2012-01-01

Modeling Large-Truck Involved Crashes: An Econometric Modeling Framework

Mouyid Islam
University of Texas at El Paso, mbislam@miners.utep.edu

MODELING LARGE-TRUCK INVOLVED CRASHES: AN ECONOMETRIC MODELING FRAMEWORK

MOUYID ISLAM

Department of Civil Engineering

APPROVED:

Salvador Hernandez, Ph.D., Chair

Kelvin Cheu, Ph.D

Carlos M. Chang-Albitres, Ph.D

Thomas M. Fullerton, Ph.D

Fred L. Mannering, Ph.D

Benjamin C. Flores, Ph.D.
Dean of the Graduate School

Dedication

To my

MOTHER and FATHER

MODELING LARGE-TRUCK INVOLVED CRASHES: AN ECONOMETRIC MODELING FRAMEWORK

by

MOUYID ISLAM, M.Sc

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Department of Civil Engineering

THE UNIVERSITY OF TEXAS AT EL PASO

December 2012

Acknowledgements

I am deeply grateful to almighty Allah who gave me the strength and patience to successfully complete my doctoral study at The University of Texas at El Paso (UTEP) from 2010 to 2012 fall.

I express sincere gratitude to my advisor, Dr. Salvador Hernandez, who helped me, gathering the crash data, devising research plan, and showing the importance of related analysis in the dissertation. His continued guidance in writing papers for national conferences and transportation journals from the dissertation helped me understand the professional and academic writing style in this area. I am also very grateful to all of my committee members who assisted me to improve the quality of my dissertation from the proposal to the final defense. I am thankful to Dr. Kelvin Cheu for his inquisitive queries related to crash data and his support in attending conferences such as TexITE and TRB to present my research work. I also thank to Dr. Carlos Chang for his constructive critiques on the chapters, which in turn helped me to present my research findings based on econometric modeling from the engineering stand point. My other committee members having expertise in econometric modeling in traffic safety area, Professor Fred Mannering and Professor Tom Fullerton helped me check my model results and explained the complex details of model estimation, and strength of modeling techniques. I appreciate Professor Mannering's sincere cooperation being the external in this committee.

I also extend my sincere thanks to all BIGLab members, particularly Alicia Romo for helping me with the review and references of my chapters and encouraging me through paper submission process to conferences and journals. I specially thank to all ITE UTEP chapter members and my officers – Alicia Romo, Edwin Varela, Lorenzo Cornejo, and David Salgado. I really enjoyed a wonderful time with BIGLab activities and ITE UTEP chapter that Professor Cheu has been continuing and promoting. I

sincerely thank to graduate school and doctoral writing tutors, particularly Marianna Hendricks for her time and sincere evaluation on my technical writing skills. I also thank our civil engineering administrative officers and Student Government Association for their support in paper works which was very important in academic progress particularly our RA and TA assignments and conferences attendance.

Finally, I am sincerely grateful to my mother and two younger sisters who have provided unwavering and unconditional support while completing my doctoral degree at UTEP. It would not be possible without their mental and financial support during my study at UTEP. I also thank my father, who is no longer in this world, for teaching me to endure through the difficult times of graduate and doctoral study and to work hard. My family truly strengthened me to achieve this degree setting up a milestone for my next generations. I also acknowledge my previous advisors – Professor Andrew Tarko at Purdue University, late Professor Yordphol Tanaboriboon at Asian Institute of Technology, and Professor Md. Mazharul Hoque at Bangladesh University of Engineering and Technology for their sincere guidance and shaping an inquisitive mind in me for knowledge to reach up to this mark and beyond.

Abstract

This research aims to analyze large-truck involved crashes (i.e., having a gross vehicle weight rating more than 10,000 pounds) through the application of advanced econometric modeling techniques—namely, random parameter models (i.e., tobit regression, mixed logit). To achieve this, various national and state specific data sets are analyzed with the goal to provide an improved understanding of the complex interactions of contributing factors (e.g., factors related to drivers and occupants, vehicle, and road-environment) influencing large-truck crashes. Additionally, the modeling techniques considered in this research account for possible unobserved effects related to the data. The aforementioned econometric techniques provide an analytical foundation for exploring the contributing factors leading to large-truck involved crashes.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Significance and motivation of the study

As the national economy continues to recover, the volume of large trucks (i.e., having a gross vehicle weight rating of more than 10,000 pounds) present on the nation's highway system will also experience slow but consistent growth. This increased growth in large truck volume poses many challenges for transportation organizations that operate, maintain, and construct the transportation system (e.g., through increased loads on roadways which may lead to increased road deterioration rates creating safety hazards). Additionally to roadway users, through problems stemming from passing sight distance due to trucks size and height (Douglas, 2003).

Recent statistical data have shown that large trucks have been responsible for more fatalities in the United States (US) than passenger vehicles based on the number of registered vehicles and vehicle-miles traveled (VMT) (FHWA, 2010; NHTSA, 2008). For example, large trucks accounted for roughly four percent of registered vehicles and about eight percent of VMT in 2008, but eleven percent of motor vehicle involved crash deaths in 2008 were due to large trucks (FHWA, 2010). To further illustrate the gravity of large-truck involved crashes, comparative statistics on the number of passenger vehicles and large trucks involved in fatal crashes over the period from 1999 to 2008 indicates that the rate for large trucks involved in fatalities dropped from 2.43 to 1.79 per 100 million VMT, whereas, that for passenger vehicles dropped from 1.94 to 1.45. However, the numbers still show the gravity of the problem concerning large trucks involved in fatal crashes over this period of time. Although the trend slopes downwards (possibly due to advancements in safety technologies and some combination of increased fuel prices and economic factors), the numbers are still concerning especially given the percentage of trucks on the nation's highways.

While fatalities are a major aftermath of large-truck involved crashes, the societal effects and cost associated with the resulting crashes are remarkably high—for example, expenses related to loss of life, medical attention, and insurance, and short term and long term physical and emotional effects (Miller, 1993). For instance, the estimated cost of police-reported crashes involving large trucks (GVWR higher than 10,000 pounds) averaged $91,112 based on 2005 dollars (Zaloshnja and Miller, 2006). This study also estimated the average cost per fatality, non-fatality and property-damage-only crash as $3,604,518, $195,258, and $15,114, respectively. An earlier study (Zaloshnja and Miller, 2004) estimated the cost associated with multiple combination trucks having the highest cost of $88,483 per crash based on 2000 dollars. Further, the crash costs based on 2000 dollars per 1000 truck miles were estimated at $157 for single unit trucks, $131 for single combination trucks, and $63 for multiple combinations (Zaloshnja and Miller, 2004). The above-mentioned costs illustrate the potential monetary impacts that these crashes have on society.

Moreover, large-truck involved crashes greatly influence the level of injury severity experienced by those involved (e.g., incapacitating, non-incapacitating, etc…) (Chang and Mannering, 1999). As such, these types of crashes are gnarring increased public and media attention as well increased interest from academia, transportation safety professionals, and the trucking industry. Consequently, large trucks drive the national economy through daily freight movements and would not be going away anytime soon. Hence, the aim of this work is to develop and present a methodology to better understand the factors influencing large-truck involved crashes.

1.2 Research objectives

The primary objective of this dissertation is to analyze the injury severities of large-truck involved crashes through advanced econometric modeling approaches, namely discrete outcome models.

More specifically, the random parameters logit model also referred to as the mixed logit model (Train, 2003). The use of this modeling approach is motivated by its ability to account for "unobserved heterogeneity" (i.e., unobserved factors that may vary across observations) in the police reported crash databases used in this dissertation. In addition and for completeness, in this dissertation a random parameters tobit regression model is also estimated to help explain large-truck involved fatalities rates. The proposed modeling approaches will provided additional insights into the contributing factors of large-truck involved crashes and provided decision makers with tools to better mitigate the injury severity that users of the transportation system may sustain. In addition, the developed models will provide the trucking industry with factors that can help improve driver awareness and preparedness in dealing with the day-to-day tasks of moving freight on the nation's highways. To our knowledge no comprehensive study has been performed for large-truck involved crashes utilizing the aforementioned econometric models and that study explicitly national highways and border region demographics. The specific problems addressed to achieve the objective this dissertation are:

(i) Develop research framework for analyzing large-truck involved crash data. The primary focus of this is to establish the necessary mechanisms to identify and to fuse current crash related databases for use in the modeling framework. Here, both national and state specific data are considered.

(ii) Estimate fatality rates of large-truck involved crashes. The main focus of this is to capture the factors influencing fatality rates for fatalities per million truck-miles traveled and fatalities per ton-miles of freight. These factors will then be used as starting points for injury severity analyses to follow.

(iii) Estimate single- and multi-vehicle large-truck involved crashes utilizing the NASS-GES crash database maintained by the National Highway Traffic Safety Administration (NHTSA). The primary focus here is to capture the factors influencing large-truck

involved crashes using a national crash data sample. Again, these factors will serve as starting points for the state specific crash data.

(iv) Estimate a holistic model of injury outcomes of crashes involving heavy vehicles on Texas highways using Texas State specific data. Here, the main focus is to model large-truck involved crashes in the State of Texas.

(v) Estimate split models for urban and rural areas using Texas State specific data. The primary focus here is to study the effects population density on large-truck involved crashes.

(vi) Estimate split models for time of day for urban areas using Texas State specific data. Here, the main focus is to study the effects of AM, PM and offpeak traffic periods on large-truck involved crashes.

(vii) Develop spatial and temporal transferability of severity outcome models utilizing Texas data for specific cities. The primary focus is to develop a model that can be used to assess large-truck involved crash severity outcomes for various cities in the State of Texas.

1.3 Contribution to large truck safety profession

To better understand the safety impacts related to increased large truck traffic on the nation's highway system, tools need to be developed that can aid transportation safety professionals as well as trucking industry operations managers in the avoidance and mitigation (i.e., the development of countermeasures) of large-truck involved crashes. With this in mind, this study aims to add to the current literature by proposing an advanced methodological approach that takes into account the highest injury rate (i.e., fatalities per million truck-miles and fatalities per ton-mile of freight), discrete nature of severity (i.e., fatal, incapacitating, non-incapacitating, possible injury, and property-damage-only) for single and multi-vehicle collisions all over the US and for rural and urban settings and time of the day in

urban areas in the state of Texas. The research findings certainly benefit the stakeholders particularly the general public, trucking companies, decision makers, and highway engineers to devise proper countermeasures to reduce likelihood of the injury severities and associated societal cost. The flexibility of time and space transferability will be a great contribution to the research of injury severity for large-truck involved crashes.

1.4 Organization of the dissertation

This dissertation consists of the following chapters with some research problems statement, objectives, literature review, data and research framework, some preliminary results and interpretations of results.

Chapter 1 entitled as Introduction consists of significance and motivation of the study, problem statement and objective of the study, and contribution to large truck safety profession.

Chapter 2 entitled as Data and Research Framework highlighting data sources and data processing for sample, research hypothesis, and overall research framework and associated tasks for this study.

Chapter 3 entitled as An Empirical Analysis of Fatality Rates for Large-truck involved Crashes on Interstate Highways mainly describes fatality rate regression models highlighting fatalities per million truck-miles, fatalities per ton-miles of freight, and their marginal effects. This chapter analyzes exposure based rates of the maximum level of injury severity (i.e., fatality rates) of large-truck involved crashes in the US interests systems.

Chapter 4 entitled as Single- and Multi-vehicle Large Trucks Involved Collisions: An Exploratory Injury Analysis describes severity models on US interstates. This analysis considers the significance of number of vehicles involved in crashes affecting injury severity rather than looking at the

injury severity with a combined dataset. This chapter also introduces the discrete outcome severity models on a nomial scale (i.e., random parameter logit model).

Chapter 5 entitled as Modeling Injury Outcomes of Crashes involving Heavy Vehicles on Texas Highways describes a mixed logit model for border a state such as Texas. This chapter shifts the research focus from national level data to state level data, specifically Texas. Additionally, this chapter deals with detailed crash database in Texas and utilizes the mixed logit modeling approach.

Chapter 6 entitled as Modeling Characteristics of Rural and Urban Injury Severity of Heavy Vehicle involved Crashes in Texas. This chapter deals injury analysis with rural and urban areas in Texas utilizing mixed logit model.

Chapter 7 entitled as Modeling Severity Models for Time of Day in Urban Texas Interstates deals with model development for different time period of the day – AM, PM, and off-peak period. This chapter deals with role of time of day on large-truck involved crashes.

Chapter 8 entitled as Spatial and Temporal Transferability of Severity Models and its Validation deals with model development with different state and time period, conduct the statistical test for spatial and temporal transferability. This chapter illustrates the applicability of severity models in safety planning. The evolution process of spatial and temporal transferability was demonstrated with estimated severity models utilizing the two geographic locations and time periods.

Chapter 9 entitled as Conclusion and Future Works deals with major findings of each model, insights of contributing factors for injury severity, the contribution of this research, and recommendations for future study.

Chapter 10 which is appended in Appendix – A, is entitled an as Large-truck involved Crashes: An Exploratory Injury Severity Analysis. This is an application exercise utilizing the random parameters ordered probit model.

Chapter 2: Data and Research Framework

2.1 Background

This chapter introduces the data utilized in this dissertation. First, national crash datasets are utilized, namely the Fatality Analysis Reporting System (FARS) and National Automotive Sampling System General Estimates System (NASS-GES). Second, In addition to the national datasets, the Texas Crash Records Information System (CRIS) dataset is also utilized due to its comprehensive and detailed format. To analyze the injury severities of large-truck involved crashes, the national and state crash datasets are split differently based on manner of collisions (i.e., single- and multi-vehicle collisions), location specific collisions (i.e., rural and urban areas), time period specific collisions (i.e., AM, PM, and off peak time period). The following sections describe the data processing and research framework in greater detail.

2.2 Data processing

Data processing is an important step in the estimation of any statistical model. As part of this dissertation, data was extracted and fused by appropriate linking variables. For example, the data was extracted from the crash databases by filtering for only crashes involving large trucks (GVWR over 10,000 pounds) and on the interstate system over a period time (dependent on the data). Then, the maximum level of injury severity recorded in the vehicle or person datasets were aggregated to represent a crash based on the individual vehicle or involved people for these observations as shown in Table 2.1. Accordingly, each observation in the sample is a crash representing the maximum level of injury of the occupants (either drivers or passengers), involving at least one large truck in the interstate system. The crash dataset was then linked to the vehicle and person datasets through appropriate linking variables, namely the crash number, and the fused vehicle and person datasets were further linked through the

vehicle and crash number – crash IDs. The linking process of three different datasets is illustrated in Figure 2.1.

Table 2.1: Injury Severity in the converted Crash Dataset

Person Injury Record	**Original Source Data**	**Converted Value**	**Remark**
Yes	Fatal	Fatal	Fatal category includes Fatal and Died prior to crash occurrence
	Died Prior		
	Incapacitating	Incapacitating	None
	Non-incapacitating	Non-incapacitating	None
	Possible	Possible	Possible category is the combination of Possible, Unknown Injury Severity , and Unknown if Injured categories
	Injured, Unknown Injury Severity		
	Unknown if Injured		
No	N/A	No injury	Absence of the injury record implies no injury and Property Damage crash was added in this category

While fusing vehicle and person datasets with the vehicle and person IDs, multiple observations corresponding to same crash ID are obtained. Then, crash dataset was linked with previously fused vehicle and person dataset with crash IDs. Later on, the duplicate observations were eliminated through the maximum severity of crash by comparing crash dataset with person dataset.

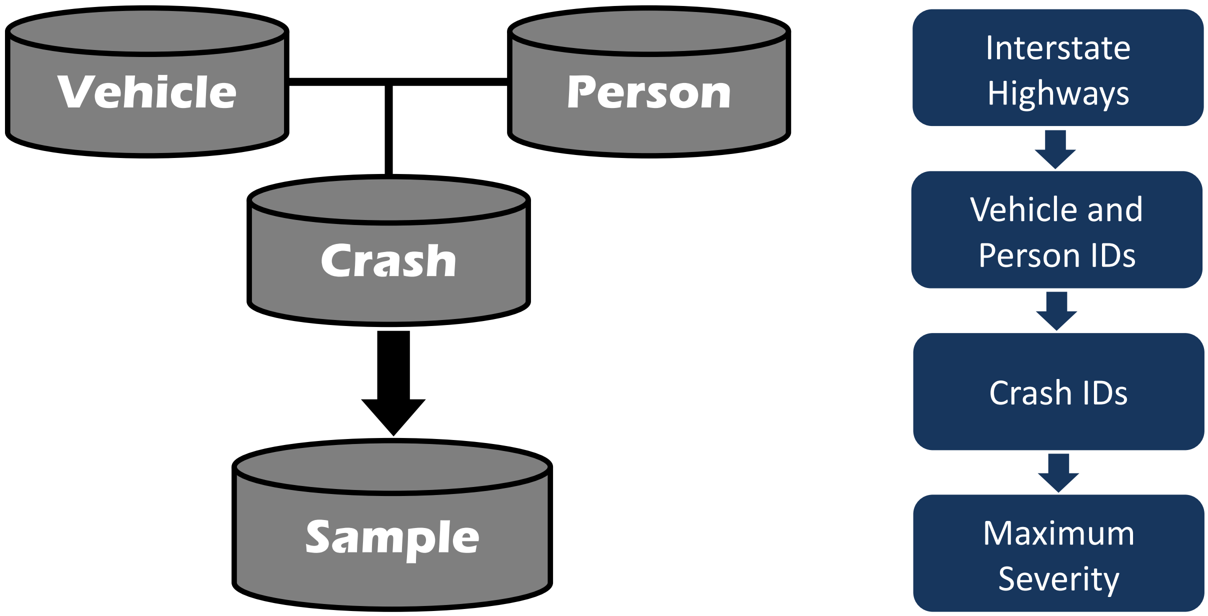

Figure 2.1: Data Process – Fusion of Datasets (left) and Criteria of Fusion (right)

2.2.1 Data processing for fatality rates in US Interstates

To estimate the fatality rates of large-truck involved crashes crash data was collected from the Fatality Analysis Reporting Systems (FARS) from 2005 to 2008. FARS is a nation-wide crash census system, where a set of files have been built documenting all qualifying fatal crashes that occurred within all the states in the US. The observation in the model is a fatal crash (A variable – *Fatals* includes the total number of fatalities in a fatal collision reported in the FARS database system) involving a motor vehicle, where at least a large truck is involved in the fatal collision traveling on U.S. interstate system resulting in a fatal (or fatalities) within 30 days of the collision. Annual average daily traffic (AADT) is not available in the dataset, and hence not considered in the estimation of the fatality rates.

The ton-miles of freight data from 2005 to 2007 was collected from the Bureau of Transportation Statistics special tabulation (BTS/RITA, 2010), whereas, truck-miles traveled data from 2005 to 2008 was collected from FHWA travel reports (FHWA, 2009) and secondary estimation procedures includes use of State supplied data. Since the crash data was limited to the U.S. interstate system, data for the truck-miles traveled and ton-miles of freight models are limited to the U.S. interstate system.

2.2.2 Data Processing for maximum severity in US interstates

To model maximum injury severity at a nation level, a sample of 8,291 data observations was processed from a national crash database known as NASS-GES crash database over a period of four years. This processed dataset was utilized in modeling random parameters ordered probit model (see Appendix for this work) where each observation is a crash representing the most severely injured occupants (i.e., maximum severity) involving at least a large truck on the interstate system from 2005 to 2008. This large-truck involved data sample (i.e., 8,291 observations) was extracted from the GES crash dataset with an average of 56,970 crashes (i.e., total truck and non-truck involved crashes) reported each year over time period from 2005 to 2008. The crash dataset was fused to vehicle and person dataset through the appropriate linking variable and crash number, while vehicle and person datasets were linked through vehicle and crash number using the Statistical Analysis System (SAS) (SAS, 2011).

