

5.2. Methodology

5.2.1. Mixed logit model

Assuming that driver injury severities are classified into K levels (in this study $K = 4$), and given the fact that the studied dataset is regarding single-vehicle crashes, the function determining the driver injury severity level k ($k \in K$) for the nth driver, Y_{kn}, is given by

$$Y_{kn} = \boldsymbol{\beta}_k \boldsymbol{X}_{kn} + \varepsilon_{kn} \tag{5-1}$$

where $\boldsymbol{\beta}_k$ is a vector of parameters to be estimated for driver injury severity level k which may vary across observations, $\boldsymbol{X}_{kn}$ is a vector of explanatory variables (light conditions, traffic controls, driver ages, etc.), and the disturbance term is notated as ε_{kn}, which is assumed to be generalized extreme value distributed (McFadden, 1981). Consequently, the standard multinomial logit model (neglecting for the error components) can be expressed as

$$P_n(k) = \frac{e^{\boldsymbol{\beta}_k \mathbf{X}_{kn}}}{\sum_{\forall k \in K} e^{\boldsymbol{\beta}_k \mathbf{X}_{kn}}} \tag{5-2}$$

where $P_n\ (k)$ is the probability of the nth driver having kth severity level. Supposing the random parameters that capture unobserved heterogeneity on driver injury severity outcomes are given by $f(\boldsymbol{\beta}_k|\boldsymbol{\varphi})$, where $\boldsymbol{\varphi}$ is a vector that representing the probability density function (PDF). According to previous studies (McFadden and Train, 2000; Train, 2003), in the MLM, the resulting outcome probabilities $P_n(k|\boldsymbol{\varphi})$ are given by

$$P_n(k|\boldsymbol{\varphi}) = \int \frac{e^{\boldsymbol{\beta}_k X_{kn}}}{\sum_{\forall k \in K} e^{\boldsymbol{\beta}_k X_{kn}}} f(\boldsymbol{\beta}_k|\boldsymbol{\varphi}) d\boldsymbol{\beta}_k \tag{5-3}$$

Consequently, the individual specific variations of the impacts of the corresponding variable vector, $\boldsymbol{X}_{kn}$, is accounted by the parameter vector, $\boldsymbol{\beta}_k$. Different with the standard multinomial logit model, $\boldsymbol{\beta}_k$ is not a fixed parameter and may be randomly distributed with various mode and skewness. Normal distribution is the most familiar, simplest assumption of the distribution of $\boldsymbol{\beta}_k$, and is specified as,

$$\boldsymbol{\beta}_k = \boldsymbol{\beta}_i + \sigma_i \nu_i,\ \nu_i \sim N(0,1) \tag{5-4}$$

where $\boldsymbol{\beta}_i$ is the mean, σ_i is the standard deviation of the distribution, and ν_i is the individual-specific heterogeneity, with mean equal to zero and standard deviation equal to one (Greene, 2012; Li et al., 2018b).

In this study, not only the normal distribution but also other types of distribution are chosen to limit the parameter values to a specific range based on engineering experience. For example, as shown in Eq. (5-5), the lognormal distribution is used to maintain the values of certain parameters (e.g., drug-impaired, seatbelt used, etc.) to be either positive or negative

$$\boldsymbol{\beta}_k = e^{\boldsymbol{\beta}_i + \sigma_i \nu_i},\ \nu_i \sim N(0,1) \tag{5-5}$$

In addition, to ensure a reasonable range of the variation of a parameter, uniform distribution is selected on certain parameters to replace the normal distribution that assumes the variation range is infinite, and is given by

$$\boldsymbol{\beta}_k = \boldsymbol{\beta}_i + \sigma_i \nu_i,\ \nu_i \sim unif(-1,1) \tag{5-6}$$

Normal distribution, lognormal distribution, and uniform distribution are separately assumed for each potential random parameter. The model is finalized based on the characters of these parameters as well as some parsimony indices, including Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). AIC and BIC are defined as follows, respectively,

$$\mathrm{AIC} = -2\ln(L) + 2p \tag{5-7}$$

$$\mathrm{BIC} = -2\ln(L) + p \times \ln(N) \tag{5-8}$$

where $\ln(L)$ is the model likelihood, p is the number of estimated model parameters in the model, and N is the total number of observations. Lower AIC or BIC value of a candidate model generally indicates the model has better fit and is closer to the "true" model.

McFadden Pseudo R-squared (McFadden and Zarembka, 1974) statistic is also applied to demonstrate the model fitness. The formula is given as

$$R^2 = 1 - \frac{\ln \hat{L}(M_{Full})}{\ln \hat{L}(M_{Constant})} \tag{5-9}$$

where $\hat{L}$ is the estimated likelihood, M_{Full} is the full model with the constant term and all predicting variables, $M_{Constant}$ is the intercept model only including the constant term. The log likelihood of the intercept model is treated as a total sum of squares (TSS), and the log likelihood of the full model is treated as the sum of squared errors of prediction (SSE). The ratio of the likelihoods suggests the level of improvement over the intercept model offered by the full model. According to previous research (Domencich and McFadden, 1975), the McFadden Pseudo R-squared value in this study indicates that the full model is a significantly better fit compared with the intercept model.

5.2.2. Latent class model

The LCM has some similarities with the MLM but embodies several critical differences as well. This model is generally used to discover potential subtypes or confirm hypothesized subtypes based on multivariate categorical data (Lazarsfeld et al., 1968). The underlying theory of the LCM posits that individual behavior depends on observable attributes and on latent heterogeneity that varies with factors that are unobserved by the analyst. In an LCM framework, a discrete distribution (e.g., multinomial distribution) is selected to model unobserved heterogeneity across observations. The latent classes that are not revealed to analyst could be treated as different bins where an individual resides in based on its own characteristics.

Assuming there are C distinct latent classes in the model, the probability of the nth crash record belonging to class c ($c \in C$) specified by an MNL model is defined as:

$$P_n(c) = \frac{e^{\boldsymbol{\theta}_c \gamma_n}}{\sum_{\forall C} e^{\boldsymbol{\theta}_c \gamma_n}} \tag{5-10}$$

where $\boldsymbol{\gamma}_n$ is a vector of characteristics that determine class c probabilities for nth crash, $\boldsymbol{\theta}_c$ is the corresponding vector of estimable parameters. In addition, the conditional probability of in the nth driver in class c having level k injury severity is given by

$$P_n(k|c) = \frac{e^{\boldsymbol{\beta}_{kc} \mathbf{X}_{knc}}}{\sum_{\forall k \in K} e^{\boldsymbol{\beta}_{kc} \mathbf{X}_{knc}}} \tag{5-11}$$

Finally, the unconditional probability of driver in the nth crash having level k injury severity is defined as:

$$P_n(k) = \sum_{\forall C} P_n(c) \times P_n(k|c) \tag{5-12}$$

As mentioned above, significant research effort has been made in search of an optimal number of classes when an LCM is developed. In this study, a statistical accrual searching process is utilized to find the optimal number of latent classes, starting with 2 latent classes and increasing by 1 in each step up to the maximum plausible number of latent classes, and AIC and BIC are selected to assess model fitness. Estimation of the variables in this study is conducted with an iterative numerical method, the maximum likelihood estimation (MLE) algorithm. The estimated asymptotic covariance matrix is based on the second derivatives of the specific utility functions. The Berndt–Hall–Hall–Hausman (BHHH) estimator is used in case that the matrix fails to be positive because of rounding error (Berndt et al., 1974).

5.2.3. *Pseudo-elasticity analysis*

Extensive studies have proved that the signs of parameter estimation results are not always consistent with the real impacts of these parameters when a multinomial response variable is applied in model design (Kim et al., 2013; Osman et al., 2016; Wu et al., 2014). Therefore, an elasticity analysis is necessary to assess the influences of statistically significant variables in each of the proposed models, given the multi-level driver injury severity outcome. The elasticity is calculated in the form of the partial derivative for each observation (Washington et al., 2011),

$$E_{X_{kni}}^{P_{kn}} = \frac{\partial P_{kn}}{\partial X_{kni}} \frac{X_{kni}}{P_{kn}} \tag{5-13}$$

where $E_{X_{kni}}^{P_{kn}}$ is the elasticity outcome for the driver of the nth crash with severity level k, X_{kni} is the value of the ith variable for the nth crash in the propensity function with respect to the kth injury severity level. However, Eq. (5-12) is not applicable for this study since the variables have already been transformed into binary forms (with 0/1 outcome), and the probabilities are not differentiable with respect to indicator variables. To address this issue, direct pseudo-elasticity is defined in Eq. (5-13) by modifying Eq. (5-12) to calculate the influence of each significant indicator variable (Kim et al., 2007),

$$E_{(p)}{}_{X_{kni}}^{P_{kn}} = \frac{P_{kn}[given\ X_{kni}=1] - P_{kn}[given\ X_{kni}=0]}{P_{kn}[given\ X_{kni}=0]} \tag{5-14}$$

where $E_{(p)}{}_{X_{kni}}^{P_{kn}}$ is the pseudo-elasticity of the probability and is defined as the percentage change in probability when an indicator variable is switched (i.e., from 0 to 1 or from 1 to 0); P_{kn} is the probability the driver of *n*th crash having an injury severity level *k* for the given value of the variable X_{kni} while holding other variables constant. The direct pseudo-elasticity in Eq. (5-14), $E_{(p)}{}_{X_{kni}}^{P_{kn}}$, is calculated for each record in the dataset, and the average pseudo-elasticity is calculated based on all data records to measure variable influence.