2.2.3 Data Processing for maximum severity in Texas interstates

State specific data was also used in this dissertation and was collected from the Texas Peace Officers' Crash Reports, commonly known as the Crash Record Information System (CRIS) database. To investigate human, vehicle, and road-environmental factors, a sample of 20,495 data observations were extracted from the CRIS database by filtering crashes involving large trucks on the interstate system over a period of five years from 2006 to 2010. Each observation in the sample is a crash representing the maximum level of injury of the drivers, involving at least one large truck in the interstate system. The crash dataset was linked to the vehicle and person datasets through appropriate linking variables, namely crash number, and the vehicle and person datasets were linked through the vehicle and crash number using the Statistical Analysis System (SAS) (SAS, 2011).

2.3 Research hypothesis

In order to reach a final model, hypothesis tests are conducted. The hypothesis testing provides a mechanism to determine if a particular hypothesis is correct or not. With regards to the work presented in this dissertation, the log likelihood ratio test will be conducted to justify the presence of unobserved heterogeneity, model splitting (e.g., rural and urban from a holistic model) and transferability (i.e., spatial and temporal). The hypothesis set-up for the all econometric modeling frameworks under the scope of this dissertation will follow a similar structure:

Null hypothesis – H_0: no heterogeneity present in the model developed from the data sample.

Alternate hypothesis – H_A: heterogeneity present in the model developed from the data sample.

$$\chi^2 = -2[LL_{FIX}(\beta^{FIX}) - LL_{RAN}(\beta^{RAN})] \tag{2.1}$$

where,

$$LL_{FIX}(\beta^{FIX})$$
$$= loglikelihood\ at\ convergence\ of\ fixed\ parameter\ model\ with\ degree\ of\ freedom, n_{FIX}$$
$$LL_{RAN}(\beta^{RAN})$$
$$= loglikelihood\ at\ convergence\ of\ random\ parameter\ model\ with\ degree\ of\ freedom, n_{RAN}$$

The test statistics is χ^2 distributed with difference in degrees of freedom between two models which provides the confidence interval to reject or accept the null hypothesis (Washington et al., 2011). In addition, fixed and random parameter models, the combined and separate models will also be tested with log-likelihood ratio.

2.4 Research framework

This section illustrates the order of steps followed to address the research presented in this dissertation. The first task involves the research motivation, objectives, significance, and organization of the dissertation. The second task involves some recent and fundamental studies related to this dissertation. That is, only previous studies that shared similar methodological approaches and similar research problems were considered. Figure 2.2 outlines the research framework and Figure 2.3 simplifies the flowchart in Figure 2.2 by demonstrating the data used for the modeling frameworks and application aspect of the developed models.

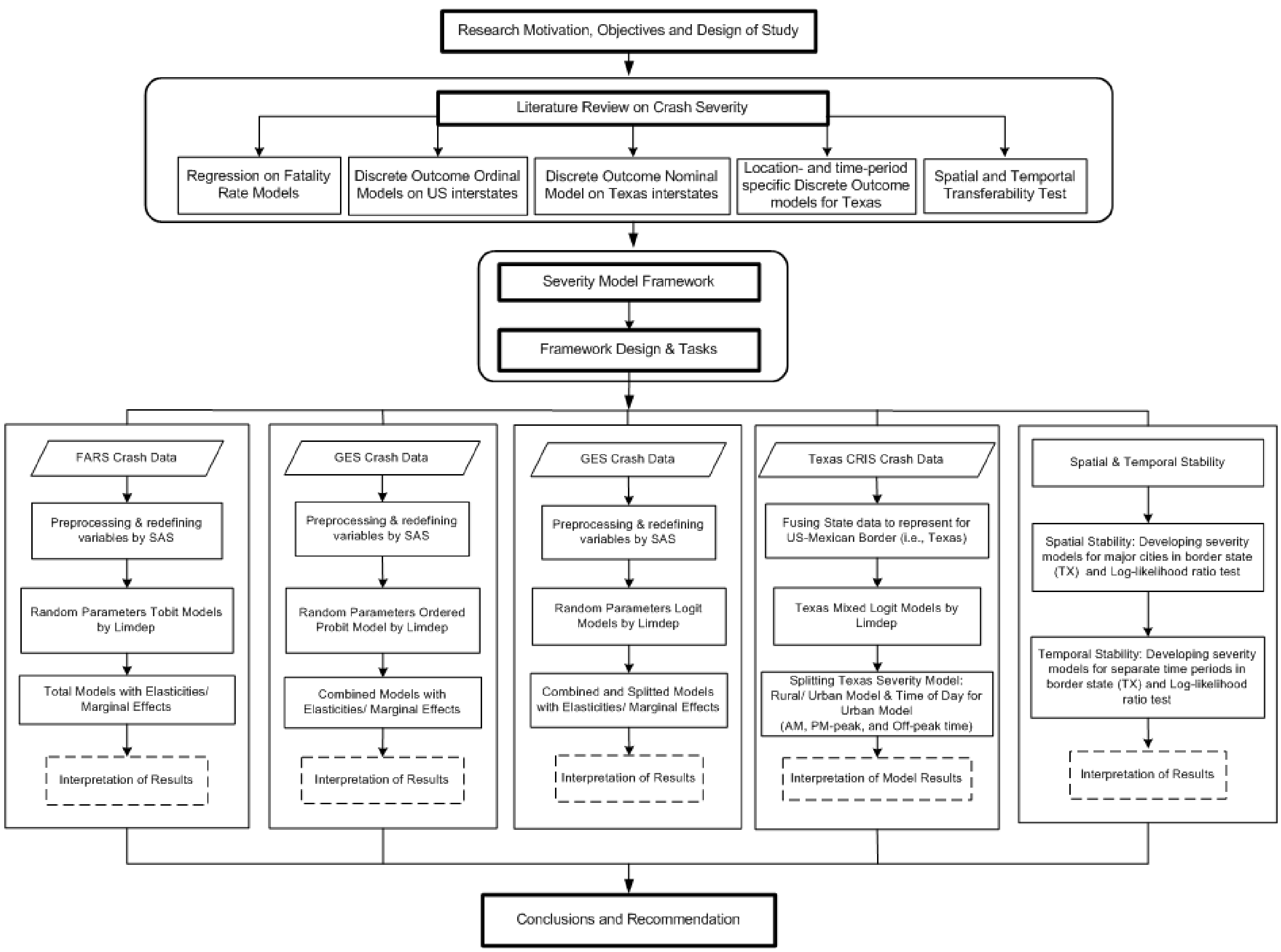

Figure 2.2: Research Framework and Tasks

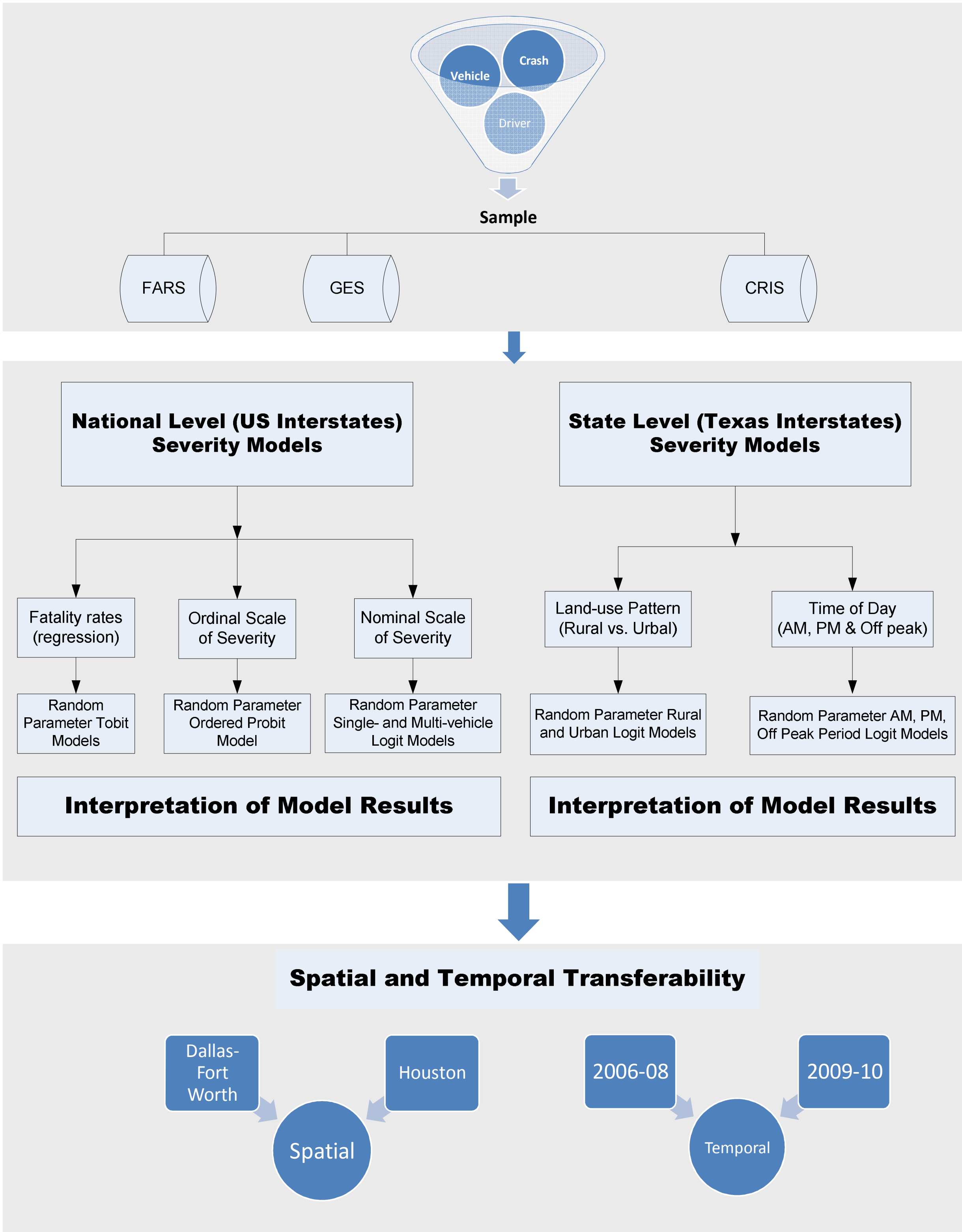

Figure 2.3: Research Framework and Model Estimation

Chapter 3: An Empirical Analysis of Fatality Rates for Large-truck involved Crashes on Interstate Highways

3.1 Introduction

To better understand the safety impacts related to increased large truck traffic on the nation's highway system, tools need to be developed that can aid transportation safety professionals as well as trucking industry operations managers in the avoidance and mitigation (i.e., aid them in the development of countermeasures) of large-truck involved crashes. With this in mind, our study aims to add to the current literature by proposing a methodological approach that takes into account fatalities per million truck-miles traveled and fatalities per ton-miles of freight for large-truck involved crashes. This is done through the application of a random parameters tobit modeling (censored at zero) framework (random parameters to account for unobserved factors that may vary across observations also known as heterogeneity). Through this, we seek to shed light on possible contributing factors to large-truck involved crashes.

Over the last two decades, crash frequency modeling approaches have been widely used in traffic safety analysis. The most frequently applied models in this regard have been the Negative Binomial and Poisson models (Shankar et al., 1995; Poch and Mannering, 1996; Abdel-Aty and Radwan, 2000; Savolainen and Tarko, 2005) and their variants the zero-inflated Poisson and zero-inflated Negative Binomial models (Shankar et al., 1997; Carson and Mannering, 2001; Lee and Mannering, 2002), random parameter Negative Binomial models (Shankar et al., 1998; Chin and Quddus, 2003; Anastasopoulos and Mannering, 2009), Markov switching of two different state of crash occurrence (Malyshkina and Mannering, 2009) and Bayesian statistics on Negative Binomial models (Park et al., 2010). Although literature in crash frequency modeling is rich, severe crash rates in terms of number of crashes per VMT has not been widely studied. Specifically, literature pertaining to the modeling of

fatalities per million truck-miles traveled or fatalities per ton-miles with respect to freight movements is relatively sparse. Using exposure-based crashes such as crashes per 100 million VMT instead of traditional crash frequency as the dependent variable carries more practical significance since crash rates are widely used in crash reporting (Anastasopoulos et al., 2008).

Although Hauer (1997) argued that expected accident rate should not be a legitimate measure of safety, the debate on crash frequency or rate is still on going. In our study traffic data particularly on interstates are not clearly free of ambiguity at the state or segment level, truck-miles and ton-miles information are considered aggregate in the sample. Here, we intend to study fatalities per truck-miles traveled or fatalities per ton-miles in response to other contributing factors, which nullify the effect of 'bias by selection' (Hauer, 1997). Because we utilized exposure data (say exposure in terms of truck-miles or ton-miles) only in the dependent variable (i.e., fatalities per million truck-miles or fatalities per ton-miles) without considering it (e.g., traffic data say AADT) as the independent variable and we are not considering a before-after safety effect study as was the case in Hauer (1997). In addition, we are not considering the traffic exposure variable (AADT) as an independent.

Trucking is important to the national economy, but it also presents a significant safety concern (Zhu and Srinivasan, 2011). In the 2007 Commodity Flow Survey trucks accounted for 70.7 percent of all freight movement, 68.8 percent by weight, and 39.8 percent by ton-miles of freight (USDOT/BTS, 2008). Zhu and Srinivasan (2011) illustrate that the unique operating characteristics, driving behavior and skills, design-weight related issues for trucks as a mode for freight movements significantly impacted the frequency of crashes, and severity of injuries sustained. This is further illustrated from the fact that 413,000 large trucks were involved in traffic crashes resulting in 4,808 fatalities, accounting for 12 percent of all fatal crashes in 2007 (NHTSA, 2008). And more recently, according to Federal Motor Carrier Safety Administration (2009), 10% of people were killed in the motor vehicle crashes involving a large truck.

In summary, the objective of this study is then to seek those factors related to drivers and passengers, vehicle and road-environment and weather that influence fatality rates as the highest level of injury severity for large-truck involved crashes. A random parameters tobit modeling framework is utilized to account for heterogeneity (Tobit model applications to transportation problems have primarily assumed fixed parameter estimates see Weiss, 1992; Talley, 1995; Nolan, 2002; Anastasopoulos et al., 2008; Anastasopoulos et al., 2012). The number of fatalities per million truck-miles traveled and number of fatalities per ton-miles for large truck freight movements is considered as a continuous variable instead of discrete integer (non-negative count) over a period of time. Since there is a likelihood of zero fatalities per million truck-miles traveled or zero fatalities per ton-miles of freight, this research is focused on fatalities higher than zero as a rate of safety indicator over a time period on US interstates, where the random parameters tobit modeling framework provides the flexibility of censoring the irrelevant count process in the regression estimation and at the same time account for unobserved factors that may vary across observations. To best of the authors' knowledge, these are the first attempts to model fatalities per million truck-miles traveled and fatalities per ton-miles for large truck freight movements utilizing a random parameters tobit modeling framework.

3.2 Method

In this section, the data is described followed by the modeling approach.

3.2.1 Data

To illustrate the application of the fixed- and random-parameters tobit models, crash data were collected from the Fatality Analysis Reporting Systems (FARS) from 2005 to 2008. FARS is a nation-wide crash census system where a set of files have been built documenting all qualifying fatal crashes that occurred within all the states in the U.S. The observation in the model is a fatal crash (A variable – *Fatals* includes the total number of fatalities in a fatal collision reported in the FARS database system)

involving a motor vehicle where at least a large truck is involved in the fatal collision traveling on U.S. interstate system resulting in a fatal (or fatalities) within 30 days for the collision.

The ton-miles of freight data from 2005 to 2007 were collected from the Bureau of Transportation Statistics special tabulation (BTS/RITA, 2010), whereas, truck-miles traveled data from 2005 to 2008 were collected from FHWA travel reports (FHWA, 2009). The second level data includes state level crash related data.

3.2.2 Data analysis

For model estimation, the truck-miles traveled and ton-miles of freight were aggregated for the range of years of 2005 to 2008 and 2005 to 2007, respectively. Then, fatalities per million truck-miles traveled and fatalities per ton-miles of freight were calculated as follows:

$$Fatality\ Rate = \left[\frac{Number\ of\ fatalities}{truck - miles\ traveled}\right] * 1{,}000{,}000 \tag{3.1}$$

$$Fatality\ Rate = \left[\frac{Number\ of\ fatalities}{ton - miles\ of\ freight * 1{,}000{,}000}\right] * 1{,}000{,}000 \tag{3.2}$$

The total number of observations utilized in this study for fatalities per million truck-miles (Eq. (3.1)) and fatality per ton-miles of freight (Eq. (3.2)) are 3498 and 2714, respectively. Here, the truck-miles of travel data was considered aggregate level for all state level by year of crashes, the same is true for ton-miles of freight. These two exposure data are very difficult to obtain at the state level or segment level, this is because these are not reported in the FARS system. So, there are four years (i.e., 2005 to 2008) of truck-miles of travel and three years (2005 to 2007) of ton-miles of freight and each is available at the aggregate level by years in the sample. The crash data were processed using the statistical software SAS. The LIMDEP (NLOGIT 4.0) software was utilized to estimate the fixed- and random-

parameter tobit models with proper codes under censored regression framework (see Table 3.1 for descriptive statistics for key variables).

3.2.3 Random parameters tobit model

To achieve a better understanding of the contributing factors associated to large-truck involved crashes, we seek to develop a statistical model that can be used to determine those influencing factors that affect the fatalities per million truck-miles traveled and fatalities per ton-miles of freight movements using a tobit modeling framework first introduced by James Tobin (1958). Most transportation safety applications of the tobit model have primarily been confined to the assumption of fixed parameter estimates as illustrated earlier; however, unobservable factors due to discretion of police officers at crash scenes and some critical factors not identified at particular fatal crashes may lead to bias estimates. Therefore, to account for unobserved heterogeneity we apply the random parameters tobit model. The following describes the general tobit model with respect to our problem and how incorporation and estimation of the random parameter to the tobit model is performed.

For this work, the standard tobit model is expressed (for large-truck involved in crash *i*) using a lower limit of zero (i.e., censored at zero) which is regarded the condition in our analysis for zero fatalities per million truck-miles traveled and zero fatalities per ton-miles of freight as (Washington et al, 2011):

$$
\begin{aligned}
&Y_i^* = \boldsymbol{\beta X}_i + \varepsilon_i, \qquad i = 1,2,\dots,N \\
&Y_i = Y_i^* \quad \text{if} \quad Y_i^* > 0 \\
&Y_i = 0 \quad \text{if} \quad Y_i^* \leq 0
\end{aligned}
\tag{3.3}
$$

where:

Y_i : is the dependent variable (fatalities per million truck-miles traveled or fatalities per ton-miles of freight),

$\boldsymbol{X}_i$: is a vector of independent variables (e.g., human, roadway segment, vehicle, and crash mechanism characteristics),

$\boldsymbol{\beta}$: is a vector of estimable parameters,

N : is the number of observations in the sample used in the model, and

ε_i : is normally and independently distributed error term with zero mean and constant variance σ^2.