5.2.4. *Temporal stability test*

In order to test and compare the temporal stability of MLM and LCM, a series of likelihood ratio tests are conducted. These tests are used to compare models developed for two different years and examine if the parameter estimates are stable between the two years. The test statistic follows a χ^2 distribution with degrees of freedom equal to the number of estimated parameters and can be written as (Washington et al., 2011)

$$\chi^2 = -2[LL(\boldsymbol{\beta}_{t_2t_1}) - LL(\boldsymbol{\beta}_{t_1})] \tag{5-15}$$

where $LL(\boldsymbol{\beta}_{t_2t_1})$ is the log-likelihood at convergence of a model containing converged parameters based on using year $t_2's$ data, while using data from year t_1, and $LL(\boldsymbol{\beta}_{t_1})$ is the log-likelihood at the convergence of the model using year $t_1's$ data. It should be noted that the parameters are no longer restricted to using year $t_2's$ converged parameters as are the case for $LL(\boldsymbol{\beta}_{t_2t_1})$. This test is also reversed such that year t_1 above becomes year t_2 and year t_2 above becomes subset year t_1. The resulting χ^2 statistic can be used to determine if the null hypothesis that the parameters are equal in the two years can be rejected.

5.3. Estimation results and discussion

The NLOGIT 5 software is utilized for model estimation. It should be noted that although we introduced temporal instability, the detailed estimation results of each year are omitted due to the high complexity.

Instead, the three-year overall parameter estimation results and corresponding pseudo-elasticity results are presented for discussing their impacts on driver injury severity.

5.3.1. Mixed logit model estimation results

A simulation-based maximum likelihood estimation (MLE) method is used to estimate model parameters in the MLMs. Three models with different assumptions on parameter distributions, including normal, lognormal, and uniform distributions, are examined respectively. By balancing the computational cost-efficiency and model goodness-of-fit, simulations with 1,000 Halton draws are applied in each model to provide an efficient estimation (Train, 2000). In addition, the O level is selected as the reference level. The comparison results of the three models are provided in Table 5-2.

Table 5-2 Comparison Results of MLMs with Different Distributions

Model No.	1	2	3
Distribution	Normal	Lognormal	Uniform
Significant random parameters (Injury Severity)	Curve (C), Male (C), Young driver (AK), Male (AK)	None	Curve (C), Male (C), Young driver (AK), Male (AK)
Log likelihood	-18138.98	-18226.73	-18130.34
Number of estimated model parameters	26	18	26
AIC	36329.96	36489.46	36312.68
BIC	36532.61	36629.76	36515.33

As illustrated in Table 5-2, the results indicate that using lognormal distribution as the prior distribution assumption (Model 2) is not appropriate for this crash dataset since there is no significant random parameter found with this assumption. In contrast, several parameters (e.g., curve, male, etc.) are found to be randomly distributed in the other two models with the uniform and normal distributions assumptions (Model 1 and Model 3), indicating that the two distributions are both applicable for analyzing the dataset. In addition, Model 2 has the highest AIC and BIC values, also demonstrating that the lognormal distribution works inferior to the other two. Based on the rule-of-thumb of AIC and BIC (Schermelleh-Engel et al., 2003), using the uniform distribution as prior shows much better performance than using the normal distribution. Thus, Model 3 is selected as the final model in this study. The estimation results using MLM with uniform distribution simulated random parameters (Model 3) are illustrated in Table 5-3.

Table 5-3 Estimation Results of MLM

Variable	Coefficient	Standard error	t-stat	95% Confidence interval	
				Lower	Upper
Constants					
C	2.43*	0.20	12.15	2.04	2.82
B	1.38*	0.18	7.67	1.03	1.73
AK	1.72*	0.19	9.05	1.35	2.09
Mean of random parameters					

Variable	Coefficient	Standard error	t-stat	95% Confidence interval	
				Lower	Upper
Male (C)	-1.04*	0.27	-3.85	-1.57	-0.51
Curve (C)	-0.52	0.49	-1.06	-1.48	0.44
Male (AK)	-1.61*	0.53	-3.04	-2.65	-0.57
Young (AK)	-1.63*	0.51	-3.20	-2.63	-0.63
Spread scale of random parameters					
Male (C)	2.96*	0.32	9.26	2.33	3.59
Curve (C)	3.91*	0.71	5.51	2.52	5.31
Male (AK)	3.69*	0.44	8.38	2.83	4.55
Young (AK)	3.00*	0.45	6.66	2.11	3.88
Fixed parameters					
Semi (C)	-0.91*	0.39	-2.33	-1.67	-0.15
Straight (C)	-0.85*	0.07	-12.14	-0.99	-0.71
Drug/Alcohol Impaired (C)	1.86*	0.16	11.63	1.55	2.17
Signal Control (B)	-0.47*	0.04	-11.75	-0.55	-0.39
Multiple Lanes (B)	-0.36*	0.07	-5.14	-0.50	-0.22
Semi (B)	-0.84*	0.31	-2.71	-1.45	-0.23
Drug/Alcohol Impaired (B)	0.41*	0.11	3.73	0.19	0.63
Male (B)	-0.31*	0.04	-7.75	-0.39	-0.23
Curve (AK)	0.88*	0.11	8.00	0.66	1.10
On grade (AK)	0.57*	0.13	4.38	0.32	0.82
Wet (AK)	-0.61*	0.20	-3.05	-1.00	-0.22
Pick-up (AK)	0.31*	0.13	2.38	0.06	0.56
Semi (AK)	-1.48*	0.68	-2.18	-2.81	-0.15
Seatbelt not used (AK)	1.88*	0.17	11.06	1.55	2.21
Drug/Alcohol Impaired (AK)	2.08*	0.21	9.90	1.67	2.49
Model statistics					
Number of observations (*N*)	17929.00				
Log-likelihood at constant	-34854.87				
Log-likelihood at convergence	-18130.34				
McFadden Pseudo R-squared	0.47				

* Level of significance $\leq$ 5%.

A variety of variables is found significantly associated with driver injury severity. The variable, *male driver*, is found as a random parameter affecting both possible injury (C) and more severe injury (AK), with a statistically significant mean and standard deviation with respect to each injury level. The variable, *curve*, is also a random parameter that has an influence on possible injury (C), although its mean is not significant based on the level of significance. This issue is not critical because whether a parameter is random or not is primarily based on the significance indication of its distribution of the standard deviation instead of its mean. The t-stat result of the standard deviation indicates that the variable, *curve*, is a random parameter. In addition, the variable, *young age*, is also found as a random parameter in the utility function of AK injury severity.

5.3.2. Latent class model estimation results

For the LCM, three different class numbers, 2 to 4, are separately tested with the same dataset, and their examination results are illustrated in Table 5-4. As the class number increases, the AIC and BIC values of the model also slightly increase, indicating that the performance becomes degraded. In addition, another two indices, the class probability and level of significance for each class, are also adopted for model selection. When the dataset is classified into two classes, the two classes involve 70% and 30% of the total data, respectively. When assuming the whole dataset contains three latent classes, the larger sub-dataset is then approximately evenly divided and the smaller one remains almost the same, as shown by the percentages 36%, 33% and 31%, respectively. While in the four-class scenario, approximately both the two sub-datasets in the two-class model are more delicately divided, and the final four classes contain 66%, 4%, 18%, and 12% of the whole data, respectively. However, the delicate division is not beneficial for improving model performance. It shows in Table 5-4 that the class probabilities in the three-class model are not statistically significant. In addition, although all the class probabilities in the four-class scenario are significant, the deficient parameter estimation results show that this kind of division is less meaningful since considerably fewer parameters (12 versus 17/18) are found to be significantly related to driver injury severity. Therefore, the two-class model is selected as the final model for studying the variables' impacts on driver injury severity outcomes in rural single-vehicle crashes under rain conditions.

Table 5-4 Comparison Results of LCMs with Different Numbers of Classes

The number of latent classes	2	3	4
Log likelihood	-18038.60	-18196.38	-18203.31
Class probability	70%*/30%*	36%/33%/31%	66%*/4%*/18%*/12%*
Number of estimated model parameters (p)	17	18	12
AIC	36111.20	36428.76	36430.62
BIC	36243.70	36569.06	36524.15

* Level of significance $\leq$ 0.05.