However, to account for heterogeneity (unobserved factors that may vary across observations), Greene (2007) has developed estimation procedures (simulation based maximum likelihood estimation) for incorporating random parameters in tobit (censored regression) models (see Moeltner and Layton, 2002 for power outage costs application). To allow for such random parameters in tobit models, estimable parameters can be written as

$$\beta_i = \beta + \gamma_i \tag{3.4}$$

where:

γ_i : is randomly distributed term (for example a normally distributed term with mean 0 and variance σ^2)

With this equation, the tobit model for large-truck involved in crash i becomes $Y_i^*|\gamma_i = \boldsymbol{\beta X}_i + \varepsilon_i$. The corresponding log-likelihood can be written as

$$LL = \sum_{\forall i} ln \int_{\gamma_i} \mathrm{g}(\gamma_i) P(Y_i^*|\gamma_i) d\gamma_i \tag{3.5}$$

where:

$\mathrm{g}(\cdot)$: is the probability density function of the γ_i, and

$P(\cdot)$:   is the probability for the tobit model.

Maximum likelihood estimation of the tobit model shown in Eq. (3.3) is undertaken with simulation approaches due to the difficulty in computing the probabilities. The most widely accepted simulation approach uses Halton draws which is a technique developed by Halton (1960) to generate a systematic non-random sequence of numbers. Halton draws have been shown to provide a more efficient distribution of the draws for numerical integration than purely random draws (Bhat, 2003; Train, 1999).

For estimation procedures of the standard tobit model and marginal effects derivations the reader is referred to Amemiya (1973, 1985), McDonald and Moffitt (1980), Roncek, (1992), Anastasopoulos et al. (2008), and Anastasopoulos et al. (2012).

3.3 Empirical results

Table 3.1 illustrates a summary of the key variables considered in this study. With regards to fatalities per million truck-miles traveled, Table 3.1 illustrates that on average of 5.7 fatalities per million truck-miles traveled on the interstates are experienced which is substantially higher than passenger vehicle million miles travelled which had an average of about 1.4 fatalities (NHTSA, 2008). Additionally as seen from Table 3.1, rear-end collisions on average account for 37% of the total fatal crash observations for fatalities per million truck-miles traveled compared to 6.5% of angled collisions. Another key observation to note from the data is that the presence of a traffic median barrier for divided highways on average accounted for 29% of the total fatal crash observations for fatalities per million truck-miles traveled. The statistics further illustrate that wet surface and rainy weather conditions on average account for 13.4% and 9.3% of the total fatal crash observations for fatalities per million truck-miles traveled, respectively.

With regards to fatalities per ton-miles of freight, Table 3.1 shows that on average about one fatality per ton-miles of freight movement is experienced. The statistics as seen in Table 3.1 illustrate that angled collision accounts for 6.9% of the total fatal crash observations. The presences of a traffic median barrier for divided highways on average account for 28% of total crash observations for fatalities per ton-miles of freight. The temporal characteristics such as winter month (i.e., December) and day of week (i.e., Friday) on average account for 7.6% and 16% of total crash observations for fatalities per ton-miles of freight, respectively.

The correlation matrix for both of the tobit models – fatalities per million truck-miles traveled and fatalities per ton-miles freight was performed. The correlation matrix for the former (fatalities per million truck-miles traveled) and later (fatalities per ton-miles of freight) models indicate that number of people not fatally injured and number of vehicles involved in the fatal crashes having the coefficients of 0.559 and 0.528, respectively, which might pose some multicollinearity issues. However, injury mechanism for both of these variables account for exposure to fatalities.

In Chang and Mannering (1999), the authors showed that occupancy is a crucial exposure measure in crashes resulting severely injured vehicle occupants and is also found to be relevant in this study. Also, the same argument stands for the number of vehicles involved in crashes.

Table 3.1: Descriptive Statistics of Key Variables

Variables	Fatalities per million truck-miles traveled		Fatalities per ton-miles of freight	
	Mean	Std. Dev.	Mean	Std. Dev.
Fatalities per million truck-miles traveled	5.676	3.261	-	-
Fatalities per ton-miles of freight	-	-	0.985	0.58
Manner of collision (1 if rear-end, 0 otherwise)	0.367	0.482	-	-
Manner of collision (1 if angle, 0 otherwise)	0.065	0.246	0.069	0.253
Ambient light condition (1 if dawn time, 0 otherwise)	0.027	0.163	0.027	0.163
Surface condition (1 if wet, 0 otherwise)	0.134	0.341	-	-
Weather condition (1 if foggy, 0 otherwise)	-	-	0.017	0.129
Weather condition (1 if rainy, 0 otherwise)	0.093	0.289	-	-
Traffic median barrier (1 if divided highway with traffic barrier, 0 otherwise)	0.291	0.454	0.28	0.449
Time of the day (1 if 5 pm in the evening, 0 otherwise)	0.041	0.198	0.043	0.203
Time of the day (1 if 6 pm in the evening, 0 otherwise)	0.034	0.183	0.034	0.181
Trailing unit (1 if two trailing unit, 0 otherwise)	0.051	0.219	0.052	0.222
State specific crash information (1 if Texas, 0 otherwise)	0.091	0.287	0.093	0.29
Month of the year (1 if month is August, 0 otherwise)	0.083	0.276	-	-
Month of the year (1 if month is December, 0 otherwise)	-	-	0.076	0.265
Day of the week (1 if Friday, 0 otherwise)	-	-	0.16	0.367
Crash related human factors (1 if driving too fast, 0 otherwise)	0.068	0.252	0.079	0.271
Driver's license type (1 if license is valid, 0 otherwise)	0.863	0.343	-	-
Number of involved vehicles in crash	1.843	1.369	1.81	1.195
Number of person not fatally insured	2.802	4.206	2.739	4.152

Table 3.2 and Table 3.3 present estimation results for the tobit fixed- and random-parameters models for fatalities per million truck-miles traveled and fatalities per ton-miles of freight, respectively on interstates system in the US. The random parameters tobit models were estimated using simulation-based maximum likelihood with 200 Halton draws. This number of draws has been empirically shown to produce accurate parameter estimates (Bhat, 2003; Milton et al., 2008; Gkritza and Mannering, 2008). With regard to the distribution of the tobit random parameters, consideration was given to the normal, lognormal (which restricts the impact of the parameters to be either negative or positive), triangular, and

uniform distributions. However, only the normal distribution was found to be significant. The estimation results in Tables 3.2 and 3.3 show the estimated parameters with their respective statistical significance (*t-stat* and *P-value*) and plausible sign based on the sample sizes of 3498 (fatalities per million truck-miles traveled) and 2714 (fatalities per ton-miles of freight) of crash observations that had complete information of all variables used.

The Madalla pseudo R^2 was estimated for both the fixed- and random-parameter tobit models (see Tables 3.2 and 3.3) (Madalla, 1983). Veall and Zimmermann (1996) show that the Madalla pseudo R^2 is good indicator of overall goodness of fit and is computed as (also see Anastasopoulos et al., 2008)

$$Madalla\ pseudo\ R^2 = 1 - e^{[-2(LL(\beta)-LL(0))/N]} \tag{3.6}$$

where:

$LL(\beta)$: is log-likelihood at convergence,

$LL(0)$: is log-likelihood at zero, and

N : is the number of observations.

For the fatalities per million truck-miles traveled model, the pseudo R^2 were found to be 0.227 and 0.355 for the fixed- and random-parameter tobit models, respectively. Similarly, for the fatalities per ton-miles of freight model, the pseudo R^2 were found to be 0.222 and 0.360 for the fixed and random parameter tobit models, respectively. The pseudo R^2 for the tobit models indicate that the random parameter tobit models are more robust in explaining unobserved heterogeneity than fixed parameter tobit models. Furthermore, a likelihood ratio test comparing the fixed- and random-parameters models for the fatalities per million truck-miles traveled ($\chi^2 = 629.51$) and fatalities per ton-miles of freight ($\chi^2 = 529.54$) indicates that random parameter models are statistically significant from their fixed parameter models for more than 99.99% (a *p-value* near zero) level of significance (see Washington et al., 2011). Therefore, the interpretation of the estimation of results will be confined to

both the fatalities per million truck-miles traveled and fatalities per ton-miles of freight random parameter tobit models.

To assess the degree of influence of specific variables, Table 3.4 illustrates the computed marginal effects for the fatalities per million truck-miles traveled and fatalities per ton-miles of freight for the random parameter tobit models, respectively. Finding the marginal effect of an independent variable on the expected value of a dependent variable for all cases, $E[Y]$, was calculated using the McDonald and Moffitt (1980) formula:

$$\partial E[Y]/(\partial X_i) = F(z) \times (\partial E[Y^*]/(\partial X_i)) + E[Y^*] \times (\partial F(z)/(\partial X_i)) \qquad (3.7)$$

where:

$F(z)$: is the cumulative normal distribution function, associated with the proportion of cases above the limit (in this case zero),

$E[Y^*]$: denotes the expected value for cases above zero,

$\frac{\partial E[Y^*]}{\partial X_i}$: denotes observations above zero which indicates fatalities per million VMT and fatalities per ton-miles of freight (not censored),

$\frac{\partial F(z)}{\partial X_i}$: is the change in the cumulative probability of being above zero associated with an independent variable.

Table 3.2: Tobit Regression Estimation for Fatalities per Million Truck-miles Traveled

Variables	Fixed Parameter Tobit			Random Parameter Tobit		
	Coeff.	*t-stat*	*P-value*	*Coeff.*	*t-stat*	*P-value*
Constant	3.430	19.263	0.000	3.916	19.435	0.000
Crash Mechanism						
Manner of collision (1 if rear-end, 0 otherwise)	-0.186	-1.744	0.081	-0.229	-2.122	0.034
Manner of collision (1 if angle, 0 otherwise)	0.591	2.901	0.004	0.725	4.961	0.000
Temporal Characteristics						
Ambient light condition (1 if dawn time, 0 otherwise)	0.598	1.995	0.046	0.920	4.175	0.000
Time of the day (1 if 5 pm in the evening, 0 otherwise)	0.756	3.063	0.002	0.898	5.046	0.000
Time of the day (1 if 6 pm in the evening, 0 otherwise)	0.809	3.040	0.002	0.806	4.574	0.000
Month of the year (1 if month is August, 0 otherwise)	0.416	2.362	0.018	0.432	3.131	0.002
Location Characteristics						
State specific crash information (1 if Texas, 0 otherwise)	0.501	2.950	0.003	0.361	2.504	0.012
Environment - Weather						
Weather condition (1 if rainy, 0 otherwise)	-0.834	-3.266	0.001	-0.741	-3.275	0.001
Road Surface condition (1 if wet, 0 otherwise)	0.565	2.606	0.009	0.478	2.609	0.009
Road - Geometry						
Traffic median barrier (1 if divided highway with traffic barrier, 0 otherwise)	-0.234	-2.182	0.029	-0.233	-2.007	0.045
Vehicle Configuration						
Trailing unit (1 if two trailing unit, 0 otherwise)	-0.517	-2.334	0.019	-0.488	-1.881	0.060
Human Factor						
Vehicle maneuver (1 if going straight, 0 otherwise)	0.398	3.527	0.000	0.174	1.508	0.132*
Crash related human factors (1 if driving too fast, 0 otherwise)	1.167	5.970	0.000	0.741	4.897	0.000
Driver's license type (1 if license is valid, 0 otherwise)	0.399	2.813	0.005	0.332	2.003	0.045
Exposure to Injury Severity						
Number of vehicles involved in the crash	0.438	10.211	0.000	0.293	8.887	0.000
Std. dev. of parameter distribution				*0.409*	*36.118*	*0.000*
Number of persons not fatally injured in the crash	0.246	17.328	0.000	0.236	19.659	0.000
Std. dev. of parameter distribution				*0.247*	*32.781*	*0.000*
Number of variables	17			17		
Log-likelihood at zero, $LL(\mathbf{0})$	-9097.403			-9097.403		
Log-likelihood at convergence, $LL(\boldsymbol{\beta})$	-8646.047			-8331.289		
$X^2 = -2[LL(\mathbf{0}) - LL(\boldsymbol{\beta})]$	902.71			1532.228		
Number of observations	3498			3498		
Madalla pseudo-R^2	0.227			0.355		

**the p-value is considered upto 0.15 indicating that we are 85% confident that coefficient estimates are significantly different from zero.*

Table 3.3: Tobit Regression Estimation for Fatalities per Ton-miles of Freight

	Fixed Parameter Tobit			Random Parameter Tobit		
Variables	Coeff.	t-stat	P-value	Coeff.	t-stat	P-value
Constant	0.679	32.485	0.000	0.696	33.487	0.000
Std. dev. of parameter distribution	-	-	-	*0.025*	*3.096*	*0.002*
Crash Mechanism Manner of collision (1 if angle, 0 otherwise)	0.136	3.469	0.000	0.135	5.038	0.000
Temporal Characteristics						
Ambient light condition (1 if dawn time, 0 otherwise)	0.110	1.916	0.055	0.108	3.146	0.002
Time of the day (1 if 5 pm in the evening, 0 otherwise)	0.147	3.025	0.003	0.137	4.205	0.000
Time of the day (1 if 6 pm in the evening, 0 otherwise)	0.185	3.391	0.001	0.164	5.232	0.000
Day of the week (1 if Friday, 0 otherwise)	0.062	2.312	0.021	0.077	3.117	0.002
Month of the year (1 if month is December, 0 otherwise)	-0.100	-2.686	0.007	-0.062	-1.775	0.076
Location Characteristics						
State specific crash information (1 if Texas, 0 otherwise)	0.109	3.194	0.001	0.092	3.226	0.001
Environment - Weather Weather condition (1 if foggy, 0 otherwise)	-0.267	-3.471	0.000	-0.214	-2.488	0.013
Road - Geometry						
Traffic median barrier (1 if divided highway with traffic barrier, 0 otherwise)	-0.052	-2.350	0.019	-0.036	-1.571	0.116*
Vehicle Configuration						
Trailing unit (1 if two trailing unit, 0 otherwise)	-0.102	-2.309	0.021	-0.082	-1.627	0.104*
Human Factor						
Crash related human factors (1 if driving too fast, 0 otherwise)	0.223	6.041	0.000	0.173	6.366	0.000
Exposure to Injury Severity Number of vehicles involved in the crash	0.087	8.923	0.000	0.075	11.489	0.000
Std. dev. of parameter distribution	-	-	-	*0.068*	*27.787*	*0.000*
Number of persons not fatally injured in the crash	0.043	15.117	0.000	0.039	17.377	0.000
Std. dev. of parameter distribution	-	-	-	*0.055*	*35.470*	*0.000*
Number of variables		14			14	
Log-likelihood at zero, $LL(\mathbf{0})$		-2373.439			-2373.439	
Log-likelihood at convergence, $LL(\boldsymbol{\beta})$		-2032.271			-1767.500	
$X^2 = -2[LL(\mathbf{0}) - LL(\boldsymbol{\beta})]$		682.336			1211.878	
Number of observations		2714			2714	
Madalla pseudo-R^2		0.222			0.360	

*****_the p-value is considered upto 0.15 indicating that we are 85% confident that coefficient estimates are significantly different from zero._

Table 3.4: Marginal Effects Comparison for Fixed- and Random-parameter Tobit Models for Fatality Rate Models

Variables	Fatalities per million truck-miles traveled		Fatalities per ton-miles of freight	
	Random	Fixed	Random	Fixed
Constant	3.869	3.349	0.688	0.661
Manner of collision (1 if rear-end, 0 otherwise)	-0.226	-0.182	-	-
Manner of collision (1 if angle, 0 otherwise)	0.716	0.576	0.133	0.132
Ambient light condition (1 if dawn time, 0 otherwise)	0.909	0.583	0.107	0.108
Surface condition (1 if wet, 0 otherwise)	0.472	0.551	-	-
Weather condition (1 if foggy, 0 otherwise)	-	-	-0.211	-0.26
Weather condition (1 if rainy, 0 otherwise)	-0.732	-0.815	-	-
Traffic median barrier (1 if divided highway with traffic barrier, 0 otherwise)	-0.23	-0.229	-0.036	-0.05
Time of the day (1 if 5 pm in the evening, 0 otherwise)	0.888	0.738	0.136	0.143
Time of the day (1 if 6 pm in the evening, 0 otherwise)	0.797	0.789	0.162	0.18
Trailing unit (1 if two trailing unit, 0 otherwise)	-0.482	-0.505	-0.081	-0.099
Vehicle maneuver (1 if going straight, 0 otherwise)	0.172	0.388	-	-
State specific crash information (1 if Texas, 0 otherwise)	0.356	0.489	0.091	0.106
Month of the year (1 if month is August, 0 otherwise)	0.426	0.406	-	-
Month of the year (1 if month is December, 0 otherwise)	-	-	-0.062	-0.097
Day of the weekend (1 if Friday, 0 otherwise)	-	-	0.076	0.061
Crash related human factors (1 if driving too fast, 0 otherwise)	0.732	1.139	0.171	0.217
Driver's license type (1 if license is valid, 0 otherwise)	0.328	0.39	-	-
Number of vehicles involved in the crash	0.29	0.428	0.074	0.006
Number of persons not fatally injured in the crash	0.233	0.24	0.039	0.084

3.3.1 Fatalities per million truck-miles traveled model

The indicator variable representing rear-end collisions decreases fatalities per million truck-miles traveled. This may be due to most occupants sitting in the front seats of their vehicle (trucks) and are afforded more full body protection from the rear seats (trailers) and head restraints (airbags) upon collision. In addition, the direction of the impact and the resulting relative movement of the occupants' minimize the chance of more serious injuries of striking more lethal objects in the vehicle (Duncan et al., 1998). The average marginal effect for this variable is -0.226 (and only a decrease of -0.182 for the

fixed parameter model). The angle collision indicator variable increases fatalities per million truck-miles traveled. In contrast to rear-end collisions, angled collisions lead to more severe injury outcomes (e.g., fatalities) especially when large trucks are involved. This may be due to the structural dynamic makeup of vehicles especially when struck in an angle—not as energy absorbing as the front or rear of vehicles (Abdel-Aty and Abdelwahad, 2004). Marginal effect for this variable is 0.716 compared to 0.576 for the fixed parameter model.

With regards to the temporal variables, all the indicator variables increase fatalities per million truck-miles traveled. First, the dawn variable (before the sunrise) increases the fatality rate. This may be a result of driver experiencing drowsiness and maybe capturing, among other factors, the effects related to long hours of driving. A similar result found in a study by Ivan et al. (2000) indicated that "dawn" time increases single vehicle crash rates. Next, the time widow between 5 to 6 pm increases fatalities per million truck-miles traveled. This may be capturing some driver related factors (as in the dawn variable) with regards to the level of alertness and fatigue. During the summer periods, in particular, the month of August increases fatalities per million truck-miles traveled. This may be reflecting vehicular interactions on highways due to preferable weather condition for outdoor activities especially during this time of year. Malyshkina and Mannering (2009) also found that the summer months (from June to August) increase the likelihood of severity of single-vehicle collisions on Indiana state routes and interstates. A marginal effect of 0.426 for the August indicator variable is observed for the random parameters tobit model compared to 0.402 for the fixed parameter tobit model.

Crashes that occurred in the State of Texas indicator variable increases fatalities per million truck-miles traveled. A possible explanation may be related to the number of trucks in the State of Texas due to it sharing a border with North American Free Trade Agreement (NAFTA) member Mexico. This variable may be capturing the driving complexities related to the diverse geographical nature of the State of Texas and network familiarity. In addition, the speed limit in Texas is higher and varies on rural

highways (i.e., 75, 80 and 85 mph) and urban highways (i.e., 75 mph) compared to other states in the US (GHSA, 2012). The Texas variable may be capturing those human factors related speeding with regards to speed limit differences (once again familiarity issues).