Table 5-5 Estimation Results of LCM

Variable	Latent Class 1					Latent Class 2				
	Coef.[a]	S.E.[b]	t-stat	95% CI[c]		Coef.[a]	S.E.[b]	t-stat	95% CI[c]	
				Lower	Upper				Lower	Upper
Constant										
C	3.62*	0.61	5.93	2.42	4.82	0.13	0.36	0.36	-0.57	0.83
B	1.67*	0.62	2.69	0.46	2.89	0.70*	0.31	2.26	0.1	1.31
AK	0.26	1.36	0.19	-2.41	2.93	3.28*	0.55	5.96	2.21	4.35
Non-constant Parameter										
Curve (C)	0.58*	0.13	4.46	0.32	0.84	-	-		-	-
Straight (C)	-	-		-	-	-1.33*	0.23	-5.78	-1.78	-0.88
Drug/Alcohol Impaired (C)	1.50*	0.13	11.54	1.25	1.77	-	-		-	-
Male (C)	-	-		-	-	-0.70*	0.22	-3.18	-1.14	-0.28
Signal Control (B)	2.52*	1.12	2.25	0.34	4.71	-4.86*	1.32	-3.68	-7.45	-2.28
Multiple Lanes (B)	-0.26*	0.11	-2.36	-0.48	-0.05	-1.71*	0.43	-3.98	-2.55	-0.87
Male (B)	-0.34*	0.1	-3.40	-0.55	-0.14	-1.51*	0.36	-4.19	-2.21	-0.81
On grade (AK)	1.16*	0.35	3.31	0.47	1.86	-	-		-	-
Wet (AK)	-1.09*	0.4	-2.73	-1.88	-0.32	-	-		-	-
Pick-up (AK)	1.55*	0.4	3.88	0.78	2.33	-	-		-	-
Young (AK)	-0.73*	0.3	-2.43	-1.33	-0.14	-0.65*	0.11	-5.91	-0.88	-0.43
Seatbelt not used (AK)	2.50*	0.47	5.32	1.58	3.43	1.01*	0.23	4.39	0.57	1.46
Drug/Alcohol Impaired (AK)	3.50*	0.54	6.48	2.51	4.61	-	-		-	-
Male (AK)	-0.88*	0.29	-3.03	-1.45	-0.31	-0.35*	0.17	-2.06	-0.69	-0.01

Variable	Latent Class 1					Latent Class 2				
	Coef.[a]	S.E.[b]	t-stat	95% CI[c]		Coef.[a]	S.E.[b]	t-stat	95% CI[c]	
				Lower	Upper				Lower	Upper
Class probability	0.70*					0.30*				
Model statistics										
Number of observations (*N*)	17929									
Log-likelihood at constant	-34854.87									
Log-likelihood at convergence	-18038.6									
McFadden Pseudo R-square	0.48									

[a] Coefficient.

[b] Standard error.

[c] 95% confidence interval of estimation results.

* Level of significance ≤ 0.05.

The detailed estimation results of this model are listed in Table 5-5. It shows that remarkable differences exist between the two latent classes, since the variables that significantly influence driver injury severity to distribute quite diversely in the two classes. For instance, Drug/Alcohol-Impaired (C) and Drug/Alcohol-Impaired (AK) are found significantly affecting driver injury severities in Latent Class 1, whereas insignificant in Latent Class 2. These distinctive outcomes suggest that the data has a multivariate categorical nature and demonstrate the latent class logit model is appropriate for analyzing the crash dataset.

5.3.3. Pseudo-elasticity analysis results

In order to explain the impacts of these variables accurately, the average pseudo-elasticity is adopted on both the MLM and the LCM, and the results are listed in Table 5-6. Twelve variables regarding road geometric characteristics, road surface conditions, vehicle types, and driver demographic information and behavior, are found significant in the MLM while the elasticity analysis results of the two models are quite similar in terms of colored magnitude category, although the estimation values are not exactly same. The detailed discussions of these variables are presented in the following sections.

Table 5-6 Average Pseudo-Elasticity Analysis for MLM and LCM

Variable	MLM				LCM			
	O	C	B	AK	O	C	B	AK
Curve	-3.62%	-24.33%	35.87%	84.14%	-3.36%	-9.43%	34.69%	20.13%
On Grade	-2.55%	-2.37%	-2.37%	52.77%	-1.52%	-1.84%	-2.93%	59.88%
Wet	2.81%	2.34%	2.63%	-35.23%	2.13%	1.83%	4.69%	-43.94%
Signal Control	9.08%	-27.20%	7.20%	6.51%	9.06%	-41.38%	21.11%	48.08%
Multiple Lanes	7.62%	-25.39%	5.92%	5.40%	8.21%	-38.20%	11.31%	23.88%
Pick-up	-1.39%	-1.11%	-1.12%	28.22%	-1.60%	-2.58%	-2.90%	68.28%
Semi	30.52%	-47.83%	-39.72%	-61.11%	-	-	-	-
Straight	10.62%	10.20%	-43.42%	9.93%	6.56%	3.99%	-29.60%	11.79%
Drug/Alcohol Impaired	-42.37%	-10.58%	152.74%	204.72%	-43.73%	-15.46%	136.33%	502.67%
Seatbelt not used	-11.73%	-10.77%	-11.34%	265.43%	-7.63%	-6.50%	-13.89%	318.38%
Male	11.19%	-18.72%	-11.49%	-5.27%	12.27%	-24.98%	-8.93%	-16.81%
Young	2.47%	1.77%	2.22%	-38.24%	1.42%	0.66%	3.94%	-42.76%

Pseudo-elasticity analysis results of the two models show that the variable, *curve*, has an influence on increasing driver injury severity in rural single-vehicle crashes under rain conditions. In the MLM, this variable increases the probabilities of driver injury serveries, B level and AK level injuries, by 35.87% and 84.14%, respectively, while the corresponding values in the LCM are 34.69% and 20.13%, respectively. These results are consistent with previous studies where it was found that severe injuries are more likely to happen on the curve roads (Holdridge et al., 2005; Ye and Lord, 2014). Considering the impacts of rain, driving on curve roads becomes more challenging. It may take more time for vehicles to decelerate to a safe speed when running on a curve roadway because of the low friction on wet road surface. In addition, also due to the low friction, vehicles may easily lose control when sudden changes are applied to speed or steering while running at high speed on wet curve roads. Accordingly, possible countermeasures on this issue include increasing the radius of the curve where possible, installing a speed indicator at the beginning of the curve, and paving the curve road with materials with high resistance.

The variable, *road grade*, is also significantly associated with driver injury severity by showing that it can increase the likelihood of the driver being severely injured (AK level) by over 50% in both two models. The reasons for these results are complicated, one of the important factors is that vehicle brakes are more frequently used to maintain the vehicle stability while driving on the graded roadway, which may increase the risk of brake failure and then lead to the vehicle losing control. Similar findings have also been discovered by previous studies (Khattak, 2001; Quddus et al., 2009; Li et al., 2018a), and the same influence exists not only in single-vehicle crashes but also in multi-vehicle crashes. Enhanced delineation treatments on the roadway can alert drivers in advance of grade roads and vary depending on the severity of the grade and the driving speed. In addition, high friction surface materials and treatments also can be implemented to help the drivers to maintain speeds when driving on grade roadways (Li et al., 2018a).

The variable, *wet*, is found to significantly decrease the possibilities of AK level injuries in both models. The results seem to be contrary to everyday experience, however, in fact many previous studies have obtained similar conclusions (Lee et al., 2015; Wu et al., 2014). The reason may be that the drivers tend to adapt their speeds to the adverse road conditions to some degree while driving on the wet roads. This behavior is due to the driver's active behavioral adjustments to adverse external conditions to reduce and maintain low perceived driving risk, which is an example of risk compensation behavior. Interested readers are referred to the paper by Mannering and Bhat (2014) and the references cited therein.

It is not surprising that the variable, *signal control*, can aggravate the driver injury severity in both models. Signal control devices are mainly located at the intersections that are among the locations with the most complex traffic conditions in a road network (Wang and Abdel-Aty, 2008). In the U.S., although only 10% of all intersections were signalized, nearly 30% of intersection-related fatalities occurred at signalized intersections (Rice, 2007). Traffic signals, especially for the left-turn and through traffic, may increase the number of conflict points as well as accident potential on a roadway. This challenge becomes more serious under rain conditions. Due to the limited visibility and low pavement friction under rain conditions, the available response distances for drivers at intersections is significantly reduced, and therefore crash risks and severe injury possibilities are notably increased.

As illustrated in Table 5-6, the variable, *multiple lanes*, slightly increases the possibilities of AK level crashes. A probable explanation is that multiple lanes are always associated with complex roadway and traffic conditions, e.g., more exclusive turning lanes, frequent lane changing behaviors, and variable speed limits across lanes, and thus may pose more challenges to the drivers in the rain. This finding is consistent with the results of previous studies (Aziz et al., 2013; Wang et al., 2006), where it was found that crashes on multi-lane roads have a higher probability of fatality. On the contrary, single lane roads were found to have a lower probability of leading to severe injuries and fatalities.

The variable, *pickup*, is found significantly associated with driver injury severity in the rural single-vehicle crashes under rain conditions in the two models. More specifically, pickup drivers are more likely to suffer serious injuries and fatalities in rain-related single-vehicle crashes since the pseudo-elasticity analysis results showed that this variable could increase the probabilities of AK level injuries by 28.22% and 68.28% in the two models, respectively. The reason is understandable given that driving a pickup requires more driving skills and experiences than driving a passenger car. Besides, the higher inertia, which results from the larger mass of pickups comparing to passenger vehicles, also makes it more difficult to maintain safe driving, especially on the slippery road surface. Due to the difficulties in vehicle operation, rollovers, collisions with fixed objects, and other crash types with severe outcomes are more likely to occur in pickup related accidents.

The variable, *Semi-trailers*, can reduce the possibilities of the driver being seriously injured according to their low pseudo-elasticity results in serious crashes in the MLM. However, this variable is not significant in the LCM. Previous studies also implied that semi-trailer is a variable that has adverse effects on driver injury severity (Carson and Mannering, 2001; Celik and Oktay, 2014; Chen et al., 2016a). For instance, Chen et al. (2016a) found that semi-trailers are less related to severe injuries, indicated by the negative estimated coefficient. Carson and Mannering (2001) concluded that semi-trailers could increase the probability of fatal injuries due to its relatively large size and weight. Therefore, more efforts should be made to figure out the underlying reason for these impacts.

This variable, *drug/alcohol-impaired*, contains two aspects, i.e., *drug* and *alcohol*. However, they have quite similar impacts in compromising drivers' sobriety and reasonable judgment but did not have enough records of presence in the studied dataset, and therefore were combined together in this study. The combined variable, representing the drivers' state of consciousness, is expected to aggravate driver injury severity significantly. As shown in Table 5-6, *driver alcohol/drug impairment* can increase the potential for injuries and fatalities (B and AK levels) by over 150% in the two models. The results are reasonable since drug and alcohol can easily affect drivers' physical and psychological functions, e.g., body balance, vision, sobriety, reaction time, etc., and bring about a series of consequences, including misjudgment, short-term memory loss, reduced information processing capability, and impaired perception. Therefore, engineering, enforcement and educational (3E) traffic safety-related countermeasures are needed, including highway safety patrol and sobriety check on the random and timely basis, enforced punishment for driving under the influence (DUI), and related defensive driving programs.

As shown in Table 5-6, when the seatbelt is not used, the likelihood of drivers suffering AK level injury dramatically increases in the two models, suggesting that using a seatbelt is an effective way of protecting the driver in a rural single-vehicle crash under rain condition. The favorable effects of seatbelt usage have also been evaluated by previous research (Abay et al., 2013; Chu, 2014; Yasmin et al., 2014). The seatbelt, which is an important part of vehicle design, can secure the occupant of a vehicle against harmful movement during a collision or a sudden stop. When faced with crashes or other urgent circumstances or in the rain, it is harder for drivers to maintain normal driving on the slippery roadway, and thus seatbelts become more necessary to secure the driver against fierce movement and potential collision impact. Therefore, 3E efforts are also needed to ensure seatbelt usage on each occupant in every vehicle ride. For instance, a useful countermeasure could be video recognition techniques through roadway cameras applied to identify seatbelt usage status on vehicle occupants and issue traffic violation tickets and fines to those not wearing seatbelts, without violating people's rights of privacy. The elasticity analysis results suggest that *male* drivers have less likelihood of serious injury and fatality when comparing with female drivers. The reason for this finding may be that male drivers have relatively experienced operation skills and additional physiological strength and can better handle complex road and environment situations under rain conditions. Other scholars have also found similar results that male drivers demonstrate better driving performance and safety levels in the areas of the complex external environment that require additional driving skills than during average-day driving (Staff et al., 2014; Yasmin et al., 2014).

The *age* of drivers involved in the crashes is found to be a significant variable on driver injury severity in both models. The pseudo-elasticity analysis results indicate that young drivers are associated with reductions in the possibilities of driver serious injuries and fatalities in rural single-vehicle crashes under rain conditions. This is because young drivers have faster reactions than drivers in other age groups due to their physical flexibility (Castro et al., 2013; Xie et al., 2012). The relative lower average speeds in the rain may also contribute to reducing the injury severity of young drivers who are more likely to conduct speeding or reckless driving (Bolderdijk et al., 2011; Ulleberg, 2001). Furthermore, it is found in the dataset that 7.64% of young drivers drive pickup trucks, while the numbers in mid-age and old drivers are 15.30% and 10.66%, respectively. Given various impacts of vehicle types in the previous section, it is also evident to conclude that on average young drivers are safer than drivers of other age groups in rural single-vehicle crashes under rain conditions.

5.3.4. Comparison between mixed logit model and latent class model

A statistical comparison can better demonstrate which model is more appropriate for this dataset. As shown in Table 5-2, AIC and BIC of the MLM are 36312.68 and 36515.33, respectively. The same indices are also presented in Table 5- 4 for the LCM, and are 36111.20 and 36243.70, respectively. The relatively lower AIC and BIC values of the LCM indicate the model has slightly better performance than the MLM. In addition, the prediction success index, McFadden pseudo r-squared (0.47 for the MLM, 0.48 for the LCM), suggests that the LCM has slightly better predictive capability than the MLM. As shown in Table 5-7, it can also be observed that the LCM predicted probabilities for the O and C levels (contain over 80% observations) are closer to the observations than the ones predicted by the MLM.

Table 5-7 Estimated MLM and LCM Outcome Probabilities Compared to Observed Severity Outcomes

Components	Mixed Logit Model	Latent Class Model			Observed
		Latent Class 1	Latent Class 2	Overall	
Crash population share		0.70	0.30		
Crash injury severity					
O	0.594 (-1.82%)	0.604	0.595	0.601 (-0.66%)	0.605
C	0.223 (3.72%)	0.213	0.226	0.217 (0.93%)	0.215
B	0.131 (-1.50%)	0.122	0.135	0.126 (-5.26%)	0.133
AK	0.052 (10.64%)	0.061	0.044	0.056 (19.15%)	0.047

Tables 5-8 and 5-9 present the likelihood ratio tests between different years of the MLM and the LCM, respectively. The results show that both models are not temporally stable. This finding is in line with some recent research (Behnood and Mannering, 2015, 2016). It is impossible to determine whether this temporal instability is due to some underlying influence of factors that affecting driver injury severity, or the result of changes induced by variations in economic conditions. However, the significant differences indicate that the effect of explanatory variables on driver injury severity of rain-related crashes has shifted over the years studied. Moreover, the LCM shows that it tends to provide lower χ^2 values in the temporal stability tests. In addition, some χ^2 values are not significant, indicating the model is temporal stable in some extent. It should also be noted that the temporal instability indicates the estimation results of these two models may not be able to fully reveal the actual effects of the variables, and thus more advanced models that can account this issue are recommended in the future studies.

Therefore, from the comparison between the MLM and the LCM, it could be concluded that the LCM is slightly superior to the MLM in the modeling process on the studied dataset regarding rural single-vehicle crashes under rain conditions. However, it is noteworthy that the differences between the two models are rather modest, as the estimated parameters of the two models and the corresponding pseudo-elastic results are not significantly different.

Table 5-8 Likelihood Ratio Test Results between Different Years based on MLM (χ^2 Values with Degrees of Freedom in Parenthesis and Confidence Level in Brackets)

t_1	t_2		
	2012	2013	2014
2012	-	47.332 (24) [>99.70%]	54.631 (26) [>99.91%]
2013	48.350 (26) [>99.50%]	-	60.232 (25) [>99.99%]
2014	55.253 (26) [>99.99%]	53.282 (24) [>99.99%]	-

Table 5-9 Likelihood Ratio Test Results between Different Years based on LCM (χ^2 Values with Degrees of Freedom in Parenthesis and Confidence Level in Brackets)

t_1	t_2		
	2012	2013	2014
2012	-	34.332 (16) [>99.50%]	22.218 (17) [>82.22%]
2013	27.622 (17) [>99.50%]	-	40.335 (17) [>99.99%]
2014	24.333 (17) [>89.00%]	33.513 (16) [>99.40%]	-

5.4. Summary

A three-year crash dataset including all rural single-vehicle crashes under rain conditions in four South Central states, i.e., Texas, Arkansas, Oklahoma, and Louisiana, from 2012 to 2014 was selected in this paper to analyze the impact factors on driver injury severity. The MLM and the LCM are both developed in this study on the identical dataset. Several ongoing debates, including distributions of random parameters and efficient iterations in the MLM, the optimal class number of LCM, and comparison of the two models, are all discussed in the study. Statistical parsimony indices, including AIC, BIC, as well as McFadden pseudo r-squared, are calculated for each model to evaluate their respective performance. In addition, their abilities to capture temporal instability are also calculated via likelihood ratio tests. Results show that choosing uniform distribution as prior for random parameters increases goodness-of-fit of the MLM more than using normal and lognormal distributions. In addition, the two-class LCM also shows superiority when compared to the three- and four-class models. Moreover, a careful comparison between the two best models of their kinds is also conducted, and the results indicate that the LCM works slightly better in analyzing the dataset in this study.

A series of significant contributing factors in terms of road geometric characters, traffic compositions and dynamics, and driver demographic features, are identified and compared with the two models. To better explain the model estimation results, pseudo-elasticity analyses of the significant factors were conducted. The results reveal that *curve*, *on grade*, *signal control*, *multiple lanes*, *pickup*, *straight*, *drug/alcohol impaired*, and *seat belt not used* can increase driver injury severity in the two models. On the other hand, *wet*, *male*, *semi*, and *young* are found to decrease driver injury outcomes. These results are not only useful for understanding the underlying risk factors of rural single-vehicle crashes under rain conditions, but can also provide meaningful references for developing appropriate countermeasures, strategies, and policies to mitigate the driver injury severity of relative crashes worldwide. In addition, effective management and planning, technical implementation guide of specific countermeasures, and political support and leadership are necessary and should be fulfilled together to improve traffic safety.