With respect to weather, the indicator variable for rain was found to be significant and decreased the fatalities per million truck-miles traveled. A possible explanation is that truck drivers are more cautious while driving through rain. This result is supported by Zhu and Srinivasan (2011) and Chen and Chen (2011) based on the risk-averse behavior of the drivers in the adverse weather conditions. On one hand, the indicator variable for surface condition being wet increases fatalities per truck-miles traveled. This is possibly capturing, among other factors, vehicular conditions (e.g., tire wear leading to hydroplaning). Chen and Chen (2011) also show for wet surface conditions due to snow/slush, increases the likelihood of collisions.

The presences of median barriers (or not) separating the opposing traffic flow decreases fatalities per million truck-miles traveled. As shown in Anastasopoulos et al. (2008) median barriers potentially reduces head-on collisions and may lower injury severity, which significantly reduces the likelihood of fatalities.

The indicator variable for a truck hauling two trailers (i.e., Truck Tractor-Semitrailer-Trailer Combinations having two 28.0 to 28.5 feet trailing units) decreases the fatalities per million truck-miles traveled. A possible reason is that professional truck drivers that receive additional practical driving and maneuvering safety training drive these types of large trucks, in contrast to single trailer unit truck drivers.

Driving straight in a traffic lane as a crash avoidance maneuver (or not) increases fatalities per million truck-miles traveled. This may be due to, among other factors, the kinematics revolving around large-truck involved crashes. Akin, driving too fast was identified as increasing the fatalities per million

truck-miles traveled. Speed (being the top factor identified in the FARS data) has been shown to increased fatalities rates due to due higher energy transfer between colliding bodies (Craft, 2010).

The indicator variable for a truck driver who poses a valid Commercial Drivers' License (CDL) (or not) increases fatalities per million truck-miles traveled. This variable may be capturing unobserved effects related to the level of experience, years of driving, drivers with none or fraudulent CDL's. A study by Blower and Kostyniuk (2007) indicated that there could be instances of fraudulent issuance of CDL licenses and because of lack of proper training may increase the likelihood of crashes.

With regard to the parameters found to be random, the number of vehicles involved in a crash varies across observations and resulted in a random parameter that is normally distributed, with a mean of 0.293 and a standard deviation of 0.409. The positive sign indicates that an increase in number of vehicles involved in a crash per million trucks-mile traveled increases the fatality rate (less than 23.7 percent of the distribution would have a negative value). One possible explanation for this finding is that crashes with many cars (e.g., pile ups) varies in severity (may not always lead to fatalities) due to some unforeseen pile up dynamics and preventive technologies present in vehicles (Chakravarthy et al., 2009). With respect to marginal effects, Table 3.4 shows that a unit increase in the number of vehicles involved in the crash results in an average 0.29 increase in the number of fatalities per million truck-miles traveled. This variable was also found to be significant by Chen and Chen (2011) for multi-vehicle collisions.

Similarly, the number of occupants not fatally injured in the crash was also found to be random and normally distribution with a mean of 0.236 and a standard deviation of 0.247. Given the distributional patterns, an increase in the number of persons not fatally injured in a crash increases fatalities per million truck-miles traveled but with varying magnitude—that is, less than 16.9 percent of the distribution (less than zero) would have a negative value (would increase fatalities). A possible reason for this finding may be due to that fact that persons dying sometime later due to serious injuries

sustained during the crash may not be updated later on the police reports. Marginal effects show that a unit increase in persons not injured in the crash results in an average 0.23 increase in the number of fatalities per million truck-miles traveled. More broadly, Islam and Mannering (2006) also indicate that the likelihood of fatality increases when one or more occupants travel with the driver.

Two parameters were found to be random with statistically significant standard deviations for their assumed distributions. Also, for the parameters whose standard deviations were not statistically different from zero, the parameters were fixed to be constant across observations. The estimation results shown in Table 2 indicate that the number of vehicles involved in the crash, and the number of persons not fatally injured in the crash were found to produce statistically significant random parameters. Both of the exposure level variables – number of occupants (non-fatally injured) and number of vehicles involved in the crashes increase the fatality rate which is consistent with findings from a study by Chang and Mannering (1999).

3.3.2 Fatalities per ton-miles of freight Model

To avoid repetition in the explanation of the specified estimates found in the two models, only variables specific to the fatalities per ton-miles of freight will be explained in this section. Turning to the model specification, three parameters were found to be random with statistically significant standard deviations for their assumed distributions. Also, for the parameters whose standard deviations were not statistically different from zero, the parameters were fixed to be constant across the observations.

The constant for fatalities per ton-miles of freight is found to be random and normally distributed with mean of 0.696 and standard deviation of 0.025. With these distributional patterns, the constant term is less than zero for 0% and more than zero for 100% of the large-truck involved fatalities per ton-miles of freight. This variability is likely capturing the unobserved heterogeneity in the severity outcomes that

could include factors such as traffic condition, among other factors, which was not directly measured in the dataset for this model.

With regards to the significant temporal variables, the day of the week (i.e., Friday) indicator variable increases fatalities per-ton miles. Although freight movements are made pretty uniformly from Monday through Friday, this variable may be capturing, among other factors, some weekend effects. Similar findings by Chang and Mannering (1999) indicated that weekend increases the likelihood of severe injuries (fatality and injury crashes) in non-truck involved crashes. However, weekend increases the likelihood of non-severe injuries for truck involved crashes as found by Chang and Mannering (1999). In addition, the indicator variable for December was found to be significant and decreased fatalities per ton-miles of freight. The significance of this variable may stem from the lower activity of freight movements due to winter (the possibility of adverse weather conditions such as snow), and seasonal effects (e.g., Christmas holidays). Typically, freight movements are at their highest in the early fall for the winter holiday season.

Consistent to Zhu and Srinivasan (2011) we find that the presence of foggy weather conditions has a negative effect on fatalities per ton-miles of freight. As was the finding with the rain indicator variable earlier, truck drivers are more cautious while driving through foggy conditions. Additionally, this variable may be capturing some risk-averse behavior of drivers.

The estimation results shown in Table 3.3 indicate that the constant, the number of vehicles involved in the crash, and the number of persons not fatally injured in the crash were found to produce statistically significant random parameters.

3.4 Summary and conclusions

This study provides a demonstration of the random parameters tobit regression as a viable methodological approach to gain new insights into factors that significantly influence fatalities per

million truck-miles traveled and fatalities per ton-miles of freight. The random-parameters tobit regression modeling framework is an important approach because it allows us to account and correct for heterogeneity that can arise from factors such as individuals (i.e., drivers and passengers), vehicle, road-environment, weather, variations in police reporting, temporal and other unobserved factors not captured. Although, victim level or the location of victims in the crash were not found to be significant they may share some unobserved effect with other variables found significant in the models.

Using four years of data for fatalities per million truck-miles traveled and three years of data for fatalities per ton-miles of freight our estimation results provide some interesting findings, respectively. For example, factors related to the type of collision were found to be significant including rear-end and angled crashes as was driving too fast. Temporal factors were also found to be significant such as the effects of dawn, evening times between 5 and 6 pm, and the months of August and December. In terms of locational variables the State of Texas was found to be a contributing factor for both models. Also, factors related to weather which included rain, foggy, and wet surfaces were significant. With regards to road geometry, the presences of traffic medians impacted both models. The hauling of two trailers by a truck was also found to be significant for both models. And, exposure variables number of vehicles involved in a crash and the number of persons not fatally injured were significant. Although traffic data such as AADT has not been incorporated in the dataset for the developed models, there are variables in both models representing the time of the day (dawn time, between 5 pm to 6 pm), day of the week (Friday) and month of the year (August, December) serve as a proxy for traffic conditions on the highway system.

Although this study is exploratory in nature, the modeling approach presented in this paper offers a flexible methodology that has considerable potential to analyze fatalities per million truck-miles traveled and fatalities per ton-miles of freight. Applying this approach to state specific datasets with available AADT (average annual daily traffic) data and for more years, would potentially provide more

information on the effects of contributing factors present and new on fatalities per million truck-miles traveled and fatalities per ton-miles of freight.

Chapter 4: An Analysis of Single- and Multi-vehicle Large-truck involved Crashes

4.1 Introduction

Past research has shown that two separate models are more reliable and more accurate in predicting performance than a single holistic (or whole) model in the context of crash severity (Ivan, 2004; Lord et al., 2005). However, most of these works are from the perspective of passenger vehicles. Here we consider the development of a random parameter logit model (mixed logit model) for single- and multi-vehicle models utilizing the General Estimate System (GES) nationwide crash database with focus on large-truck involved crashes in the US interstate system. These models potentially identify the factors associated with likelihood of crashes resulting in different levels of injury outcomes on the highways systems for crashes involving large trucks. The injury severities of the occupants in both single- and multi-vehicle collisions involving at least a large truck are quite alarming considering the national level crash statistics.

According to NHTSA Safety Facts (2012) there was a 9.0% increase in fatalities in 2010 as compared to 2009 in crashes involving large trucks. This increase in fatalities includes large-truck occupants, occupants of other vehicles and non-occupants in large truck crashes. There is 14% of large-truck occupants killed in single- and multi-vehicle crashes, in contrast to a 76% of fatalities of other vehicle occupants involved in those multivehicle crashes (NHTSA, 2012).

With this in mind, this chapter provides the first attempts at modeling single- and multi-vehicle collisions for large-truck involved crashes.

4.2 Background

This section provides a background on single- and multi-vehicle collisions.

Krantz (1979) provided the first documented glimpse on the difference of characteristics between single and multi-vehicle collision in Sweden. In their study, the authors analyzed 458 fatal crashes that occurred in 1975 throughout Sweden. The crashes were based on time of occurrence and weather condition, crash scene, vehicle types, drivers' characteristics – age, drinking habit, social and criminal background. Their study found that drivers' characteristics in single vehicle crashes were young in age, intoxicated, and socially deviant including low seat belt usage and low license possessions. Another Swedish study by Ostrom and Eriksson (1993) also found similar findings for their study that analyzed 597 fatal crashes in Northern Sweden from 1980 to 1989. The study found that single-vehicle fatalities are more likely to occur on weekends and during dark hours. Also, the study established the fact of previous study that majority (58%) of drivers in single-vehicle crashes were intoxicated, compared to only very small number (10%) of the multi-vehicle crashes.

A study on single vehicle collisions in rural settings by Xie et al. (2012) found that driver age, driving under influence, seat belt usage, points of impact, lighting condition, speed, first and second harmful events, and ethnicity are found to be closely related to driver injury severity levels.

Yau et al. (2006) show that road type, speed limit, and number of vehicles involved in the accident usually play a critical role in multiple vehicle collisions, which is not the case for single-vehicle collisions. Using Traffic Accident Data System in Hong Kong, they found that the district board, time of the accident, driver's gender, vehicle type, road type, speed limit, and the number of vehicles involved are contributing factors in injury severity.

Ivan et al. (1999) developed single- and multi-vehicle crash frequency prediction models on two lane rural highways. The study found different explanatory variables that influence single and multi-vehicle crashes. For example, single-vehicle crash rates increase with decreasing traffic intensity, shoulder width, and sight distance. On the other hand, multi-vehicle crash rates increase with the number of signals, the daily single-unit truck percentage, and shoulder width, and decreased on principal

arterials compared to other roadway classes. Another study by Ivan et al. (2000) also analyzed single and multi-vehicle crash frequency models and found traffic density and land use, ambient light conditions, and time of day significant. Time of day (6 am to 7 pm), segment volume-capacity ratio, and number of intersections are negatively associated; whereas, percent of segment without passing-zone, and shoulder width are positively associated with single-vehicle collisions with driveway having mixed effects. On the other hand, day light condition (10 am to 3 pm, 3 pm to 7 pm), and driveways are positively associated; whereas, the number of intersections are negatively associated with multi-vehicle collisions.

Based on exploratory data analyses and regression methods, Persaud and Mucsi (1995) show that single- and multi-vehicle crashes have significantly different characteristics. Their study indicated that separate models predict more crashes than aggregate models (i.e., combining single- and multi-vehicle crashes). Also the crash mechanism for single- and multi-vehicle crashes is quite different from each other (Chen & Chen, 2010; Baker, 1991) considering different vehicle types involved in the crashes. Savolainen and Mannering (2007) also found the crash risk factors such as poor visibility, (horizontal curvature, vertical curvature, darkness); unsafe speed (citations for speeding); alcohol use; not wearing a helmet; right-angle and head-on collisions; and collisions with fixed objects vary based on the types of crashes (2273 single vehicle – motorcycle and 2213 multi-vehicle crashes involving motorcycles in Indiana) to be significant.

Jonsson et al. (2007) analyzed single- and multi-vehicle collisions on four-lane rural highways considering different crash scenarios: opposite-direction, same-direction, intersecting-direction and single-vehicle collisions using California crash data over a period from 1993 to 2002. They found that Annual Average Daily Traffic (AADT) of major and minor roads as separate variables affects a multi-vehicle collision model; whereas, the sum of major- and minor-road AADT jointly affects a single-vehicle collision model. Aggregation of data affects some of the covariates in the single-vehicle crash

model. Also, they indicated that there are differences in the crashes types by crash-flow relationship, variables in the models, and severity distribution.

Geedipally and Lord (2010) analyzed Texas multi-lane undivided highway crashes and found out that modeling single- and multi-vehicle crashes separately provides wider confidence intervals than a single model jointly. As such, this difference is much higher for fatal and injury crash models than for models for all severity levels.

Loeb and Clarke (2007) conducted time series analysis from 1970 to 2001 on truck accidents in the US. They found that alcohol consumption, the unemployment rate, and railroad activity significantly affected truck accidents, while deregulation of the trucking industry did not adversely affect these accidents.

Based on ten years of crash data from Highway Safety Information System (HSIS) involving trucks on rural highways, Chen and Chen (2011) found that there are substantial differences between single- and multi-vehicle collisions analyzing drivers' injury severity by mixed logit model. Their findings also recommend that more rational and effective injury preventive strategies may be developed for single- and multi-vehicle collisions where driving conditions vary significantly.

4.3 Mathematical approach

Many research studies utilized a number of methodological approaches such as the multinomial logit, ordered probit and Bayesian Ordered Probit, nested logit, and mixed logit models to model injury severities (Chistoforou et al., 2010; Lemp et al., 2011; Milton et al., 2008; Zhu & Srinivasan, 2011; Duncan et al., 1998; Xie et al., 2009; Khorashadi et al., 2005; Chang & Mannering, 1999). In this study, we estimate a mixed logit model for injury severity of crashes involving large trucks following the similar logical research framework as of past research studies (Milton et al., 2008, Gkritza & Mannering, 2008, Chen & Chen, 2011). The level of injury is discrete in nature as coded in the injury

scale KABCO (i.e., ‘K’ for Fatal, ‘A’ for Incapacitating injury, ‘B’ for Non-incapacitating Injury, ‘C’ for Possible Injury and ‘O’ for Property-Damage-Only), and a mixed logit model has been widely accepted to model the effects of several contributory factors on the levels of injury severity. Several research studies conducted by Revelt and Train (1997, 1999), Train (1997), Brownstone and Train (1999), McFadden and Train (2000), and Bhat (2001) have shown the effectiveness of this methodological approach. Although discrete outcome severity could be modeled by a multinomial logit model, heterogeneous effects and correlation in unobserved factors could still be potential limitations in the assumption behind utilizing this model for injury severity (Train, 2009). So, a mixed logit model overcomes all of these limitations by generalizing multinomial logit structure, allowing for the parameters $\boldsymbol{\beta}_i$ vector to vary across the observation of crashes (Savolainen et al., 2010). The assumption regarding IID (independently and identically distributed errors), IIA (independence of irrelevant alternatives) and unobserved heterogeneity associated with observations in multinomial logit model is completely relaxed (Jones and Hensher, 2007).

In order to achieve a better understanding of the injury severity of large-truck involved crashes on the US interstate system, we seek to develop a statistical model that can be used to determine the contributing factors that influence the likelihood of severity outcomes in large-truck involved crashes. To do so, we start with a linear function S_{in} that determines discrete injury severity outcome i (fatality, incapacitating injury, non-incapacitating injury, possible injury, and property damage only) for large-truck involved incident *n* such that (Washington et al., 2011):

$$S_{in} = \boldsymbol{\beta}_i \mathbf{X}_{in} + \varepsilon_{in} \tag{4.1}$$

where $\mathbf{X}_{in}$ is vector of explanatory variables (driver, vehicle, road, and environment variables), $\boldsymbol{\beta}_i$ is vector of estimable parameters, ε_{in} is the error term. If ε_{in}'s are assumed to be generalized extreme value distributed, McFadden (1981) has shown that the multinomial logit results such that:

$$P_n(i) = \frac{EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]}{\sum_I EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]} \tag{4.2}$$

where $P_n(i)$ is probability of large-truck involved incident *n* having severity outcome i ($i \in I$ with I denoting all possible injury severity outcomes)

As GES crash data are likely to have a significant amount of unobserved heterogeneity because the information regarding any of the factors are not obtained from the in-depth crash investigation or reconstruction studies (for example, relating to police reporting, roadway, vehicle, and driver factors), we consider the possibility that elements of the parameter vector $\boldsymbol{\beta}_i$ may vary across observations of each large-truck involved crash by using a random-parameters logit model (also known as the mixed logit model). Previous works by McFadden and Rudd (1994), Geweke et al. (1994), Revelt and Train (1997, 1999), Train (1997), Stern (1997), Brownstone and Train (1999), McFadden and Train (2000), and Bhat (2001) have shown the development and effectiveness of the mixed logit approach which can explicitly account for the variations (across crash observations) of the effects that variables have on the severity outcomes (or choices) considered in this study. The mixed logit model is written as (see Train, 2003):

$$P_{in} = \int \frac{EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]}{\sum_I EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]} f(\boldsymbol{\beta}_i|\boldsymbol{\varphi}) d\boldsymbol{\beta}_i \tag{4.3}$$

where, $f(\boldsymbol{\beta}_i|\varphi)$is the density function of $\boldsymbol{\beta}_i$, $\boldsymbol{\varphi}$ is a vector of parameters of the density function (mean and variance), and all other terms are as previously defined. This model can now account for severity outcome specific variations of the effect of $\mathbf{X}$ on large-truck involved crash probabilities, with the density function $f(\boldsymbol{\beta}_i|\boldsymbol{\varphi})$ used to determine $\boldsymbol{\beta}_i$. Mixed logit probabilities are then a weighted average for different values of $\boldsymbol{\beta}_i$ across the observations where some elements of the vector $\boldsymbol{\beta}_i$ may be fixed and some randomly distributed. If the parameters are random, the mixed logit weights are determined by the

density function $f(\boldsymbol{\beta}_i|\boldsymbol{\varphi})$ (Milton et al., 2008; Washington et al., 2011).

Maximum likelihood estimation of the mixed logit model shown in Equation (5.3) is undertaken with simulation approaches due to the difficulty in computing the probabilities. The most widely accepted simulation approach uses Halton draws which is a technique developed by Halton (1960) to generate a systematic non-random sequence of numbers. Halton draws have been shown to provide a more efficient distribution of the draws for numerical integration than purely random draws (Bhat, 2003; Train, 1999).

A study by Ye and Lord (2011) showed the influence of the sample size on injury severity modeling. Although the analysis by Ye and Lord (2011) regarding the sample size corresponding to each severity model is simulation driven, there are still a few findings that could be generalized in terms of sample size for the three commonly used models. Crash severity models with sample sizes below 1,000 should not be estimated. An ordered probit model is the one that requires the least samples more than 1000, a mixed logit is the most demanding on samples having more than 5000, and a multinomial logit model requirement are located between the ordered probit and mixed logit models – somewhere more than 2000. In our study, the single-vehicle crash sample size is 1,703 and multi-vehicle crash sample size is 6,588, which are below and over the safe threshold identified by Ye and Lord (2011), respectively.