CHAPTER 6. CONCLUSIONS AND RECOMMENDATIONS

6.1. Conclusions

This report documents the research activities to investigate the traffic crashes in RITI communities involving considerable incapacitating injuries and losses. The traffic crashes occurring in RITI communities are related more to features like speeding, low safety devices application (for instance, seatbelt), adverse weather conditions and lacking maintenance and repair for road conditions, and inferior lighting conditions, than urban crashes. Thus, it is necessary to study the properties and attributes of traffic crashes at the RITI area using data analysis methods, such as statical methods, data-driven methods, and so on. Unfortunately, there exists not only the unobserved heterogeneities but also the temporal instability in traditional crash data analysis. To address the research gap, this project employed the mixed logit model to examine the risk factors in determining driver injury severity in four crash configurations in two-vehicle rear-end crashes on state roads based on seven-years of data from the Washington State Department of Transportation. The research team developed a latent class mixed logit model with temporal indicators to investigate highway single-vehicle crashes and the effects of significant contributing factors to driver injury severity. In addition, this project also investigated the differences between the MLM and the LCM for exploring the relationships between driver injury severity in the rain-related rural single-vehicle crash and its corresponding risk factors.

Four mixed MNL models for each crash configuration and one model for the overall data were constructed. The general (Winter), environment-specific (Daylight, Surface wet, straight but not level), driver-specific (Male, Age under 24), and vehicle-specific (In front, Airbag not ejected) variables show heterogeneity on the injury but only in specific groups. Each model's elasticity analysis is conducted to determine the sensitivity of the possibility of severity to the change if the key factors are estimated in these five mixed MNL models. The similarities across all the models include drinking alcohol (whether impaired or not impaired) raises the risk of injury and even fatality in all cases; male drivers reduce the probability of injury in all circumstances; straight but not level contributes to unsafe driving; the vehicle in front significantly relates to injury; effects of airbag not ejecting and dark without light are related to the injury. Besides similarity, each configuration has specific characteristics. For example, daylight driving is safer for all the cases except the TT crashes; In front and Curve and no level are impact factors of fatality only in PT crashes. Age above 65 is a risk factor of fatality only for PP crashes; surface condition influences TP crashes the most.

In addition, a latent class mixed logit model with temporal indicators is developed. The proposed model is able to interpret both within- and across- class unobserved heterogeneity and temporal instability. Model goodness-of-fit measurements, including AIC and BIC, are conducted to compare the models with different numbers of latent classes. The two-class model outperformed the other models with higher number of classes in terms of lower AIC and BIC. The temporal indicators, including Year 2015 and Year 2016, show significant influence on latent class membership, indicating that the effects of the explanatory variables on injury severity varies significantly in 2015 and 2016. Urban indicator and principal indicator are identified to be random parameters and have significant heterogeneity in the means as different functions of male, driver's age indicator for [45, 65) and driver's age for above 65. The model also includes a wide variety of factors relating general crash characteristics (collision object type and seasonal indicator), environment characteristics (road function, surface type, and speed limit), vehicle characteristics (vehicle age, airbag condition, and ejection condition), and driver characteristics (driver age, belt usage, liability condition, and sobriety condition). The effects of the significant factors on driver injury severities are analyzed using pseudo elasticity estimations. Our results are generally in line with past studies that investigated the factors affecting single-vehicle crash severity. Based on the temporal elasticity analysis results, it is found that elasticity estimations of some significant variables reduce during the studying periods (such as overturn collision, off-road collision, winter, snow, turning movement, deployed airbag, child seat, no liability and old driver indicator on serious injury and fatality), while others increase (such as wet surface, ice surface, old vehicle, lane-change movement, no-airbag, partial and totally ejection, impaired driver, and male driver on serious injury and fatality) or show stable value.

Also, the MLM and the LCM are both developed in this study. This project shows that choosing uniform distribution as prior for random parameters could better increase goodness-of-fit of the MLM than using normal and lognormal distributions. In addition, the two-class LCM also shows superiority when compared to the three- and four-class models. Moreover, a careful comparison between these two best models of their kinds is also conducted, and the results indicate that the LCM works slightly better in analyzing the aforementioned dataset in this study. A series of significant contributing factors in terms of road geometric characters, traffic compositions and dynamics, driver demographic features, etc., are identified and compared with the two models. To better explain the model estimation results, pseudo-elasticity analyses of the significant factors are conducted. The results reveal that *curve*, *on grade*, *signal control*, *multiple lanes*, *pickup*, *straight*, *drug/alcohol impaired*, and *seat belt not used* can increase driver injury severity in the two models. On the other hand, *wet*, *male*, *semi*, and *young* are found to decrease driver injury outcomes. These results are not only useful for understanding the underlying risk factors of rural single-vehicle crashes under rain conditions, but can also provide meaningful references developing appropriate countermeasures, strategies, and policies to mitigate the driver injury severity of relative crashes worldwide.

6.2. Recommendations

To facilities future research, the following recommendations are made:

(1) The differences in the key elements and those with heterogeneity imply that specific strategies should be adopted in each configuration accordingly. It is worth digging deeper into each crash configuration to understand their risk factors better. More research efforts could be made, including more potential risk factors such as the vehicle velocities, passengers in the vehicle, and so on. Time and space heterogeneity could also be a further consideration.

(2) Effective management and planning, technical implementation guide of specific countermeasures, and political support and leadership are necessary and should be fulfilled together to improve traffic safety.

(3) Based on the time-varying effects and previous engineering experience, appropriate countermeasures and police recommendations could be implemented to reduce highway single-vehicle crashes.

REFERENCES

Abdel-Aty, M., 2003. Analysis of driver injury severity levels at multiple locations using ordered probit models. Journal of Safety Research 34(5), 597–603.

Abdel-Aty, M., Ekram, A.-A., Huang, H., Choi, K., 2011. A study on crashes related to visibility obstruction due to fog and smoke. Accident Analysis and Prevention 43(5), 1730–1737.

Abdel-Aty, M., Hassan, H., Ahmed, M., Al-Ghamdi, A., 2012. Real-time prediction of visibility related crashes. Transportation Research Part C 24, 288–298.

Adams, K., Drenner, R., Chumchal, M., Donato, D., 2016. Disparity between state fish consumption advisory systems for methylmercury and US Environmental Protection Agency recommendations: A case study of the south central United States. Environmental Toxicology and Chemistry 35(1), 247–251.

Afghari, A., Haque, M., Washington, S., Smyth, T., 2016. Bayesian latent class safety performance function for identifying motor vehicle crash black spots. Transportation Research Record: Journal of the Transportation Research Board 2601, 90–98.

Ahmed, M., Franke, R., Ksaibati, K., Shinstine, D., 2018. Effects of truck traffic on crash injury severity on rural highways in Wyoming using Bayesian binary logit models. Accident Analysis and Prevention 117, 106–113.

Alves, L., Fasolo, A., 2015. Not Just Another Mixed Frequency Paper (No. 400), Central Bank of Brazil, Research Department.

Amarasingha, N., Dissanayake, S., 2014. Gender differences of young drivers on injury severity outcome of highway crashes. Journal of Safety Research 49, 113 -120.

Anarkooli, A., Hosseinlou, M., 2016. Analysis of the injury severity of crashes by considering different lighting conditions on two-lane rural roads. Journal of Safety Research 56, 57–65.

Anastasopoulos, P.C., 2016. Random parameters multivariate tobit and zero-inflated count data models: Addressing unobserved and zero-state heterogeneity in accident injury-severity rate and frequency analysis. Analytic Methods in Accident Research 11, 17–32. doi:https://doi.org/10.1016/j.amar.2016.06.001

Andrey, J.C., Mills, B.E., Vandermolen, J., 2001. Weather information and road safety.

Ankargren, S., Unosson, M., Yang, Y., 2018. A mixed-frequency Bayesian vector autoregression with a steady-state prior. Department of Statistics, Uppsala University.

Barua, S., El-Basyouny, K., Islam, M., 2016. Multivariate random parameters collision count data models with spatial heterogeneity. Analytic Methods in Accident Research 9, 1–15.

Bauer, C., Wakefield, J., Rue, H., Self, S., Feng, Z., Wang, Y., 2016. Bayesian penalized spline models for the analysis of spatio - temporal count data. Statistics in medicine 35, 1848 - 1865.

Behnood, A., Mannering, F., 2017. Determinants of bicyclist injury severities in bicycle-vehicle crashes: A random parameters approach with heterogeneity in means and variances. Analytic Methods in Accident Research 16, 35–47.

Behnood, A., Mannering, F., 2016. An empirical assessment of the effects of economic recessions on pedestrian-injury crashes using mixed and latent-class models. Analytic Methods in Accident Research 12, 1–17.

Behnood, A., Mannering, F., 2015. The temporal stability of factors affecting driver-injury severities in single-vehicle crashes: some empirical evidence. Analytic Methods in Accident Research 8, 7–32.

Behnood, A., Roshandeh, A., Mannering, F., 2014. Latent class analysis of the effects of age, gender, and alcohol consumption on driver-injury severities. Analytic Methods in Accident Research 3, 56–91.

Berndt, E., Hall, B., Hall, R., Hausman, J., 1974. Estimation and inference in nonlinear structural models. Annals of Economic and Social Measurement 3(4), 653–665.

Bellos, V., Ziakopoulos, A., Yannis, G., 2019. Investigation of the effect of tourism on road crashes. Journal of Transportation Safety & Security 1–18.

Bhat, C., 2003. Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. Transportation Research Part B 37(9), 837–855.

Bhat, C.R., Born, K., Sidharthan, R., Bhat, P.C., 2014. A count data model with endogenous covariates: formulation and application to roadway crash frequency at intersections. Analytic Methods in Accident Research 1, 53–71.

Bhat, C., Dubey, S., 2014. A new estimation approach to integrate latent psychological constructs in choice modeling. Transportation Research Part B 67, 68–85.

Blackburn, T., Gaston, K., 2001. Local avian assemblages as random draws from regional pools. Ecography 24(1), 50–58.

Blazquez, C.A., Celis, M.S., 2013. A spatial and temporal analysis of child pedestrian crashes in Santiago, Chile. Accident Analysis & Prevention 50, 304–311.

Brazil, N., Kirk, D.S., 2016. Uber and metropolitan traffic fatalities in the United States. American journal of epidemiology 184, 192–198.

Brijs, T., Karlis, D., Wets, G., 2008. Studying the effect of weather conditions on daily crash counts using a discrete time-series model. Accident Analysis & Prevention 40, 1180–1190.

Buddhavarapu, P., Scott, J., Prozzi, J., 2016. Modeling unobserved heterogeneity using finite mixture random parameters for spatially correlated discrete count data. Transportation Research Part B 91, 492–510.

Bureau of Economic Analysis, 2019. Washington State Income [WWW Document]. URL https://www.bea.gov/

Bureau of Labor Statistics, 2019. Washington State Economy at a Glance [WWW Document]. URL https://www.bls.gov/eag/eag.wa.htm

Canova, F., 2011. Methods for applied macroeconomic research. Princeton university press.

Canova, F., Ciccarelli, M., Ortega, E., 2007. Similarities and convergence in G-7 cycles. Journal of Monetary economics 54, 850–878.

Cappellari, L., Jenkins, S., 2003. Multivariate probit regression using simulated maximum likelihood. The Stata Journal 3(3), 278–294.

Carriero, A., Clark, T.E., Marcellino, M., 2015. Bayesian VARs: specification choices and forecast accuracy. Journal of Applied Econometrics 30, 46–73.

Carter, C.K., Kohn, R., 1994. On Gibbs sampling for state space models. Biometrika 81, 541–553.

Chen, C., Zhang, G., Huang, H., Wang, J., Tarefder, R., 2016a. Examining driver injury severity outcomes in rural non-interstate roadway crashes using a hierarchical ordered logit model. Accident Analysis and Prevention 96, 79–87.

Chen, C., Zhang, G., Tian, Z., Bogus, S., Yang, Y., 2015. Hierarchical Bayesian random intercept model-based cross-level interaction decomposition for truck driver injury severity investigations. Accident Analysis and Prevention 85, 186–198.

Chen, C., Zhang, G., Yang, J., Milton, J., 2016b. An explanatory analysis of driver injury severity in rear-end crashes using a decision table/Naïve Bayes (DTNB) hybrid classifier. Accident Analysis and Prevention 90, 95–107.

Chen, E., Tarko, A., 2014. Modeling safety of highway work zones with random parameters and random effects models. Analytic Methods in Accident Research 1, 86–95.

Chen, F., Chen, S., 2011. Injury severities of truck drivers in single- and multi-vehicle accidents on rural highways. Accident Analysis & Prevention 43, 1677–1688. doi:https://doi.org/10.1016/j.aap.2011.03.026

Chen, F., Chen, S., Ma, X., 2018. Analysis of hourly crash likelihood using unbalanced panel data mixed logit model and real-time driving environmental big data. Journal of Safety Research 65, 153–159. doi:https://doi.org/10.1016/j.jsr.2018.02.010

Chen, F., Chen, S., Ma, X., 2016. Crash frequency modeling using real-time environmental and traffic data and unbalanced panel data models. International journal of environmental research and public health 13, 609.

Chen, F., Ma, X., Chen, S., 2014. Refined-scale panel data crash rate analysis using random-effects tobit model. Accident Analysis & Prevention 73, 323–332. doi:https://doi.org/10.1016/j.aap.2014.09.025

Cheng, W., Gill, G.S., Dasu, R., Xie, M., Jia, X., Zhou, J., 2017. Comparison of Multivariate Poisson lognormal spatial and temporal crash models to identify hot spots of intersections based on crash types. Accident Analysis & Prevention 99, 330–341. doi:https://doi.org/10.1016/j.aap.2016.11.022.

Chiou, Y.-C., Fu, C., 2015. Modeling crash frequency and severity with spatiotemporal dependence. Analytic Methods in Accident Research 5, 43–58.

Chiou, Y.-C., Fu, C., Chih-Wei, H., 2014. Incorporating spatial dependence in simultaneously modeling crash frequency and severity. Analytic Methods in Accident Research 2, 1–11.

Chow, K., Manners, S., Meuleners, L., 2016. Risk factors for killed and serious injury intersection crashes in metropolitan Perth: 2006-2015. Accident Analysis and Prevention 42(6), 1908–1915.

Coruh, E., Bilgic, A., Tortum, A., 2015. Accident analysis with aggregated data: The random parameters negative binomial panel count data model. Analytic Methods in Accident Research 7, 37–49.

Daniels, S., Brijs, T., Nuyts, E., Wets, G., 2010. Explaining variation in safety performance of roundabouts. Accident Analysis & Prevention 42, 393–402.

Del Negro, M., Schorfheide, F., 2011. Bayesian macroeconometrics.

Depaire, B., Wets, G., Vanhoof, K., 2008. Traffic accident segmentation by means of latent class clustering. Accident Analysis and Prevention 40(4), 1257–1266.

Domencich, T.A., McFadden, D., 1975. Urban travel demand, A behavioral analysis. North-Holland Publishing, Oxford.

Dong, B., Ma, X., Chen, F., Chen, S., 2018. Investigating the Differences of Single-Vehicle and Multivehicle Accident Probability Using Mixed Logit Model. Journal of Advanced Transportation 2018.

Dong, N., Huang, H., Lee, J., Gao, M., Abdel-Aty, M., 2016. Macroscopic hotspots identification: A Bayesian spatio-temporal interaction approach. Accident Analysis & Prevention 92, 256–264. doi:https://doi.org/10.1016/j.aap.2016.04.001

Donmez, B., Liu, Z., 2015. Associations of distraction involvement and age with driver injury severities. Journal of Safety Research 52, 23–28.

Duan, Y., Lv, Y., Liu, Y.-L., Wang, F.-Y., 2016. An efficient realization of deep learning for traffic data imputation. Transportation research part C: emerging technologies 72, 168–181.

Eluru, N., Bagheri, M., Miranda-Moreno, L., Fu, L., 2012. A latent class modeling approach for identifying vehicle driver injury severity factors at highway-railway crossings. Accident Analysis and Prevention 47, 119–127.

Eraker, B., Chiu, C.W., Foerster, A.T., Kim, T.B., Seoane, H.D., 2014. Bayesian mixed frequency VARs. Journal of Financial Econometrics 13, 698–721.

Eustace, D., Aylo, A., Mergia, W.Y., 2015. Crash frequency analysis of left-side merging and diverging areas on urban freeway segments–A case study of I-75 through downtown Dayton, Ohio. Transportation research part C: emerging technologies 50, 78–85.

Federal Reserve Bank of St. Louis, 2019. Washington State Economic Data [WWW Document]. URL https://fred.stlouisfed.org/categories/27331

Feng, S., Li, Z., Ci, Y., Zhang, G., 2016. Risk factors affecting fatal bus accident severity: Their impact on different types of bus drivers. Accident Analysis and Prevention 86, 29–39.

Freeman, D.G., 2007. Drunk driving legislation and traffic fatalities: new evidence on BAC 08 laws. Contemporary Economic Policy 25, 293–308.

García-Laencina, P.J., Sancho-Gómez, J.-L., Figueiras-Vidal, A.R., Verleysen, M., 2009. K nearest neighbours with mutual information for simultaneous classification and missing data imputation. Neurocomputing 72, 1483–1493.

Giannone, D., Lenza, M., Primiceri, G.E., 2015. Prior selection for vector autoregressions. Review of Economics and Statistics 97, 436–451.

Granger, C.W.J., Newbold, P., 2014. Forecasting economic time series. Academic Press.

Gray, R., Quddus, M., Evans, A., 2008. Injury severity analysis of accidents involving young male drivers in Great Britain. Journal of Safety Research 39(5), 483–495.

Greene, W., 2012. NLOGIT 5 Reference Guide. Econometric Software. Inc., Plainview, NY.

Haleem, K., Gan, A., 2015. Contributing factors of crash injury severity at public highway-railroad grade crossings in the US. Journal of Safety Research 53, 23–29.

Hapfelmeier, A., Hothorn, T., Ulm, K., Strobl, C., 2014. A new variable importance measure for random forests with missing data. Statistics and Computing 24, 21–34.

Hermans, E., Brijs, T., Stiers, T., Offermans, C., 2006. The impact of weather conditions on road safety investigated on an hourly basis. Transportation Research Board

Heydari, S., Fu, L., Miranda-Moreno, L., Jopseph, L., 2017. Using a flexible multivariate latent class approach to model correlated outcomes: a joint analysis of pedestrian and cyclist injuries. Analytic Methods in Accident Research 13, 16–27.

Holdridge, J., Shankar, V., Ulfarsson, G., 2005. The crash severity impacts of fixed roadside objects. Journal of Safety Research 36(2), 139–147.

Huth, V., Sanchez, Y., Brusque, C., 2015. Drivers' phone use at red traffic lights: A roadside observation study comparing calls and visual-manual interactions. Accident Analysis and Prevention 74, 42–48.

Jones, M.P., 1996. Indicator and stratification methods for missing explanatory variables in multiple linear regression. Journal of the American statistical association 91, 222–230.