In order to estimate the impact of particular variables on the injury-outcome likelihood, elasticities (or direct-pseudo elasticity) are computed. In the context of the current injury severity model, most of the variables are indicator in nature; direct-pseudo elasticities are estimated to measure the marginal effects of indicator variables when any particular indicator variable switches from 0 to 1 or reverse (Washington et al., 2011). Also, this is translated to percentage change in the likelihood of while the indicator variables switching between 0 and 1 or 1 to 0. For binary indicator variables, the direct-pseudo elasticity is estimated as follows (Kim et al., 2010):

$$E_{x_{nk}}^{P_{in}} = \frac{P_{in}[given\ x_{nk} = 1] - P_{in}[given\ x_{nk} = 0]}{P_{in}[given\ x_{nk} = 0]} \tag{4.4}$$

where, P_{in} is given the Equation (4.3) and simulated as shown in Equation (4.5).

x_{nk} = the k-th independent variable associated with injury severity i for observation n.

The unconditional probability in Equation (4.3) (Kim et al., 2010) can be estimated with an unbiased and smooth simulator (McFadden & Train, 2000) that is computed as (Walker & Ben-Akiva, 2002):

$$\hat{P}_{in} = \frac{1}{R}\sum_{r=1}^{R} P_{in} = \frac{1}{R}\sum_{r=1}^{R}\int \frac{EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]}{\sum_I EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]} f(\boldsymbol{\beta}_i|\boldsymbol{\varphi})d\boldsymbol{\beta}_i \tag{4.5}$$

where, R = the total number of draws.

Since the direct pseudo-elasticity is calculated for each observation, it is usually reported as the average direct pseudo-elasticity (taking average over the sample) as a measure of the marginal effect of an indicator variable on the likelihood of a particular injury severity outcome (Kim et al., 2010).

With the simulator in Equation (4.5), Maximum Simulated Likelihood Estimation (MSLE) can be used to estimate parameters and this MSLE estimator is asymptotically normal and consistent (Lee, 1992):

$$max_{\boldsymbol{\beta}_{in}} \sum_{n=1}^{N}\sum_{i=1}^{I} y_{in} ln\hat{P}_{in} \tag{4.6}$$

where, N = the total number of observations (i.e., crashes in the sample)

y_{in} = 1 if individual n suffers from injury severity i, 0 otherwise.

4.4 Empirical settings

The data for large-truck involved crashes was obtained from the nationwide NASS-GES crash database maintained by the National Highway Traffic Safety Administration (NHTSA). A large truck is

commonly classified as a tractor-trailer, single-unit truck, or cargo van having GVWR greater than 10,000 pounds (IIHS, 2009). The GES database is based on a nationally representative probability sample selected from the estimated 5.8 million police-reported crashes resulting in a fatality or injury and those involving major property damage annually (NASS-GES, 2008). It is very likely and traditional to analyze injury severity utilizing police reported crash data. However, this police reported crash data is generally subjected to under reporting in case of minor or no personal injury as evidenced from a technical report by NHTSA, in 2009 that 25% of minor injury crashes and half of no injury crashes are unreported (Savolainen et al., 2011). In this study, we considered GES subset of 1,703 observations for single-vehicle (20.5%) and 6,588 observations multi-vehicle (79.5%) large-truck involved crashes over a period of four years (i.e., 2005 to 2008) from an annual average of 56,970 total crashes over this time period. Despite the issues of under reporting for minor and no personal injury along with the multi-stage sampling scheme in GES database, GES focuses on the crashes of the greatest concerns to the highway safety community and the general public (NASS-GES, 2008). As a result, GES is a representative sample of the crashes from police reports all over the US and it is a fairly common practice in the modeling approach to assume that sample data selected from the population has equal likelihood of being considered in the sample (Savolainen et al., 2011).

To investigate contributing factors relating to human, vehicle, and road-environment, two samples of 1,703 and 6,588 data observations representing crashes involving a large truck for single-vehicle collision and at least a large truck and other vehicles (i.e., number of vehicles involved is two or more than two) for multi-vehicle collision on the interstate highway system from 2005 to 2008 were extracted from the NASS-GES database. The maximum level of injury severity recorded in the vehicle or person dataset was aggregated to represent a crash. Each observation in the sample is a crash representing the maximum level of injury of the occupants, involving at least one large truck with one or more vehicles on interstate highways. The crash dataset was fused to the vehicle and person datasets

through appropriate linking variable and crash number, while the vehicle and person dataset were linked through the vehicle and crash number using the Statistical Analysis System (SAS). The mixed logit framework was modeled in Limdep (NLOGIT 4.0).

Descriptive statistics of the variables used in the models are presented in Table 4.1 (single-vehicle model) and 4.2 (multi-vehicle model).

Table 4.1: Descriptive Statistics of the Key Variables in Single-vehicle Model (N = 1,703)

Meaning of Variables in the Model	Mean	Std. Dev.	Severity outcome
Lap and shoulder restraint usage (1 if used by the occupants, 0 otherwise)	0.795	0.404	Fatal
The most harmful event in crash consequences (1 if rollover, 0 otherwise)	0.772	0.419	
Number of directional lane (1 if three lanes in each direction, 0 otherwise)	0.199	0.399	
Age group (1 if age group of 55 to 65, 0 otherwise)	0.119	0.323	Incapacitating Injury
The most harmful event in crash consequences (1 if rollover, 0 otherwise)	0.772	0.419	
Day of week (1 if Friday, 0 otherwise)	0.153	0.360	
Light condition (1 if dark but lighted outside, 0 otherwise)	0.149	0.356	
Previous citation records (1 if speed related violation, 0 otherwise)	0.171	0.377	
Alignment on highways (1 if curvature, otherwise)	0.341	0.474	Non-Incapacitating Injury
Number of directional lane (1 if three lanes in each direction, 0 otherwise)	0.199	0.399	
Gender of the occupants (1 if male, 0 otherwise) (standard error of parameter distribution)	0.941	0.236	
States of USA (1 if border states between US and Mexico, 0 otherwise)	0.281	0.449	
Time of day (1 if time between 10 am to 12 pm, 0 otherwise)	0.151	0.358	
State specific (1 if Texas, 0 otherwise)	0.108	0.310	Possible Injury
Surface of highway (1 if Level, 0 otherwise)	0.501	0.500	
Harmful vehicular event (1 if jackknife, 0 otherwise)	0.174	0.3799	
The most harmful event in vehicular event (1 if rollover, 0 otherwise)	0.772	0.419	
Driving along the roadway geometry (1 if negotiating a curve, 0 otherwise)	0.273	0.445	
Pre-crash maneuver (1 if running off the roadway left or right, 0 otherwise)	0.621	0.485	No-injury (PDO)
Month of year (1 if summer months, 0 otherwise)	0.251	0.4341	
Harmful vehicular event (1 if jackknife, 0 otherwise)	0.174	0.379	

Meaning of Variables in the Model	Mean	Std. Dev.	Severity outcome
Truck trailer configuration (1 if trailing only one unit, 0 otherwise)	0.757	0.429	

Table 4.2: Descriptive Statistics of the Key Variables in Multi-vehicle Model (N = 6,588)

Meaning of Variables in the Model	Mean	Std. Dev.	Severity outcome
Light condition (1 if there is no lighting, 0 otherwise)	0.109	0.312	Fatal
Time of day (1 if time between 1pm to 3 pm, 0 otherwise)	0.193	0.394	
Lap and shoulder restraint usage (1 if used by the occupants, 0 otherwise)	0.830	0.375	
Role of vehicle in crash (1 if the role of vehicle is being struck, 0 otherwise)	0.466	0.499	
The most harmful event in crash consequences (1 if rollover, 0 otherwise)	0.809	0.392	Incapaci-tating Injury
Time of day (1 if from 12 am to 5 am, 0 otherwise)	0.123	0.328	
Region specific (1 if US and Mexico border states, 0 otherwise)	0.225	0.418	
Roadway condition (1 if surface is dry, 0 otherwise)	0.777	0.416	
Role of vehicle in crash (1 if the role of vehicle is striking, 0 otherwise)	0.427	0.495	
Manner of collision (1 if sideswipe in the same direction, otherwise)	0.331	0.471	Non-incapacitating Injury
Pre-crash maneuver (1 if going straight – lane keeping driving behavior, 0 otherwise)	0.652	0.476	
Time of Day (1 if between 4 pm to 7 pm, 0 otherwise)	0.184	0.388	
Pre-crash maneuver (1 if lane changing, 0 otherwise)	0.118	0.323	
Gender of occupant (1 if Male, 0 otherwise)	0.938	0.240	Possible Injury
State specific (1 if Texas, 0 otherwise)	0.100	0.300	
Manner of collision (1 if angle collision, 0 otherwise)	0.124	0.329	
Surface condition of roadway (1 if wet surface, 0 otherwise)	0.152	0.359	
Age group (1 if age from 45 to 55 years, 0 otherwise)	0.264	0.441	
Months of year (1 if summer months (June to August), 0 otherwise)	0.235	0.424	No-injury (PDO)
Light condition (1 dark but lighted outside, 0 otherwise)	0.156	0.364	
Truck trailer configuration (1 if trailing only one unit, 0 otherwise)	0.751	0.432	
Collision manner (1 if rear-end collision, otherwise)	0.424	0.494	
Pre-crash maneuver (1 if changing lanes, 0 otherwise)	0.118	0.323	

4.5 Empirical results

A regular multinomial logit model was developed first then a mixed logit model was estimated using the maximum likelihood and simulation–based maximum likelihood methods for parameter vector ($\boldsymbol{\beta_i}$), respectively. First, a combined mixed logit model was developed, and then the sample was split between single and multi-vehicle sub-samples. In all three cases, non-injury (i.e., Property-Damage-Only) was taken as a base case with respect to other injury categories. Although triangular and uniform distribution were considered for the functional distribution of parameter density function for random parameter models, normal distribution was found to have the best statistical significance. All the variables estimated in the three models are statistically significant with 90% confidence level. The pseudo R-squared (McFadden R-squared) value for the combined, single and multi-vehicle injury model was estimated to be 0.617, 0.347, and 0.709, respectively. Turning to random parameter components in the models, one variable in single- and three constant terms in multi-vehicle injury models were found as normally distributed random parameters. For single-vehicle injury model, the only random parameter is male for non-incapacitating injury category. Also, for multi-vehicle injury model, three random parameters are constants for fatality, incapacitating, and non-incapacitating injury.

The findings are described in light of the contributing factors below and model results are presented in Table 4.3 (single-vehicle model) and 4.4 (multi-vehicle model):

Table 4.3: Model Estimates for Single-vehicle Model (N = 1,703)

Meaning of Variable	Estimate	t-stat	*p*-value
Fatal outcome			
Constant	-1.909	-4.532	0.000
Lap and shoulder restraint usage (1 if used by the occupants, 0 otherwise)	-1.605	-3.948	0.000
The most harmful event in crash consequences (1 if rollover, 0 otherwise)	-1.418	-3.438	0.000
Number of directional lane (1 if three lanes in each direction, 0 otherwise)	0.791	1.832	0.066
Age group (1 if age group of 55 to 65, 0 otherwise)	1.112	2.395	0.017
Incapacitating Injury Outcome			
Constant	-1.047	-4.549	0.000
The most harmful event in crash consequences (1 if rollover, 0 otherwise)	-1.108	-5.461	0.000
Day of week (1 if Friday, 0 otherwise)	0.446	1.899	0.058
Light condition (1 if dark but lighted outside, 0 otherwise)	-0.587	-1.691	0.090
Previous citation records (1 if speed related violation, 0 otherwise)	-0.696	-2.103	0.036
Non-incapacitating Injury Outcome			
Constant	-1.406	-3.272	0.001
Alignment on highways (1 if curvature, otherwise)	1.510	3.291	0.001
Number of directional lane (1 if three lanes in each direction, 0 otherwise)	-1.397	-2.086	0.037
Gender of the occupants (1 if male, 0 otherwise) (standard error of parameter distribution)	-4.715 (5.813)	-2.210 (2.502)	0.027
Region specific (1 if US and Mexico border states, 0 otherwise)	0.991	2.043	0.041
Time of day (1 if time between 10 am to 12 pm, 0 otherwise)	0.841	1.829	0.067
Possible Injury Outcome			
Constant	-1.840	-6.556	0.000
State specific (1 if Texas, 0 otherwise)	0.899	4.247	0.000
Surface of highway (1 if Level, 0 otherwise)	-0.536	-3.282	0.001
Harmful vehicular event (1 if jackknife, 0 otherwise)	0.596	1.827	0.068
The most harmful event in vehicular event (1 if rollover, 0 otherwise)	0.473	2.201	0.027
Driving along the roadway geometry (1 if negotiating a curve, 0 otherwise)	0.617	3.680	0.000
Non-Injury Outcome (Property-Damage-Only)			
Pre-crash maneuver (1 if running off the roadway left or right, 0 otherwise)	-0.439	-3.320	0.001
Month of year (1 if summer months, 0 otherwise)	-0.415	-3.053	0.002
Harmful vehicular event (1 if jackknife, 0 otherwise)	0.888	3.263	0.001
Truck trailer configuration (1 if trailing only one unit, 0 otherwise)	0.671	4.906	0.000
Number of observations		1,703	
Restricted log-likelihood		-2740.873	
Log-likelihood at convergence		-1790.035	
Chi-squared value		1901.676	
McFadden pseudo R-squared (ρ^2)		0.347	

Table 4.4: Model Estimates for Multi-vehicle Model (N = 6,588)

Meaning of Variable	Estimate	t-stat	*P*-Value
Fatal outcome			
Constant (standard error of parameter distribution)	-12.244 (5.741)	-2.922 (2.768)	0.004
Light condition (1 if there is no lighting, 0 otherwise)	4.219	2.721	0.007
Time of day (1 if time between 1 pm to 3 pm, 0 otherwise)	2.141	1.877	0.061
Lap and shoulder restraint usage (1 if used by the occupants, 0 otherwise)	-4.761	-3.102	0.002
Role of vehicle in crash (1 if the role of vehicle is being struck, 0 otherwise)	-3.594	-2.295	0.022
Incapacitating Injury Outcome			
Constant (standard error of parameter distribution)	-5.422 (2.365)	-3.352 (2.237)	0.001
The most harmful event in crash consequences (1 if rollover, 0 otherwise)	-1.546	-3.712	0.000
Time of day (1 if from 12 am to 5 am, 0 otherwise)	1.442	3.529	0.000
Region specific (1 if US-Mexico border states, 0 otherwise)	-0.849	-2.181	0.029
Roadway condition (1 if surface is dry, 0 otherwise)	0.662	1.809	0.071
Role of vehicle in crash (1 if the role of vehicle is striking, 0 otherwise)	0.715	2.686	0.007
Non-incapacitating Injury Outcome			
Constant (standard error of parameter distribution)	-6.606 (3.419)	-2.653 (2.158)	0.008
Manner of collision (1 if sideswipe in the same direction, otherwise)	-1.149	-2.020	0.043
Pre-crash maneuver (1 if going straight – lane keeping driving behavior, 0 otherwise)	0.717	2.049	0.041
Time of Day (1 if between 4 pm to 7 pm, 0 otherwise)	-0.760	-1.786	0.074
Pre-crash maneuver (1 if lane changing, 0 otherwise)	0.912	1.709	0.087
Possible Injury Outcome			
Constant	-2.154	-9.842	0.000
Gender of occupant (1 if male, 0 otherwise)	-0.452	-2.292	0.022
State specific (1 if Texas, 0 otherwise)	0.779	5.416	0.000
Manner of collision (1 if angle collision, 0 otherwise)	0.423	2.486	0.013
Surface condition of roadway (1 if wet surface, 0 otherwise)	0.452	3.297	0.001
Age group (1 if age from 45 to 55 years, 0 otherwise)	-0.227	-1.728	0.084
Non-Injury Outcome (Property-Damage-Only)			
Months of year (1 if summer months (June to August), 0 otherwise)	-0.281	-2.691	0.007
Light condition (1 dark but lighted outside, 0 otherwise)	0.393	2.241	0.025
Truck trailer configuration (1 if trailing only one unit, 0 otherwise)	0.804	7.826	0.000
Collision manner (1 if rear-end collision, otherwise)	-0.440	-4.048	0.000
Pre-crash maneuver (1 if changing lanes, 0 otherwise)	0.519	2.582	0.009
Number of observations		6,588	
Restricted log-likelihood		-10602.98	
Log-likelihood at convergence		-3085.250	
Chi-squared value		15035.45	
McFadden pseudo R-squared (ρ^2)		0.709	

In the single-vehicle model, the variable "male" for non-incapacitating injury is normally distributed with a mean of 5.813 and a standard deviation of 2.502. With these estimates, the constant term is more than zero for 98.98% of the large-truck involved crashes. This variability is likely capturing the physical tolerance of the body in response to crash/injury in the severity outcomes that could include

factors such as age, injury mechanism, seating arrangements, and crash types which were not directly measured in the dataset for this model. Likewise, the constant term in the multi-vehicle model for fatal, incapacitating and non-incapacitating injury outcome is normally distributed with means of 5.741, 2.365, and 3.419, respectively and standard deviations of 2.768, 2.237, and 2.158, respectively. With these estimates, the constant term is less than zero for 1.92%, 14.46%, and 5.71% of the large-truck involved crashes. This variability is likely capturing the unobserved heterogeneity (i.e., unobserved factors in the crashes) in the severity outcomes that could include factors such as traffic condition and roadway characteristics which was not directly measured in the dataset.

The elasticities of the variables (i.e., average direct pseudo elasticities) for both single- and multi-vehicle severity models are presented in Table 4.5 (single-vehicle model) and Table 4.6 (multi-vehicle model).

Table 4.5: Elasticities of Single-vehicle Model

Variables	Elasticity (%)				
	PDO/No Injury	Possible Injury	Non-incapacitating Injury	Incapacitating Injury	Fatal
Lap and shoulder restraint usage (1 if used by the occupants, 0 otherwise) [F]	1.22	1.22	0.37	2.10	-69.73
The most harmful event in crash consequences (1 if rollover, 0 otherwise) [F]	1.21	1.39	0.35	0.97	-65.35
Number of directional lane (1 if three lanes in each direction, 0 otherwise) [F]	-0.44	-0.42	-0.10	-0.71	24.49
Age group (1 if age group of 55 to 65, 0 otherwise) [F]	-0.47	-0.47	-0.18	-0.75	26.92
The most harmful event in crash consequences (1 if rollover, 0 otherwise) [INCAP]	5.22	6.02	1.55	-55.88	3.61
Day of week (1 if Friday, 0 otherwise) [INCAP]	-0.74	-0.72	-0.27	8.03	-1.35
Light condition (1 if dark but lighted outside, 0 otherwise) [INCAP]	0.42	0.41	0.12	-4.49	0.78
Previous citation records (1 if speed related violation, 0 otherwise) [INCAP]	0.50	0.59	0.16	-5.44	0.64
Alignment on highways (1 if curvature, otherwise) [NONINCAP]	-3.13	-4.88	17.03	-3.40	-4.07
Number of directional lane (1 if three lanes in each direction, 0 otherwise) [NONINCAP]	1.07	1.10	-5.41	1.12	1.81
Gender of the occupants (1 if male, 0 otherwise) [NONINCAP]	-3.25	-1.65	14.52	-2.66	-3.16
States of USA (1 if border states between US and Mexico, 0 otherwise) [NONINCAP]	-1.61	-2.26	8.59	-1.94	-1.62
Time of day (1 if time between 10 am to 12 pm, 0 otherwise) [NONINCAP]	-0.77	-1.02	4.05	-0.92	-0.73
State specific (1 if Texas, 0 otherwise) [POSS]	-1.90	11.36	-0.84	-1.96	-1.69
Surface of highway (1 if Level, 0 otherwise) [POSS]	2.92	-17.03	0.86	3.16	2.81
Harmful vehicular event (1 if Jackknife, 0 otherwise) [POSS]	-1.29	6.76	-0.28	-0.50	-0.51
The most harmful event in crash consequences (1 if rollover, 0 otherwise) [POSS]	-5.61	31.71	-1.78	-4.35	-3.74
Driving along the roadway geometry (1 if negotiating a curve, 0 otherwise) [POSS]	-3.01	17.91	-1.35	-2.78	-2.96
Pre-crash maneuver (1 if running off the roadway left or right, 0 otherwise) [PDO]	-8.37	18.25	5.19	19.96	20.23
Month of year (1 if summer months, 0 otherwise) [PDO]	-3.27	7.45	2.00	7.57	6.73
Harmful vehicular event (1 if jackknife, 0 otherwise) [PDO]	3.58	-8.98	-2.78	-5.88	-5.92
Truck trailer configuration (1 if trailing only one unit, 0 otherwise) [PDO]	13.95	-31.67	-10.15	-28.85	-29.41

Table 4.6: Elasticities for Multi-vehicle Model

Variables	Elasticity (%)				
	PDO/No Injury	Possible Injury	Non-incapacitating Injury	Incapacitating Injury	Fatal
Light condition (1 if there is no lighting, 0 otherwise) [F]	-0.26	-0.28	-0.22	-0.38	53.52
Time of day (1 if time between 1pm to 3 pm, 0 otherwise) [F]	-0.12	-0.13	-0.10	-0.12	24.12
Lap and shoulder restraint usage (1 if used by the occupants, 0 otherwise) [F]	0.42	0.45	0.34	0.45	-86.80
Role of vehicle in crash (1 if the role of vehicle is being struck, 0 otherwise) [F]	0.13	0.15	0.09	0.12	-25.97
The most harmful event in crash consequences (1 if rollover, 0 otherwise) [INCAP]	1.49	1.62	1.03	-72.78	0.89
Time of day (1 if from 12 am to 5 am, 0 otherwise) [INCAP]	-0.48	-0.45	-0.36	23.41	-0.49
Region specific (1 if US and Mexico border states, 0 otherwise) [INCAP]	0.20	0.30	0.15	-10.07	0.11
Roadway condition (1 if surface is dry, 0 otherwise) [INCAP]	-0.83	-0.87	-0.60	40.44	-0.54
Role of vehicle in crash (1 if the role of vehicle is striking, 0 otherwise) [INCAP]	-0.56	-0.60	-0.41	27.48	-0.47
Manner of collision (1 if sideswipe in the same direction, otherwise) [NONINCAP]	0.49	0.37	-12.67	0.32	0.21
Time of Day (1 if between 4 pm to 7 pm, 0 otherwise) [NONINCAP]	0.21	0.25	-5.61	0.17	0.10
Pre-crash maneuver (1 if going straight, 0 otherwise) [NONINCAP]	-1.03	-1.20	27.65	-1.04	-0.68
Pre-crash maneuver (1 if lane changing, 0 otherwise) [NONINCAP]	-0.16	-0.10	4.10	-0.09	-0.10
Gender of occupant (1 if male, 0 otherwise) [POSS]	2.32	-38.05	1.52	1.91	1.00
State specific (1 if Texas, 0 otherwise) [POSS]	-0.77	12.58	-0.52	-0.47	-0.36
Manner of collision (1 if angle collision, 0 otherwise) [POSS]	-0.33	5.40	-0.22	-0.22	-0.12
Surface condition of roadway (1 if wet surface, 0 otherwise) [POSS]	-0.52	8.46	-0.33	-0.32	-0.21
Age group (1 if age from 45 to 55 years, 0 otherwise) [POSS]	0.28	-4.56	0.18	0. 22	0. 12
Months of year (1 if summer months (June to August), 0 otherwise) [PDO]	-0.73	7.07	3.53	5.21	2.61
Light condition (1 if dark but lighted outside, 0 otherwise) [PDO]	0.46	-3.94	-2.59	-4.17	-1.29
Truck trailer configuration (1 if trailing only one unit, 0 otherwise) [PDO]	4.76	-42.41	-27.52	-35.68	-21.11
Collision manner (1 if rear-end collision, otherwise) [PDO]	-2.08	19.65	11.47	14.07	7.12
Pre-crash maneuver (1 if changing lanes, 0 otherwise) [PDO]	0.37	-3.09	-2.19	-2.93	-2.85

4.5.1 Human factors

Single-vehicle model: Considering the human factors in a single-vehicle model, male occupants are less likely to be involved in non-incapacitating injury. A similar study by Chen and Chen (2011) showed that female drivers are more likely to be involved in fatal or incapacitating injury. Also, using lap and shoulder belts reduces the likelihood of fatality by 69.7%. This fact is indirectly supported by Chen and Chen (2011) that not wearing a seat-belt by the drivers result in both fatal and incapacitating and non-incapacitating or possible injury. Drivers of age group between 55 to 65 years are more likely to be involved in fatalities which increase by 26.9%. A study conducted by the Trauma Center of the University of Michigan (Wang, 1998) shows that older drivers (i.e., 65 and older) are at higher risk of chest injuries because of the greater likelihood of being involved in driver-side-impact crashes, coupled with higher percentages of restraint use where the chest is the major load-bearing area for restraint systems (NHTSA, 2006). Drivers' residing or working in the vicinity of a particular region has effects on injury outcome. This study found that drivers residing or registered in the state of Texas are more likely to be involved in single vehicle collisions resulting in possible injuries which increase by 11.4%. Likewise, drivers working or living in the proximity of US-Mexican Border States are more likely to be involved in single vehicle collisions resulting in non-incapacitating injuries. An earlier study by Smith and O'Day (1982) on single-vehicle passenger car crashes in Texas for a 2-year period (1975-76) found that young drivers were much more likely to be in smaller cars, whereas, older drivers were more likely to be in larger cars. This relationship interacts with the fatality likelihood estimation, because older persons are more likely to be killed than young persons in a given crash situation. Drivers with previous speed citations are less likely to be involved in incapacitating injuries. This driving behavior clearly indicates that they are more concerned to avoid more citations and drive within the speed limit. Driving along the curved highway sections increases the likelihood of possible injuries by 17.9%. This result is also supported by Savolainen and Tarko (2005) with the negative binomial model showing that the

intersections located on curves experience a greater proportion of severe crashes than the intersections in the tangent sections.

Multi-vehicle model: Similar to single-vehicle model, male occupants are less likely to be involved in possible injuries which reduced by 38% as supported by what Chen and Chen (2011) found to be significant in both fatal or incapacitating and non-incapacitating or possible injuries. Restraint usage is also found to be life saving in a multi-vehicle model. Lap and shoulder seat-belt by the occupants resulted in less likelihood of fatalities which reduced by 86.8%. Similar findings by Chen and Chen (2011) also indicated the fact of not wearing a seat belt resulted in fatal or incapacitating and non-incapacitating or possible injuries. Drivers of age group of 45 to 55 years are less likely to be involved in possible injuries. This might capture the experienced driver groups of all vehicles. Drivers living and working in particular region also have significant effects on injury outcomes. As such, drivers residing or registered in the state of Texas are more likely to be involved with possible injuries which increase by 12.6%. Unlike a single-vehicle model, drivers working or residing in the proximity of US-Mexican Border States are less likely to be involved in multi-vehicle collisions resulting in incapacitating injuries which reduce by 10.1%.

4.5.2 Road and environmental factors

Single-vehicle model: Roadway geometrics and environment (lighting condition, time of day, day of week, and month of year, etc.) have also significant effects on injury outcomes. For example, curved segments result in non-incapacitating injuries with an increase of 17%. A similar study by Chen and Chen (2011) also indicated sharp curves (degree of curve more than 5) increase the likelihood of both fatal or incapacitating and non-incapacitating or possible injuries. Three lanes in each direction of highway experience single vehicle crashes resulting in less non-incapacitating injuries but in more fatalities by 24.5%. This is because most truck drivers intend to drive on the right-most lane or

sometimes pass other vehicles on the middle lane. However, it could lead to running off the roadway and hitting the roadside fixed objects resulting in fatalities. A study by Milton and Mannering (1998) found that number of crashes increase with the increase of number of lanes. Level road alignment decreases the likelihood of possible injuries by 17% which is captures the better visibility in the driving environment. Dark but lighted condition decreases the likelihood of incapacitating injuries because of professional drivers' training and caution under such situations. The time period between 10 am to 12 pm increases the likelihood of non-incapacitating injuries. This is because of non-peak hours and traffic on the highways is relatively lighter than peak hours for the large truck drivers to drive safely. A study by Qin et al. (2006) indicated that with increasing the traffic density (vehicle/hour), the likelihood of single vehicle occurrence decrease. It is related to time period of 10 am to 12 pm indicating non-peak hours and more single-vehicle crash occurrence. Crashes occurring on Fridays are more likely to be involved in incapacities injuries. On the start of weekend (i.e., Fridays), there could be detours from the regular work-routes for large trucks for some leisure or rest (e.g., entertainment) which might cause such incidents. Summer months from June to August also indicate less likelihood of non-injury crashes which indirectly implies more likelihood of serious injuries.

Multi-vehicle model: Dark condition leads to greater likelihood of fatalities which increase by 53.5% since other vehicles might be completely blinded by such unfavorable driving conditions. A similar result was also found by Morgan and Mannering (2011) that male of less than 45 years old have more likelihood of fatalities driving on dry surface in dark condition. In addition, the dark but lighted condition also leads to non-injury crashes. This fact is supported by Chen and Chen (2011) where dark but lighted condition increases both fatal or incapacitating injuries and non-incapacitating or possible injuries. The time period from 1 pm to 3 pm in the afternoon also increases the likelihood of fatalities which increase by 24.1%. This time period clearly indicates some drowsy driving by other vehicles (other than trucks) because of some after-lunch effects. Similarly, the time period from 12 am to 5 am

also increases the likelihood of incapacitating injuries by 23.4%, which clear indicates sleepy or drowsy driving condition. A similar result was found in a study by Morgan and Mannering (2011) that male drivers less than 45 years old driving on a snowy surface have more likelihood of fatalities. However, the time period from 4 pm to 7 pm reduces the likelihood of non-incapacitating injuries because of evening peak hours where congestions and slow vehicle movements are experienced. Summer months (from June to August) decreases the likelihood of no injuries crashes because of more traffic on the highways and great chances of interaction of vehicles leading to collisions which might lead to severe injuries. Dry pavement condition increases likelihood of incapacitating injuries by 40.4% because of the higher traveling speed assuming better friction between tires and pavement in case of emergency braking. Similarly, wet pavement condition increases the likelihood of possible injuries because of unfavorable slippery road condition for other vehicles and also different braking characteristics of large trucks and smaller passenger vehicles.

4.5.3 Vehicular factors

Single-vehicle model: Trucks with single trailer are more likely to be involved with no-injuries by 14% which indirectly indicate that they are less likely to be involved in serious injuries.

Multi-vehicle model: Like the single-vehicle model, trucks with single trailer are more likely to be involved with no-injuries which indirectly indicate that they are less likely to be involved in serious injuries. A similar result was also found by Islam and Hernandez (2011) that trucks with a single trailer are less likely to be involved in fatalities. Vehicular dynamics in terms of vehicular role in the crashes is crucial in injury outcomes. Large trucks being actively hit by other vehicles increase the likelihood of incapacitating injuries; whereas, other vehicles hitting large trucks decrease likelihood of fatalities.

4.5.4 Crash characteristics/mechanism

Single-vehicle model: Rollover decreases the likelihood of fatalities and incapacitating injuries by 65.4% and 55.9%, respectively, but increases likelihood of possible injuries by 31.7%. This fact captures the situation that the truck drivers are highly protected inside the occupant compartment. However, this fact is contradicted by a study by Chen and Chen (2011) where vehicle overturn increases the likelihood of both fatal or incapacitating and non-incapacitating or possible injuries. Departing the roadway (by left or right side of roadway) decreases the likelihood of non-injury crashes which indirectly implies more likelihood of severe injuries. A similar finding by Chen and Chen (2011) also indicates that running-off-the roadway increases both fatal or incapacitating and non-incapacitating or possible injuries. Jackknife for large trucks increases the likelihood of possible injuries and non-injuries (property-damage-only) which is along the same lines as a study by Chen and Chen (2011) where jackknife decreases the likelihood of fatal or incapacitating injuries.

Multi-vehicle model: Like a single-vehicle collision, a rollover situation decreases the likelihood of incapacitating injuries by 72.8% which is complex in nature for multi-vehicle collision dynamics. This fact is contradicted by Chen and Chen (2011) findings on truck overturn collisions. Other types of collisions such as rear-end, lane changing, going straight, and angle has significant impact on injury outcomes for multi-vehicle collisions. Rear-end collision decreases the likelihood of non-injuries crashes but increases the likelihood of severe injuries. Sideswipe in the same direction decreases the likelihood of non-incapacitating injuries by 12.7%. Lane changing behavior increases the likelihood of non-incapacitating injuries by 4.1% and non-injury crashes which are obviously true for the situation of multiple vehicle interactions. Going straight along the lane increases the likelihood of non-incapacitating injuries by 26.7%. Angle collision increases the likelihood of possible injuries by 5.4%.

4.6 Model specification tests

The likelihood ratio test was conducted to justify the statistical significance of single- and multi-vehicle injury model separately from the combined model. The null hypothesis is the separate models are not statistically and significantly different from the combined model. The following likelihood ratio test was conducted to test the hypothesis (Washington et al., 2011):

$$\chi 2 = -2 * [LL_T(\beta^C) - LL_S(\beta^S) - LL_M(\beta^M)] \tag{4.7}$$

where,

$LL_T(\beta^C)$ $is\ loglikelihood\ at\ convergence\ of\ the\ combined\ model$

$(-5105.041, degree\ of\ freedom, n_T = 29)$

$LL_S(\beta^S)$ $is\ loglikelihood\ at\ convergence\ of\ single - vehicle\ model$

$(-1790.035, degree\ of\ freedon, n_S = 27)$

$LL_M(\beta^M)$ $is\ loglikelihood\ at\ convergence\ of\ multi - vehicle\ model$

$(-3085.250, degree\ of\ freedom\ n_M = 31)$

The test follows χ^2 distribution with degrees of freedom equal to the sum of single- and multi-vehicle model minus that of total model (combined of single and multi-vehicle model). With 29 degrees of freedom and χ^2 value of 459.512 the test statistic indicates that we reject the null hypothesis. This leads to the fact that separate models (i.e., single- and multi-vehicle models) are statistically and significantly different from the combined model with *P*-value being less than 0.0001 (two-tailed *P*-value). So, the test result indicates that significance of injury severity should be modeled as single- and multi-vehicle models separately rather than combined one.

Then, we also conducted the log likelihood ratio to test whether random parameter models are statistically significant from their corresponding base models (i.e., fixed parameter multinomial models from all three models – combined, single and multi-vehicle injury models). The null hypothesis is that a mixed logit model is not statistically and significantly different from multinomial logit model. The following likelihood ratio test was adopted (Washington et al., 2011):

$$\chi 2 = -\ 2 * [LL_{MNL}(\beta) - LL_{MXL}(\beta)] \tag{4.8}$$

where, $LL_{MNL}(\beta)$ and $LL_{MXL}(\beta)$ are log-likelihood at convergence of multinomial logit and mixed logit model of the combined model, single- and multi-vehicle injury models. The test statistic is χ^2 distributed with degree of freedom is the number of parameters to be random in the mixed logit model. So, for the combined model χ^2 is 80.862 with degree of freedom 4, single- and multi-vehicle injury model has χ^2 of 16.592 and 17.378 with degree of freedom 1 and 3, respectively which leads to rejection of the null hypothesis. The corresponding two-tailed P-value for combined, single- and multi-vehicle injury model is less than 0.0001, 0.0001, and equal to 0.0006, respectively. Therefore, it is clear that there is statistically significant difference between random and fixed parameter models. This clearly indicates that it is better to model mixed logit where we can address heterogeneity in data reporting system and unknown and unobserved factors in the at-scene crash investigation by the police officers.

4.7 Discussions and conclusions

This study provides a demonstration of the mixed logit model as a viable methodological approach to gain some new insights into the factors that significantly influence injury severity for single- and multi-vehicle large trucks involved crashes over a period from 2005 to 2008 on the US interstate system. This random parameter mixed logit formulation allows to correct for unobserved heterogeneity that could result from data limitations or sampling from total population of crash statistics, the factors such as human, vehicle, road-environment, crash mechanism or any variations in the police reporting not fully captured because of discretion of the investigating police officers and the factors not obviously identified from the in-depth crash investigation and reconstruction studies.

Using four years of a nationally representative subsample of crash data maintained by NHTSA, the model results provide some interesting findings on injury severity of large-truck involved crashes. Although traffic data has not been incorporated in the dataset for the developed models, there

are variables representing the time of the day (i.e., 12 am to 5 am, 10 am to 12 pm, 1 pm to 3 pm, 4 pm to 7 pm), day of week (i.e., Friday), and months of the year (i.e., June to August) that serve as a proxy for traffic condition on the highway system. In addition, in the single-vehicle model, “male” variable is found to be a random parameter which indicates the age, perception of risks, driving experience and behavior, injury mechanism, seating arrangements, and crash types which are not captured in the crash data, that could influence male drivers or passengers to be less likely involved in severe injuries. On other hand, three constant terms of specific to fatal, incapacitating and non-incapacitating injury outcomes are random in nature. These random parameters very likely to indicate the traffic condition particularly the intensity of traffic which is a crucial factor for multi-vehicle collisions on the highway systems. Since the methodological approach particularly the correlations of the error terms have been greatly relaxed, the results obtained from the mixed logit could be highly beneficial to identify the contributing factors. Understanding those factors could highly benefit the trucking companies and transportation agencies to devise the truck safety strategies in terms of drivers’ safety training, highway design (e.g., warning signs), and vehicular design.

Compared to combined model, single- and multi-vehicle injury models were segregated which provide some more insights through important statistically significant variables specific to single- and multi-vehicle collisions. It is shown by statistical tests (i.e., likelihood ratio test) that modeling separately by single- and multi-vehicle provides better understanding of the factors which were not properly captured in the combined model. Similarly, random parameter logit (i.e., mixed logit) models are more statistically significant than their corresponding multinomial logit models.

There are variables which are very specific to single-vehicle collisions such as running-off the roadway, jackknifing, and negotiating along the curves that are also found to be statistically significant in the combined model. Likewise, variables such as vehicle being struck by other vehicles, going straight, rear-end collisions, time of day like 12 am to 5 am and 4 pm to 7 pm are specific to multi-

vehicle collisions and also found statistically significant in the combined model. On the other hand, there are variables that are significant in both models (i.e., single- and multi-vehicle models) such as male occupants (i.e., drivers or passengers), crashes occurring in Texas, trucks with single trailer, summer months (i.e., June to August), and rollover.

In the future, it could be more interesting to consider spatial and temporal characteristics of single- and multi-vehicle collisions. Urban and rural collisions could capture some of the spatial characteristics of single- and multi-vehicle collisions. Similarly, peak hours and non-peaks hours could also unveil some of the interesting findings towards single- and multi-vehicle collisions.