Juselius, K., 2006. The cointegrated VAR model: methodology and applications. Oxford university press.

Karlsson, S., 2013. Forecasting with Bayesian vector autoregression, in: Handbook of Economic Forecasting. Elsevier, pp. 791–897.

Khan, G., Bill, A.R., Noyce, D.A., 2015. Exploring the feasibility of classification trees versus ordinal discrete choice models for analyzing crash severity. Transportation Research Part C: Emerging Technologies 50, 86–96.

Kim, J.-K., Kim, S., Ulfarsson, G., Porrello, L., 2007. Bicyclist injury severities in bicycle-motor vehicle accidents. Accident Analysis and Prevention 39(2), 238–251.

Kim, J.-K., Ulfarsson, G., Kim, S., Shankar, V., 2013. Driver-injury severity in single-vehicle crashes in California: a mixed logit analysis of heterogeneity due to age and gender. Accident Analysis and Prevention 50, 1073–1081.

Kim, J.-K., Ulfarsson, G., Shankar, V., Mannering, F., 2010. A note on modeling pedestrian-injury severity in motor-vehicle crashes with the mixed logit model. Accident Analysis and Prevention 42(6), 1751–1758.

Koop, G., 2003. Bayesian Econometrics. Wiley, Chichester, UK.

Korobilis, D., 2013. VAR forecasting using Bayesian variable selection. Journal of Applied Econometrics 28, 204–230.

Kweon, Y.-J., 2011. Development of crash prediction models with individual vehicular data. Transportation research part C: emerging technologies 19, 1353–1363.

Landerman, L.R., Land, K.C., Pieper, C.F., 1997. An empirical evaluation of the predictive mean matching method for imputing missing values. Sociological Methods & Research 26, 3–33.

Leigh, J.P., Waldon, H.M., 1991. Unemployment and highway fatalities. Journal of health politics, policy and law 16, 135–156.

Lee, C., Li, X., 2014. Analysis of injury severity of drivers involved in single-and two-vehicle crashes on highways in Ontario. Accident Analysis and Prevention 71, 286–295.

Li, L., Li, Y., Li, Z., 2013. Efficient missing data imputing for traffic flow by considering temporal and spatial dependence. Transportation research part C: emerging technologies 34, 108–120.

Li, Z., Chen, C., Ci, Y., Zhang, G., Wu, Q., Liu, C., Qian, Z., 2018. Examining driver injury severity in intersection-related crashes using cluster analysis and hierarchical Bayesian models. Accident Analysis and Prevention 120, 139–151.

Li, Z., Chen, X., Ci, Y., Chen, C., Zhang, G., 2019a. A hierarchical Bayesian spatiotemporal random parameters approach for alcohol/drug impaired-driving crash frequency analysis. Analytic Methods in Accident Research. doi:https://doi.org/10.1016/j.amar.2019.01.002

Li, Z., Ci, Y., Chen, C., Zhang, G., Wu, Q., Qian, Z. (Sean), Prevedouros, P.D., Ma, D.T., 2019b. Investigation of driver injury severities in rural single-vehicle crashes under rain conditions using mixed logit and latent class models. Accident Analysis & Prevention 124, 219–229. doi:https://doi.org/10.1016/j.aap.2018.12.020

Liu, C., Chen, C.-L., Utter, D., 2005. Trend and pattern analysis of highway crash fatality by month and day.

Liu, C., Sharma, A., 2018. Using the multivariate spatio-temporal Bayesian model to analyze traffic crashes by severity. Analytic methods in accident research 17, 14–31.

Liu, J., Khattak, A., Richards, S., Nambisan, S., 2015. What are the differences in driver injury outcomes at highway-rail grade crossings? Untangling the role of pre-crash behaviors. Accident Analysis and Prevention 85, 157–169.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transportation Research Part A: Policy and Practice 44, 291–305.

Lütkepohl, H., 2005. New introduction to multiple time series analysis. Springer Science & Business Media.

Lynn, C., Schreiner, C., Campbell, R., 2002. Reducing fog-related crashes on the Afton and Fancy Gap Mountain sections of I-64 and I-77 in Virginia. Virginia Transportation Research Center, Charlottesville, 2002.

Ma, X., Chen, F., Chen, S., 2015. Modeling crash rates for a mountainous highway by using refined-scale panel data. Transportation research record 2515, 10–16.

Ma, X., Chen, S., Chen, F., 2017. Multivariate space-time modeling of crash frequencies by injury severity levels. Analytic Methods in Accident Research 15, 29–40.

Ma, X., Tao, Z., Wang, Y., Yu, H., Wang, Y., 2015. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transportation Research Part C: Emerging Technologies 54, 187–197.

Males, M.A., 2009. Poverty as a determinant of young drivers' fatal crash risks. Journal of safety research 40, 443–448.

Malyshkina, N. V, Mannering, F.L., 2010. Zero-state Markov switching count-data models: An empirical assessment. Accident Analysis & Prevention 42, 122–130.

Malyshkina, N., Mannering, F., Tarko, A., 2009. Markov switching negative binomial models: An application to vehicle accident frequencies. Accident Analysis and Prevention 41(2), 217–226.

Malyshkina, N. V, Mannering, F.L., 2009. Markov switching multinomial logit model: an application to accident-injury severities. Accident Analysis & Prevention 41, 829–838.

Mannering, F., 2018. Temporal instability and the analysis of highway accident data. Analytic Methods in Accident Research 17, 1–13.

Mannering, F., Bhat, C., 2014. Analytic methods in accident research: Methodological frontier and future directions. Analytic Methods in Accident Research 1, 1–22.

Mannering, F., Shankar, V., Bhat, C., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Analytic Methods in Accident Research 11, 1–16.

McCann, K., Fontaine, M., 2016. Assessing driver speed choice in fog with the use of visibility data from road weather information systems. Transportation Research Record: Journal of the Transportation Research Board 2551, 90–99.

Miaou, S.-P., 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. Accident Analysis & Prevention 26, 471–482.

Milton, J., Shankar, V., Mannering, F., 2008. Highway accident severities and the mixed logit model: an exploratory empirical analysis. Accident Analysis and Prevention 40(1), 260–266.

Morgan, A., Mannering, F., 2011. The effects of road-surface conditions, age, and gender on driver-injury severities. Accident Analysis and Prevention 43(5), 1852–1863.

Mothafer, G.I.M.A., Yamamoto, T., Shankar, V.N., 2016. Evaluating crash type covariances and roadway geometric marginal effects using the multivariate Poisson gamma mixture model. Analytic Methods in Accident Research 9, 16–26. doi:https://doi.org/10.1016/j.amar.2015.11.001

NHTSA, 2016. 2015 motor vehicle crashes: overview. Traffic Safety Facts Research Note 2016, 1–9.

Norros, I., Kuusela, P., Innamaa, S., Pilli-Sihvola, E., Rajamäki, R., 2016. The Palm distribution of traffic conditions and its application to accident risk assessment. Analytic Methods in Accident Research 12, 48–65.

Office of the Washington State Climatologist, 2019. Climate Analysis Toolbox - Time Series [WWW Document]. URL https://cefa.dri.edu/Westmap/Westmap_home.php?page=timeseries.php

Ohn-Bar, E., Tawari, A., Martin, S., Trivedi, M., 2015. On surveillance for safety critical events: In-vehicle video networks for predictive driver assistance systems. Computer Vision and Image Understanding 134, 130–140.

Oregon Department of Transportation, 2018. ODOT Statewide crash data system 2017 motor vehicle traffic crash analysis and code manual

Osman, M., Paleti, R., Mishra, S., Golias, M., 2016. Analysis of injury severity of large truck crashes in work zones. Accident Analysis and Prevention 97, 261–273.

Page, Y., 2001. A statistical model to compare road mortality in OECD countries. Accident Analysis & Prevention 33, 371–385.

Peng, Y., Abdel-Aty, M., Shi, Q., Yu, R., 2017. Assessing the impact of reduced visibility on traffic crash risk using microscopic data and surrogate safety measures. Transportation research part C: emerging technologies 74, 295–305.

Pour-Rouholamin, M., Zhou, H., 2016. Investigating the risk factors associated with pedestrian injury severity in Illinois. Journal of Safety Research 57, 9–17.

Prado, R., West, M., 2010. Time series: modeling, computation, and inference. Chapman and Hall/CRC.

Quddus, M., Noland, R., Chin, H., 2002. An analysis of motorcycle injury and vehicle damage severity using ordered probit models. Journal of Safety Research 33(4), 445–462.

Russo, B., Savolainen, P., Schneider IV, W., Anastasopoulos, P., 2014. Comparison of factors affecting injury severity in angle collisions by fault status using a random parameters bivariate ordered probit model. Analytic Methods in Accident Research 2, 21–29.

Schorfheide, F., Song, D., 2015. Real-time forecasting with a mixed-frequency VAR. Journal of Business & Economic Statistics 33, 366–380.

Schorfheide, F., Song, D., Yaron, A., 2014. Identifying long-run risks: A bayesian mixed-frequency approach. National Bureau of Economic Research.

Scuffham, P.A., Langley, J.D., 2002. A model of traffic crashes in New Zealand. Accident Analysis & Prevention 34, 673–687.

Seraneeprakarn, P., Huang, S., Shankar, V., Mannering, F., Venkataraman, N., Milton, J., 2017. Occupant injury severities in hybrid-vehicle involved crashes: A random parameters approach with heterogeneity in means and variances. Analytic Methods in Accident Research 15, 41–55.

Shankar, V., Mannering, F., Barfield, W., 1995. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. Accident Analysis & Prevention 27, 371–389.

Shaheed, M., Gkritza, K., 2014. A latent class analysis of single-vehicle motorcycle crash severity outcomes. Analytic Methods in Accident Research 2, 30–38.

Shaheed, M., Gkritza, K., Carriquiry, A., Hallmark, S., 2016. Analysis of occupant injury severity in winter weather crashes: A fully Bayesian multivariate approach. Analytic Methods in Accident Research 11, 33–47.

Statisticat, L.L.C., 2015. Laplacesdemon: complete environment for Bayesian inference. R package version 15.

Tan, H., Feng, G., Feng, J., Wang, W., Zhang, Y.-J., Li, F., 2013. A tensor-based method for missing traffic data completion. Transportation Research Part C: Emerging Technologies 28, 15–27.

Tang, J., Zhang, G., Wang, Y., Wang, H., Liu, F., 2015. A hybrid approach to integrate fuzzy C-means based imputation method with genetic algorithm for missing traffic volume data estimation. Transportation Research Part C: Emerging Technologies 51, 29–40.

Tarel, J.-P., Hautiere, N., Caraffa, L., Cord, A., Halmaoui, H., Gruyer, D., 2012. Vision enhancement in homogeneous and heterogeneous fog. IEEE Intelligent Transportation Systems Magazine 4(2), 6–20.

Train, K., 2000. Halton sequences for mixed logit. Department of Economics, University of California, Berkeley.

Traynor, T.L., 2009. The impact of state level behavioral regulations on traffic fatality rates. Journal of safety research 40, 421–426.

Uc, E., Rizzo, M., Anderson, S., Dastrup, E., Sparks, J., Dawson, J., 2009. Driving under low-contrast visibility conditions in Parkinson disease. Neurology 73, 1103–1110.

Usman, T., Fu, L., Miranda-Moreno, L.F., 2012. A disaggregate model for quantifying the safety effects of winter road maintenance activities at an operational level. Accident Analysis & Prevention 48, 368–378.

Vahdati, S.S., Ala, A., Falaki, R., Fahimi, R., Safapour, A., 2016. Demographic Study of Maxillofacial Injury in Multiple Trauma Patients. Emerg Med (Los Angel) 6, 2.

Valent, F., Schiava, F., Savonitto, C., Gallo, T., Brusaferro, S., Barbone, F., 2002. Risk factors for fatal road traffic accidents in Udine, Italy. Accident Analysis and Prevention 34(1), 71–84.

Van Lint, J.W.C., Hoogendoorn, S.P., van Zuylen, H.J., 2005. Accurate freeway travel time prediction with state-space neural networks under missing data. Transportation Research Part C: Emerging Technologies 13, 347–369.

Venkataraman, N., Shankar, V., Ulfarsson, G.F., Deptuch, D., 2014. A heterogeneity-in-means count model for evaluating the effects of interchange type on heterogeneous influences of interstate geometrics on crash frequencies. Analytic methods in accident research 2, 12–20.

Wang, X., Abdel-Aty, M., 2008. Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models. Accident Analysis and Prevention 40(5), 1674–1682.

Washington, S., Karlaftis, M., Mannering, F., 2011. Statistical and econometric methods for transportation data analysis. Chapman and Hall/CRC.

Washington State Departement of Transportation, 2010. Guide for interpreting short duration traffic count reports.

Washington State Department of Transportation, 2018. State Highway Log Planning Report 2017.

Washington State Department of Transportation, 2016. 2015 Annual Collision Summary.

Weakliem, D., 1999. A critique of the Bayesian information criterion for model selection. Sociological Methods and Research 27(3), 359–397.

Weiss, H., Kaplan, S., Prato, C., 2014. Analysis of factors associated with injury severity in crashes involving young New Zealand drivers. Accident Analysis and Prevention 65, 142–155.

Wu, Q., Chen, F., Zhang, G., Liu, X., Wang, H., Bogus, S., 2014. Mixed logit model-based driver injury severity investigations in single-and multi-vehicle crashes on rural two-lane highways. Accident Analysis and Prevention 72, 105–115.

Wu, Q., Ci, Y., Chen, C., Zhang, G., 2018. Examining driver injury severity in single-vehicle crashes: a two-step study using cluster analysis and mixed logit model. 97th Annual Meeting of the Transportation Research Board, CD-ROM, Washington, D.C.

Wu, Q., Zhang, G., Chen, C., Tarefder, R., Wang, H., Wei, H., 2016. Heterogeneous impacts of gender-interpreted contributing factors on driver injury severities in single-vehicle rollover crashes. Accident Analysis and Prevention 94, 28–34.

Wu, Y., Abdel-Aty, M., Lee, J., 2018. Crash risk analysis during fog conditions using real-time traffic data. Accident Analysis and Prevention 114, 4-11.

Xiong, Y., Mannering, F., 2013. The heterogeneous effects of guardian supervision on adolescent driver-injury severities: A finite-mixture random-parameters approach. Transportation Research Part B 49, 39–54.

Xiong, Y., Tobias, J., Mannering, F., 2014. The analysis of vehicle crash injury-severity data: A Markov switching approach with road-segment heterogeneity. Transportation Research Part B 67, 109–128.

Xu, C., Wang, W., Liu, P., Guo, R., Li, Z., 2014. Using the Bayesian updating approach to improve the spatial and temporal transferability of real-time crash risk prediction models. Transportation research part C: emerging technologies 38, 167–176.

Yamaoka, K., Nakagawa, T., Uno, T., 1978. Application of Akaike's information criterion (AIC) in the evaluation of linear pharmacokinetic equations. Journal of Pharmacokinetics and Pharmacodynamics 6(2), 165–175.

Yang, K., Wang, X., Yu, R., 2018. A Bayesian dynamic updating approach for urban expressway real-time crash risk evaluation. Transportation Research Part C: Emerging Technologies 96, 192–207. doi:https://doi.org/10.1016/j.trc.2018.09.020

Yannis, G., Papadimitriou, E., Folla, K., 2014. Effect of GDP changes on road traffic fatalities. Safety Science 63, 42–49. doi:https://doi.org/10.1016/j.ssci.2013.10.017

Yasmin, S., Eluru, N., 2018. A mixed grouped response ordered logit count model framework. Analytic Methods in Accident Research 19, 49–61. doi:https://doi.org/10.1016/j.amar.2018.06.002

Yasmin, S., Eluru, N., Bhat, C., Tay, R., 2014. A latent segmentation based generalized ordered logit model to examine factors influencing driver injury severity. Analytic Methods in Accident Research 1, 23–38.

Ye, F., Lord, D., 2014. Comparing three commonly used crash severity models on sample size requirements: Multinomial logit, ordered probit and mixed logit models. Analytic Methods in Accident Research 1, 72–85.

Yu, H., Li, Z., Zhang, G., Liu, P., 2019. A latent class approach for driver injury severity analysis in highway single vehicle crash considering unobserved heterogeneity and temporal influence. Analytic Methods in Accident Research 24, 100110.

Yu, H., Liu, P., Chen, J., Wang, H., 2014. Comparative analysis of the spatial analysis methods for hotspot identification. Accident Analysis & Prevention 66, 80-88.

Yu, H., Yuan, R., Li, Z., Zhang, G., Ma, D., 2020. Investigate Factors Affecting Driver Injury Severity in Snow-Related Rural Single-Vehicle Crashes. Accident Analysis & Prevention, In Press.

Yu, R., Abdel-Aty, M., Ahmed, M., 2013. Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors. Accident Analysis & Prevention 50, 371–376.

Yu, R., Xiong, Y., Abdel-Aty, M., 2015. A correlated random parameter approach to investigate the effects of weather conditions on crash risk for a mountainous freeway. Transportation research part C: emerging technologies 50, 68–77.

Yu, R., Wang, X., Abdel-Aty, M., 2017. A hybrid latent class analysis modeling approach to analyze urban expressway crash risk. Accident Analysis and Prevention 101, 37–43.

Zeng, Q., Huang, H., Pei, X., Wong, S.C., 2016. Modeling nonlinear relationship between crash frequency by severity and contributing factors by neural networks. Analytic methods in accident research 10, 12–25.

Zeng, Q., Sun, J., Wen, H., 2017. Bayesian hierarchical modeling monthly crash counts on freeway segments with temporal correlation. Journal of Advanced Transportation 2017.

Zhang, Q., Yu, H., Li, Z., Zhang, G., Ma, D., 2020. Assessing Potential Likelihood and Impacts of Landslides on Transportation Network Vulnerability. Transportation Research Part D: Transport and Environment 82, 102304.

Zhong, M., Lingras, P., Sharma, S., 2004. Estimation of missing traffic counts using factor, genetic, neural, and regression techniques. Transportation Research Part C: Emerging Technologies 12, 139–166.

Zhu, X., Srinivasan, S., 2011. A comprehensive analysis of factors influencing the injury severity of large-truck crashes. Accident Analysis and Prevention 43(1), 49–57.

Zou, Y., Zhang, Y., Lord, D., 2014. Analyzing different functional forms of the varying weight parameter for finite mixture of negative binomial regression models. Analytic Methods in Accident Research 1, 39–52.