Chapter 5: Modeling Injury Outcomes of Crashes involving Heavy Vehicles on Texas Highways

5.1 Introduction

A growing concern related to large-truck crashes has increased in the State of Texas (and globally) in recent years due to the potential economic impacts and level of injury severity that can be sustained (United Nations, 2010). These concerns are well founded since the total economic losses due to vehicular crashes were estimated to be $107.4 billion dollars between 2006 and 2010 in the US—in 2008, the State of Texas faced the greatest economic losses of any state at $22.9 billion dollars (NHTSA, 2009; TxDOT, 2011; IIHS, 2011). Statistically, recent data indicated that in 2010 the State of Texas experienced 3,023 deaths along with 59,660 serious injury crashes which resulted in 82,685 people sustaining serious injuries (TxDOT, 2011). Consequently, any increase in the number and level of crash injury severity is of great concern to transportation organizations that operate, maintain, and construct the Texas transportation system as well as to trucking companies (NIOSH, 2007; USDOT, 2005).

Therefore, to better understand the safety impacts related to large trucks on the Texas highway system, tools need to be developed that can aid transportation safety professionals as well as trucking industry operations managers in the avoidance and mitigation of large-truck involved crashes. Currently, severity prediction models are still regarded as the primary tools (Geedipally and Lord, 2010) to understand and estimate factors involving vehicular crashes. However, very few studies have investigated injury severities associated with large-truck involved crashes, especially utilizing state specific crash databases (Islam and Hernandez, 2011; Islam and Hernandez, 2012).

With this in mind, the main objective of this study is to analyze the contributing factors related to injury severity by utilizing Texas's crash database and applying the data to a discrete outcome model. This discrete outcome model accounts for possible unobserved heterogeneity (i.e., unobserved factors that may influence an injury outcome) related to human, vehicle, and road-environment. We estimate a random parameter logit model (i.e., mixed logit) to predict the likelihood of five standard injury severity scales commonly used in Crash Records Information System (CRIS) in Texas – fatal, incapacitating, non-incapacitating, possible, and no injury (property damage only), a first for large-truck involved crash modeling. To the best of the authors' knowledge these are the first attempts at modeling large truck injury severity utilizing the CRIS dataset. Although the mixed logit model has been applied to large-truck crash severity from different modeling perspectives, this research extends the current literature and introduces additional significant variables in regards to large-truck crashes.

The remainder of this paper is organized as follows. Section 6.2 reviews literature related to crash models in general and the mixed logit model with respect to large truck crashes. Section 3 discusses the empirical setting and descriptive statistics. Section 6.4 describes the methodological approach. Section 6.5 summarizes and discusses the insights from the empirical results and Section 6. 6 presents some concluding comments.

5.2 Background of crash models

5.2.1 Crash models in general

Crash frequency, likelihood and severity modeling approaches have been widely used in traffic safety analyses. The most frequently applied models relate to crash frequency models such as Negative Binomial and Poisson models (Shanker et al., 1995; Poch and Mannering, 1996; Abdel-Aty and Radwan, 2000; Savolainen and Tarko, 2005), zero-inflated Poisson and zero-inflated Negative Binomial models (Shanker et al., 1997; Carson and Mannering, 2001; Lee and Mannering, 2002), random parameter Negative Binomial models (Shanker et al., 1998; Chin and Quddus, 2003; Anastasopoulos

and Mannering, 2009), Markov switching of two different state of crash occurrence (Malyshkina and Mannering, 2009), and Bayesian statistics on Negative Binomial models (Park et al., 2010). Also, crash likelihood and severity models have been studied, but are relatively sparse when it comes to analyzing large truck crashes (Patil et al., 2012; Geedipally and Lord, 2010; Golob et al., 1987; Islam and Mannering, 2006; Kim et al., 2010; Moore et al., 2011; Malyshkina and Mannering, 2010; Shanker and Mannering, 1996; Ulfarsson and Mannering, 2004).

5.2.2 Mixed logit model

The modeling of crash frequency and injury severity conditioned on the crash occurrence is not new. Yet, research efforts have mainly focused on modeling crash frequency by considering the severity at each specific severity level simultaneously (Milton et al., 2008). Nonetheless, this modeling approach significantly introduces potential estimation error due to the correlation among specific severity crash counts – fatal, incapacitating, non-incapacitating, possible injury, and no-injury (Milton et al., 2008; Anastasopoulos and Mannering, 2011). To overcome such limitations in the modeling approach, Milton et al. (2008) applied a mixed logit model where the proportion of crashes of each severity level on a specific roadway segment over a specified time period was analyzed.

From the perspective of traffic safety, the mixed logit approach has been successfully applied to determine the likelihood of seat-belt usage in the presence of passengers for single and multi-occupant vehicles (Gkritza and Mannering, 2008) where data limitations could be a potential hindrance to select other modeling approaches. Additionally, this approach has been found to be very promising in the modeling of pedestrian injury severities for pedestrian-vehicle interactions (Kim et al., 2010). The flexibility of this modeling approach makes it attractive because it accounts for variation over the observations known as unobserved heterogeneity, indicating that units of observation are more different than the descriptive variables (Kim et al., 2010) that may be present in limited data such as roadway

geometrics, pavement condition and general weather and traffic characteristics (Anastasopoulos and Mannering, 2011).

A study by Anastasopoulos and Mannering (2011) applied a mixed logit model to injury categories; however, they combined the intermediate injury categories such as incapacitating, non-incapacitating and possible injury to be injury, and considered fatal and no-injury separately that result in three injury levels (i.e., fatal, injury and no-injury). Likewise, Chen and Chen (2011) applied a mixed logit model to injury severities for truck drivers for single and multi-vehicle collision separately rather than a combined model in rural settings in Illinois over a time period of ten years (1991-2000). In addition, they combined possible injury and non-incapacitating injury into one category, incapacitating injury and fatality into another category and no-injury separately, resulting in three response outcome variables. However, this is not the case in our study where we considered all five injury levels separately – fatal, incapacitating, non-incapacitating, possible and no-injury as separate response outcome variables in the mixed logit formulation.

5.3 Empirical settings

The data used in this study were collected from the Texas Peace Officers' Crash Reports, commonly known as the CRIS database. To investigate human, vehicle, and road-environmental factors, a sample of 20,495 data observations were extracted from the CRIS database by filtering crashes involving large trucks on the Texas interstate system over a period of five years from 2006 to 2010. In the data processing stage, the vehicle body style (i.e., VEH_BODY_STYL_ID in the main data recording system) was set to truck, truck tractor, semi-trailers and highway facility type and processed using the SAS statistical software (SAS, 2011). Each observation in the sample is a crash representing the maximum level of injury sustained by the drivers, involving at least one large truck in the interstate system. The crash dataset was linked to the vehicle and person datasets through appropriate linking variables, namely the crash number, and the vehicle and person datasets were linked through the vehicle

and crash number using the SAS statistical software. The linking of the three data components (i.e., crash, vehicle, and person) was processed to have a single observation with a maximum injury level sustained by the drivers involved the crash (done through the Crash ID).

Table 5.1 presents the descriptive statistics of the variables used in the model (i.e., dependent variable – MAX_SEVS and all the independent variables). As seen from Table 5.1, the dependent variable (i.e., MAX_SEVS) has five levels of injury severity outcomes: fatality, incapacitating, non-incapacitating, possible, and no injury (or PDO) which accounted for about 0.38% (N=78), 2.26% (N=463), 2.93% (N=601), 5.46% (N=1,120), and 88.96% (N=18,223) of the total observations (N=20,495), respectively. By not grouping injury severity categories we can gain further insights on the contributing factors influencing each injury outcome sustained in large truck crashes.

With regards to spatial characteristics, rural and urban settings accounted for about 26.8%, and 56.4% of the total observations, respectively. For temporal characteristics, time of day such as 3-7 pm and 12-6 am accounted for 20.3% and 15.9% of the total observations, respectively. In addition, months of the year such as summer (June to August), fall (September to December), and spring (January to May) accounted for 25.6%, 31.8%, and 42.5% of the total observations, respectively. Yet, the summer and fall month's variables were the only variables found to be statistically significant in the model. Exposure variables dealing with traffic flow in terms of Average Daily Traffic (i.e., ADT) had a mean value of 15,397 vehicles per day per lane. While the variable dealing with ADT more than 2,000 vehicles per day per lane was found to be statistically significant and accounted for 98.5% of the total observations.

The geometry of highway such as the average right shoulder (i.e., outer shoulder) width of 19.7 feet, and 4 lanes in each direction accounted for about 42.8% of the total observations. Environmental aspect at the time of a crash such as dry surface, clear weather and dark highway sections (lighted outside) were 81.0%, 68.2%, and 11.9%, of the total observations, respectively. In regards to

demographics of the occupants, male occupants accounted for 94.4% of the total observations. Furthermore, two age groups of 25-35 years and 45-55 years, accounted for 17.8%, and 23.9% of the total observations, respectively. These variables were also found to be statistically significant in the model. But, other age groups such as 35-45 years and 55-65 years, accounted for 26.0%, and 12.7%, of the total observations, respectively, but were not found to be statistically significant.

Table 5.1: Descriptive Statistics of Key Variables (N = 20,495)

Meaning of Variables in the Model	**Mean**	**Std. Dev.**	**Severity outcome**
Terrain of roadway (1 if level roadway surface, 0 otherwise)	0.764	0.425	Fatal
Time of day (1 if between 3 pm to 7 pm, 0 otherwise)	0.203	0.402	
Weather condition (1 if clear weather, 0 otherwise)	0.682	0.466	
Land-use pattern at crash location (1 if rural area, 0 otherwise)	0.268	0.443	
Traffic flow at the time of crash (ADT per lane in each direction – veh/day/ln)	15397.3	8845.23	
Month of the year (1 if summer months (June to August), 0 otherwise)	0.257	0.437	Incapacitating Injury
Weather condition (1 if clear weather, 0 otherwise)	0.682	0.466	
Shoulder width (right shoulder width (ft))	19.727	3.555	
Time of the day (1 if between 3 pm to 7 pm, 0 otherwise)	0.203	0.402	
Land-use pattern at crash location (1 if urban area, 0 otherwise)	0.564	0.496	
Number of lanes on highways (1 if 4 lanes in one direction, otherwise)	0.428	0.495	Non-Incapacitating Injury
Time of the day (1 if between 12 am to 6 am, 0 otherwise)	0.159	0.366	
Months of year (1 if from September to December, 0 otherwise)	0.241	0.428	
Land-use pattern at crash location (1 if rural area, 0 otherwise)	0.268	0.443	
Gender of the occupants (1 if male, 0 otherwise) (standard error of parameter distribution)	0.944	0.229	Possible Injury
Time of the day (1 if between 3 pm to 7 pm, 0 otherwise) (standard error of parameter distribution)	0.203	0.402	
Traffic flow at the time of crash (ADT per lane in each direction – veh/day/ln) (1 if veh/day/ln >2000, 0 otherwise)	0.985	0.122	
Surface condition at the time of crash (1 if dry surface, 0 otherwise)	0.810	0.392	
Age group (1 if age between 25 to 35, 0 otherwise)	0.178	0.382	
Weather condition at the time of crash (1 if clear weather condition, 0 otherwise)	0.682	0.466	No-injury (PDO)
Light condition of street (1 if the surrounding area is dark but	0.119	0.324	

Meaning of Variables in the Model	Mean	Std. Dev.	Severity outcome
outside is lighted, 0 otherwise)			
Land-use pattern at crash location (1 if urban area, 0 otherwise)	0.564	0.496	
Age group (1 if age between 45 to 55, 0 otherwise) (standard error of parameter distribution)	0.239	0.427	
Surface condition at the time of crash (1 if dry surface, 0 otherwise)	0.810	0.392	
Traffic flow at the time of crash (ADT per lane in each direction – veh/day/ln) (standard error of parameter distribution)	15397.3	8845.23	

5.4 Methodological approach

Past research studies utilized a number of methodological approaches such as the multinomial logit, ordered probit and Bayesian Ordered Probit, Nested logit, and Mixed logit models (Chistoforou et al., 2010; Lemp et al., 2011; Milton et al., 2008; Zhu and Srinivasan, 2011; Duncan et al., 1998; Xie et al., 2009; Khorashadi et al., 2005; Chang and Mannering, 1999; Morgan and Mannerring, 2011) to model injury severities. In this study, we developed the mixed logit model for injury severity of crashes involving large trucks by considering random parameters in the developed model following the similar logical research framework of past research studies (Milton et al., 2008, Gkritza and Mannering, 2008, Chen and Chen, 2011). The level of injury is discrete in nature as coded in the injury scale – KABCO (i.e., 'K' for Fatal, 'A' for Incapacitating injury, 'B' for Non-incapacitating Injury, 'C' for Possible Injury and 'O' for Property-Damage-Only), and the mixed logit model has been widely accepted to model the effects of several contributory factors on the levels of injury severity. Several research studies conducted by Revelt and Train (1997, 1999), Train (1997), Brownstone and Train (1999), McFadden and Train (2000), and Bhat (2001) have clearly pointed out the effectiveness of this methodological approach. Although discrete outcome severity could be modeled by a multinomial logit model, heterogeneous effects and correlation in unobserved factors could still be a potential limitation in the assumption behind utilizing this model for injury severity (Train, 2009). So, a mixed logit model overcomes all of these limitations by generalizing the multinomial logit structure allowing for the

parameters βi vector to vary across observations of crashes (Savolainen et al., 2011). The assumption regarding Independent and Identically Distributed Errors (IID), Independence of Irrelevant Alternatives (IIA) and unobserved heterogeneity associated with observations in a multinomial logit model is completely relaxed by the introduction of the mixed logit approach (Jones and Hensher, 2007).

In order to achieve a better understanding of the injury severity of large-truck involved crashes in the Texas interstate system, an econometric model is utilized to determine the contributing factors that influence the likelihood of severity outcomes in large-truck involved crashes. In the framework of the model, a linear function that determines discrete injury severity outcome i (i.e., fatality, incapacitating injury, non-incapacitating injury, possible injury, and property damage only) for large-truck involved crashes n such that (Washington et al., 2011):

$$S_{in} = \boldsymbol{\beta}_i \mathbf{X}_{in} + \varepsilon_{in} \tag{5.1}$$

where $\mathbf{X}_{in}$ is vector of explanatory variables (e.g., driver, vehicle, road, and environment variables), $\boldsymbol{\beta}_i$ is vector of estimable parameters, ε_{in} is the error term. If ε_{in}'s are assumed to be generalized extreme value distributed; McFadden (1981) has shown that the multinomial logit results such that:

$$P_n(i) = \frac{EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]}{\sum_I EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]} \tag{5.2}$$

where $P_n(i)$ is probability of large-truck involved incident n having severity outcome i ($i \in I$ with I denoting all possible injury$\int$ severity outcomes as hereunto presented).

Since the Texas Crash Records Information System (CRIS) crash data are based on the law enforcement form CR-3, commonly known as Texas Peace Officer's Crash Report, there exists' the possibility of the unobserved heterogeneity. This may be related to police reporting, and unobserved factors related to roadway, vehicle, and driver not captured but may influence the crash outcome. Here we consider the possibility that elements of the parameter vector $\boldsymbol{\beta}_i$ may vary across observations of

each large-truck involved crash by using a random parameters logit model (also known as the mixed logit model) to account for the unobserved heterogeneity.

The chronological works by McFadden and Rudd (1994), Geweke et al. (1994), Revelt and Train (1997, 1999), Train (1997), Stern (1997), Brownstone and Train (1999), McFadden and Train (2000), and Bhat (2001) have shown the development and effectiveness of the mixed logit approach which can explicitly account for the variations (across crash observations) of the effects that variables have on the severity outcomes (or choices) considered in this study. With respect to this work, the mixed logit model is represented by (see Train, 2003),

$$P_{in} = \int \frac{EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]}{\sum_I EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]} f(\boldsymbol{\beta}_i|\boldsymbol{\varphi}) d\boldsymbol{\beta}_i \tag{5.3}$$

where $f(\boldsymbol{\beta}_i|\varphi)$is the density function of $\boldsymbol{\beta}_i$, $\boldsymbol{\varphi}$ is a vector of parameters of the density function (mean and variance), and all other terms are as previously defined. This model can now account for severity outcome specific variations of the effect of $\mathbf{X}_{in}$ on large-truck involved crash probabilities, with the density function $f(\boldsymbol{\beta}_i|\boldsymbol{\varphi})$ used to determine $\boldsymbol{\beta}_i$. The mixed logit probabilities are then a weighted average for different values of $\boldsymbol{\beta}_i$ across the observations where some elements of the vector $\boldsymbol{\beta}_i$ could be fixed and some randomly distributed. If the parameters are random, the mixed logit weights can be determined by the density function $f(\boldsymbol{\beta}_i|\boldsymbol{\varphi})$ (Milton et al., 2008; Washington et al., 2011). Maximum likelihood estimation of the mixed logit model shown in Equation (6.3) is performed through a simulation based approach due to the difficulty in computing the probabilities. The most widely accepted simulation approach utilizes Halton draws which is a technique developed by Halton (1960) to generate a systematic non-random sequence of numbers. Halton draws have been shown to provide a more efficient distribution of the draws for numerical integration than purely random draws (Bhat, 2003; Train, 1999).

To apply the mixed logit, the sample size of the data used has to be considered. In regard to

sample size, a study by Ye and Lord (2011) showed the influence of the sample size on three most commonly used injury severity models. Although the analysis by Ye and Lord (2011) on the sample size corresponding to severity models is simulation driven, the findings could be generalized for the mixed logit models. They found that crash severity models with sample sizes below 1,000 should not be estimated (Ye and Lord, 2011). An ordered probit model requires at least a sample size of more than 1000, and mixed logit is the most demanding requiring more than 5000 observations; whereas, a multinomial logit model stands in between the ordered probit and mixed logit models (i.e., 2000 and greater) (Ye and Lord, 2011). In this work, the sample size is 20,495 which is over the safe threshold (i.e., 5000) identified by Ye and Lord (2011).

In order to estimate the impact of a particular variable on the injury-outcome likelihood, elasticities (or direct-pseudo elasticity) are computed. Since most of the variables are indicator in nature in our study, direct-pseudo elasticities are estimated to measure the marginal effects when any particular indicator variable switches from 0 to 1 or reverse (Washington et al., 2011). Also, this is translated to percentage change in the likelihood of injury outcomes, while the indicator variables switching between 0 and 1 or 1 to 0. For binary indicator variables, the direct-pseudo elasticity is estimated as follows (Kim et al., 2010):

$$E_{x_{nk}}^{P_{in}} = \frac{P_{in}[given\ x_{nk} = 1] - P_{in}[given\ x_{nk} = 0]}{P_{in}[given\ x_{nk} = 0]} \qquad (5.4)$$

where, P_{in} is given the Equation (5.3) and simulated as shown in Equation (5.6).

x_{nk} = the k-th independent variable associated with injury severity i for observation n.

On the other hand, direct average elasticities of are estimated for any continuous variable using Equation (5.5). This measures the percentage change in the likelihood of injury outcome when the continuous variable changes one unit (Washington et al., 2011).

$$E_{x_{nki}}^{P_n(i)} = \frac{\frac{\partial P_n(i)}{P_n(i)}}{\frac{\partial x_{nk(i)}}{x_{nk(i)}}} = \frac{\partial P_n(i)}{P_n(i)} \cdot \frac{x_{nk(i)}}{\partial x_{nk(i)}} \tag{5.5}$$

where, $P_n(i)$ is given the Equation (5.3) and simulated as shown in Equation (5.6).

$x_{nk}(i)$ = the k-th independent variable associated with injury severity i for observation n.

An unbiased and smooth simulator (McFadden and Train, 2000), estimates unconditional probability in Equation (5.3) (Kim et al., 2010), which can be computed as (Walker and Ben-Akiva, 2002):

$$\hat{P}_{in} = \frac{1}{R}\sum_{r=1}^{R} P_{in} = \frac{1}{R}\sum_{r=1}^{R} \int \frac{EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]}{\sum_I EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]} f(\boldsymbol{\beta}_i | \boldsymbol{\varphi}) d\boldsymbol{\beta}_i \tag{5.6}$$

where, R = the total number of draws.

Since the direct pseudo-elasticity is calculated for each observation, it is usually reported as the average direct pseudo-elasticity (taking average over the sample) as a measure of the marginal effect of an indicator variable on the likelihood of a particular injury severity outcome (Kim et al., 2010).

With the simulator in Equation (5.6), Maximum Simulated Likelihood Estimation (MSLE) can be used to estimate the parameters and this MSLE estimator is asymptotically normal and consistent (Lee, 1992):

$$max_{\boldsymbol{\beta_{in}}} \sum_{n=1}^{N} \sum_{i=1}^{I} y_{in} ln \hat{P}_{in} \tag{5.7}$$

where, N = the total number of observations (i.e., crashes in the sample)

y_{in} = 1 if individual n suffers from injury severity i, 0 otherwise.

5.5 Empirical results with discussion

Maximum likelihood and simulation-based maximum likelihood methods are utilized to estimate parameter vector $\boldsymbol{\beta}_{\mathbf{i}}$ for fixed- and random-parameters logit models, respectively. We considered normal, lognormal, triangular, and uniform distributions for the distribution of the random parameters in our analysis. However, the normal distribution was found to be statistically significant. With regard to random parameters estimation, 200 Halton draws has been empirically shown to produce accurate parameter estimates which was used for the simulation-based maximum likelihood estimation (Milton et al., 2008; Bhat 2003).

The estimated variables in both models were found to be statistically significant within a 95% confidence level. A likelihood ratio test comparing the multinomial logit (i.e., fixed parameter model) and mixed logit (i.e., random parameter model) was performed to test the null hypothesis that the fixed parameter model is statistically equivalent to the random parameters model and the procedure is as follows (Washington et al., 2011):

$$\chi^2 = -2[LL_{MNL}(\beta^{MNL}) - LL_{ML}(\beta^{ML})] \tag{5.6}$$

where:

$LL_{MNL}(\beta^{MNL})$:is the log-likelihood at convergence of the fixed parameters model (-9231.729)

$LL_{ML}(\beta^{ML})$:is the log-likelihood at convergence of the random parameters model (-9200.257)

The Chi-square statistic for the likelihood ratio test with six degrees of freedom gave a value greater than the 99.99% ($\chi^2 = 62.944$) confidence interval, indicating that the mixed logit model (i.e., random parameter model) is statistically superior to the corresponding multinomial model (i.e., fixed

parameter model). This means that the null hypothesis that the random parameters estimated model is no better than the fixed model comparison model is rejected.

The discussion of the models results are described in the following subsections along with the elasticities (both in terms of direct pseudo and average direct elasticity) as presented in Table 5.2 and Table 5.3, respectively.

Table 5.2: Model Estimates for Severity Model

Meaning of Variable	Estimate	t-stat	P-value
Fatal outcome			
Constant (standard error of parameter distribution)	-6.911 (1.691)	-6.209 (2.458)	0.000
Terrain of roadway (1 if level roadway surface, 0 otherwise)	-1.032	-3.840	0.000
Time of day (1 if between 3 pm to 7 pm, 0 otherwise)	-0.803	-1.998	0.045
Land-use pattern at crash location (1 if rural area, 0 otherwise)	0.986	2.937	0.003
Traffic flow at the time of crash (ADT per lane in each direction – veh/day/ln)	-0.468x10^(-04)	-2.050	0.040
Weather condition (1 if clear weather, 0 otherwise)	0.901	2.499	0.013
Incapacitating Injury Outcome			
Constant	-3.567	-11.406	0.000
Month of the year (1 if summer months (June to August), 0 otherwise)	0.224	2.043	0.041
Weather condition (1 if clear weather, 0 otherwise)	-1.045	-5.436	0.000
Shoulder width (1 if right shoulder width (ft), 0 otherwise)	0.028	1.953	0.050
Time of the day (1 if between 3 pm to 7 pm, 0 otherwise)	-0.365	-2.627	0.008
Land-use pattern at crash location (1 if urban area, 0 otherwise)	-1.159	-6.924	0.000
Non-incapacitating Injury Outcome			
Constant (standard error of parameter distribution)	-14.403 (6.924)	-4.683 (4.494)	0.000
Number of lanes on highways (1 if 4 lanes in one direction, otherwise)	1.295	3.185	0.000
Time of the day (1 if between 12 am to 6 am, 0 otherwise)	1.978	3.804	0.000
Months of year (1 if fall months (September to December), 0 otherwise)	-0.680	-1.920	0.054
Land-use pattern at crash location (1 if rural area, 0 otherwise)	1.908	3.474	0.000
Possible Injury Outcome			
Constant (standard error of parameter distribution)	-0.939 (3.083)	-1.731 (3.289)	0.083
Gender of the occupants (1 if male, 0 otherwise) (standard error of parameter distribution)	-3.936 (1.622)	-4.417 (2.353)	0.000
Time of the day (1 if between 3 pm to 7 pm, 0 otherwise) (standard error of parameter distribution)	-0.542	-2.881	.0040
Traffic flow at the time of crash (ADT per lane in each direction – veh/day/ln) (1 if veh/day/ln >2000, 0 otherwise)	-0.889	-1.879	0.060
Surface condition at the time of crash (1 if dry surface, 0 otherwise)	-1.112	-5.010	0.000
Age group (1 if age between 25 to 35, 0 otherwise)	-0.388	-2.115	0.034
Non-Injury Outcome (Property-Damage-Only)			
Weather condition at the time of crash (1 if clear weather condition, 0 otherwise)	0.379	2.337	0.019
Light condition of street (1 if the surrounding area is dark but outside is lighted, 0 otherwise)	-0.564	-5.631	0.000
Land-use pattern at crash location (1 if urban area, 0 otherwise)	-0.377	-2.633	0.008

Meaning of Variable	Estimate	t-stat	P-value
Age group (1 if age between 45 to 55, 0 otherwise) (standard error of parameter distribution)	1.097 (1.866)	4.194 (6.800)	0.000
Surface condition at the time of crash (1 if dry surface, 0 otherwise)	-0.78347032	-6.264	0.000
Traffic flow at the time of crash (ADT per lane in each direction – veh/day/ln) (standard error of parameter distribution)	0.249x10^(-04) [0.294x10^(-04)]	2.347 (2.049)	0.018
Number of observations	20,495		
Restricted log-likelihood	-32985.43		
Log-likelihood at convergence	-9231.729		
Chi-squared value	47,570.35		
McFadden pseudo R-squared (ρ^2)	0.721		

Table 5.3: Average Direct Pseudo Elasticities

Variables	Elasticity (%)				
	PDO/No Injury	Possible Injury	Non-incapacitating Injury	Incapacitating Injury	Fatal
Terrain of roadway (1 if level roadway surface, 0 otherwise) [F]	0.21	0.15	0.11	0.50	-54.10
Time of day (1 if between 3 pm to 7 pm, 0 otherwise) [F]	0.03	0.02	0.01	0.06	-7.60
Weather condition (1 if clear weather, 0 otherwise) [F]	-0.26	-0.18	-0.13	-0.48	66.86
Land-use pattern at crash location (1 if rural area, 0 otherwise) [F]	-0.19	-0.14	-0.13	-0.52	49.65
Traffic flow at the time of crash (ADT in each direction – veh/day) [F]	0.17	0.13	0.07	0.36	-44.35
Month of the year (1 if summer months (June to August), 0 otherwise) [INCAP]	-0.14	-0.11	-0.07	6.00	-0.31
Weather condition (1 if clear weather, 0 otherwise) [INCAP]	1.03	0.71	0.43	-43.48	3.26
Shoulder width (1 if right shoulder width (ft), 0 otherwise) [INCAP]	-1.22	-0.93	-0.61	51.73	-2.66
Time of the day (1 if between 3 pm to 7 pm, 0 otherwise) [INCAP]	0.12	0.08	0.05	-5.10	0.15
Land-use pattern at crash location (1 if urban area, 0 otherwise) [INCAP]	0.89	0.73	0.31	-37.58	1.23
Number of lanes on highways (1 if 4 lanes in one direction, otherwise) [NONINCAP]	-0.70	-0.64	23.02	-1.35	-1.48
Time of the day (1 if between 12 am to 6 am, 0 otherwise) [NONINCAP]	-0.47	-0.49	15.61	-0.95	-1.06
Months of year (1 if from September to December, 0 otherwise) [NONINCAP]	0.13	0.12	-4.38	0.20	0.22
Land-use pattern at crash location (1 if rural area, 0 otherwise) [NONINCAP]	-0.76	-0.71	25.08	-1.64	-1.98
Weather condition (1 if clear weather, 0 otherwise) [NONINCAP]	0.61	0.49	-19.50	0.58	1.17
Gender of the occupants (1 if male, 0 otherwise) (standard error of parameter distribution) [POSS]	6.58	-113.33	3.35	10.39	9.18

Variables	Elasticity (%)				
	PDO/No Injury	Possible Injury	Non-incapacitating Injury	Incapacitating Injury	Fatal
Time of the day (1 if between 3 pm to 7 pm, 0 otherwise) (standard error of parameter distribution) [POSS]	0.26	-4.44	0.12	0.29	0.17
Traffic flow at the time of crash (ADT per lane in each direction – veh/day/ln) (1 if veh/day/ln >2000, 0 otherwise) [POSS]	2.43	-41.68	1.30	3.48	3.03
Surface condition at the time of crash (1 if dry surface, 0 otherwise) [POSS]	2.39	-41.06	1.24	3.38	3.52
Age group (1 if age between 25 to 35, 0 otherwise) [POSS]	0.17	-2.95	0.09	0.19	0.18
Weather condition at the time of crash (1 if clear weather condition, 0 otherwise) [PDO]	1.38	-11.02	-6.53	-14.71	-25.39
Light condition of street (1 if the surrounding area is dark but outside is lighted, 0 otherwise) [PDO]	-0.63	3.48	3.02	10.37	11.48
Land-use pattern at crash location (1 if urban area, 0 otherwise) [PDO]	-1.10	9.93	4.50	11.45	10.35
Age group (1 if age between 45 to 55, 0 otherwise) (standard error of parameter distribution)	-0.45	-0.41	-2.87	19.97	15.21
Surface condition at the time of crash (1 if dry surface, 0 otherwise) [PDO]	-3.83	27.29	16.95	52.28	56.66
Traffic flow at the time of crash (ADT per lane in each direction – veh/day/ln) (standard error of parameter distribution) [PDO]	1.47	-13.11	-7.90	-13.41	-12.37

Here, F= Fatal, INCAP = Incapacitating, NONINCAP = Non-incapacitating, POSS = Possible and PDO = Property-Damage-Only/ No-injury

5.5.1 Model constants

The constant term for fatal outcome is found to be random and normally distributed with mean of -6.911 and standard deviation of 1.691. Given these values, this constant is less than zero for 99.9% for large-truck involved crashes which result in fatalities. That is, about 99% of large-truck involved crashes on average were less likely to result in fatal outcomes. Similarly, the constant for non-incapacitating injury outcome is found to be random and normally distributed with mean of -14.404 and standard deviation of 6.924. This leads to the constant being less than zero for 98.1% for large-truck involved crashes for this injury category and implying that about 98.1% of large-truck involved crashes on

average were less likely to result in non-incapacitating outcomes. Finally, the constant for possible injury outcome was also found to be random and normally distributed with mean of -0.938 and standard deviation of 3.083. Given these estimates, this constant is less than zero for 38% of large-truck involved crashes and implying that about 38% of large-truck involved crashes on average were more likely to result in possible injuries.

5.5.2 Drivers' characteristics

Turning to the results in Table 5.2, the male indicator is negative indicating that males are less likely to be involved in possible injuries as opposed to female occupants. Also, this parameter was found to be random and normally distributed with a mean of -3.936 and a standard deviation of 1.622 for possible injuries. Given these estimates, this parameter is above zero for 0.7% of large-truck involved crashes. This suggests that about 0.7% of large-truck involved crashes involving male occupants (drivers or passengers or both) on average were more likely to result in possible injuries. Usually, injury tolerance of the male body is higher than that of female body. Chen and Chen (2011) found a similar result as females are more likely to be involved in fatal or incapacitating injuries in single and multi-vehicle collisions. Additionally, this evidence is also supported by a study Abdel-Aty and Abdelwahab (2001).

The age group of 45 to 55 years old is more likely to be involved in non-injury crashes. The elasticity estimate indicates that this age group is 15.21% and 19.97% more likely to be involved in fatalities and incapacitating injuries, respectively. A possible explanation is that this age group may be more experienced when it comes to driving and handling large trucks, but may have slower reaction times under avoidance maneuvering. This parameter was also found to be random for the non-injury outcome and normally distributed with a mean of 1.097 and a standard deviation of 1.866. Given these values, this parameter is below zero for 72.2% of large-truck involved crashes. This implies that about 72.2% of large-truck involved crashes with this age group (45 to 55 years' old drivers) on average were

less likely to result in non-injury crashes. A study by Driscoll et al. (2005) found that about 70% of the traffic deaths were people between 25 to 54 years of age. The authors also point out that the largest proportion of these deaths occurred in Australia, New Zealand, and the US. The model result also indicate that the age group of 25 to 35 years of age is 2.9% (as indicated by the elasticity) less likely to be involved in possible injury crashes. This may be due to possibly faster reaction times by this age group under crash avoidance maneuvering compared to the older age group. In their study, Stamatiadis and Deacon (1995) indicated that motorist between 25 to 35 years of age were on average more likely to be involved in the crashes compared to motorist who were between 45 to 55 years old. Also, younger drivers performed relatively poorly because of limited driving experience, risk-taking behaviors, and attitudinal factors particularly at night which lead to a significant finding that middle age drivers are safer than younger group (Stamatiadis and Deacon, 1995).

5.5.3 Land use characteristics

Crashes occurring in the vicinity of a rural area (with a population of less than 5,000) resulted in 49.6% likelihood of fatalities and 25.2% likelihood of non-incapacitating injuries. This may be a result of rural areas having greater speed limits, longer stretches of roads with less traffic, and less enforcement compared to urban settings. A more detailed study by Khorashadi et al. (2005) found similar findings for rural settings where the probability of drivers' injuries in crashes involving excessive speeds, improper lane passing and single vehicle collisions increased likelihood of severe injuries or fatalities. Another study by Khattak et al. (2002) also found similar findings and that injury to older drivers increase to more severe if the crash occurred in rural settings compared to urban settings.

In contrast to rural crashes, crashes occurring in an urban area having a population of over 200,000 reduced the likelihood of incapacitating injuries by 37.6%. This may be a result of drivers driving slower in urban settings as a result of lower speed limits. Additionally, this variable may be reflecting the existence of higher enforcement levels. A similar finding by Khorashadi et al. (2005)

indicated that in an urban setting there in a greater likelihood of non-injuries crashes because of improper lane passing and multi-vehicle collisions.

5.5.4 Temporal characteristics

When crashes occurred between the time period of 3 pm to 7 pm the likelihood of fatalities, incapacitating injuries and possible injuries is reduced by 7.6%, 5.1% and 4.4%, respectively. This time period captures the afternoon peak and it may be reflecting the effect of congestion during this time period. That is, speeds tend to be slower during congested periods. A similar result is also reflected in a study by Doherty et al. (1998) with a fact that crash rates for all severity levels increases towards the evening (8:00 pm to 11:59) and late night to early morning (12:00 am to 4:59 am) time periods. In contrast, the time period from 12 am (midnight) to 6 am (morning) increases the likelihood of non-incapacitating injuries by 15.6%. This may be a result of late night driving behavior resulting from drowsy driving or being fatigued. A similar result is also found in Doherty et al. (1998).

Fall (September to December) months decreased the likelihood of non-incapacitating injuries by 4.4%. This may be due to more cautious driving due to climate and traffic pattern changes. For example, traffic volumes tend to reduce during the fall season as compared to summer season where interactions between vehicles on roads are usually relatively high. However, for the summer months between June to August increased the likelihood of incapacitating injuries by 6%. This could be reflecting the increased traffic interactions in the highways between passenger vehicles and other large trucks. Similar results were found by Brown and Baass (1997) during for these time periods. Ulfarsson and Mannering (2004) also found that summer months increased incapacitating injuries for female drivers in Sport Utility Vehicle/minivan for single vehicle collisions.

5.5.5 Traffic characteristics

Higher traffic flow in terms of average daily traffic per lane in each direction reduced the likelihood of fatalities by 44.3%. This could be a result of higher traffic volumes and the presence of congested periods. Further, traffic flow on average higher than 2000 vehicles per lane in each direction also reduced the likelihood of possible injuries by 41.7%. Again this may be a result of higher traffic volume and the presence of congested traffic conditions. A similar study by Chen and Chen (2011) showed that low AADT (traffic volume < 2000 AADT per lane) increases possible or non-incapacitating injuries for single and multi-vehicle collisions.

5.5.6 Weather characteristics

Turning to weather conditions, clear weather increased the likelihood of fatal outcomes by 66.9%. A possible explanation may stem from drivers being more relaxed and possibly invoking risk-taking behaviors due to better visible driving conditions. A similar result is supported by Edwards (1998). On the other hand, clear weather condition reduced the likelihood of incapacitating injuries by 43.5%.

5.5.7 Road geometry characteristics

In regards to road geometry, the indicator for four lanes in each direction was found to increase the likelihood of non-incapacitating injuries by 23%. This variable maybe capturing the effects of the rightmost lane of a highway; that is, the rightmost lane typically is a slower lane with a greater number of truck traffic. Furthermore, the existence of a wider right shoulder indicator increased the likelihood of incapacitating injuries by 51.7%. This may be reflecting risk compensating behavior due to drivers feeling comfortable in having increased space to drive on. The presence of a level surface increased the likelihood of non-fatalities by 54.1%. A possible explanation may be that level surfaces may instill increased driver awareness due to favorable driving visibility.

Crashes occurring on dry pavement reduce the likelihood of possible injuries by 41.1%. On dry surface any evasive actions are very effective as opposed to wet surface in terms of effective skidding of tires on dry surface.

5.5.8 Lighting characteristics

Crashes occurring in the dark highway section (but the in lighted condition) increased the likelihood of fatalities by 11.5%, and likelihood of incapacitating injuries by 10.4%. Similar results were found by Morgan and Mannering (2011) for single vehicle crashes for female less than 45 years on dry surfaces and male less than 45 years on wet surfaces, Malyshkina and Mannering (2009) for single vehicle collisions on Indiana interstates, and Anastasopoulos and Mannering (2011) for rural interstates in Indiana.

In summary, the results provided insights into the complex interactions of various human, vehicle, and road-environment factors. In addition, the results indicate that some of the variables in the model varied across observations validating the utilization of the mixed logit model.

5.6 Model validation

We conducted out-of-sample sampling test to validate the model results for multinomial logit model. Although mixed logit model was estimated for the analysis, this model validation process demonstrates step-wise procedures as a rule-of-thumb for base model (i.e., multinomial logit model). Figure 5.1 shows the step-y-step procedures and Table 5.4 indicates the results of the out-of-sample test performed on CRIS data over a period of five years (2006 to 2010).

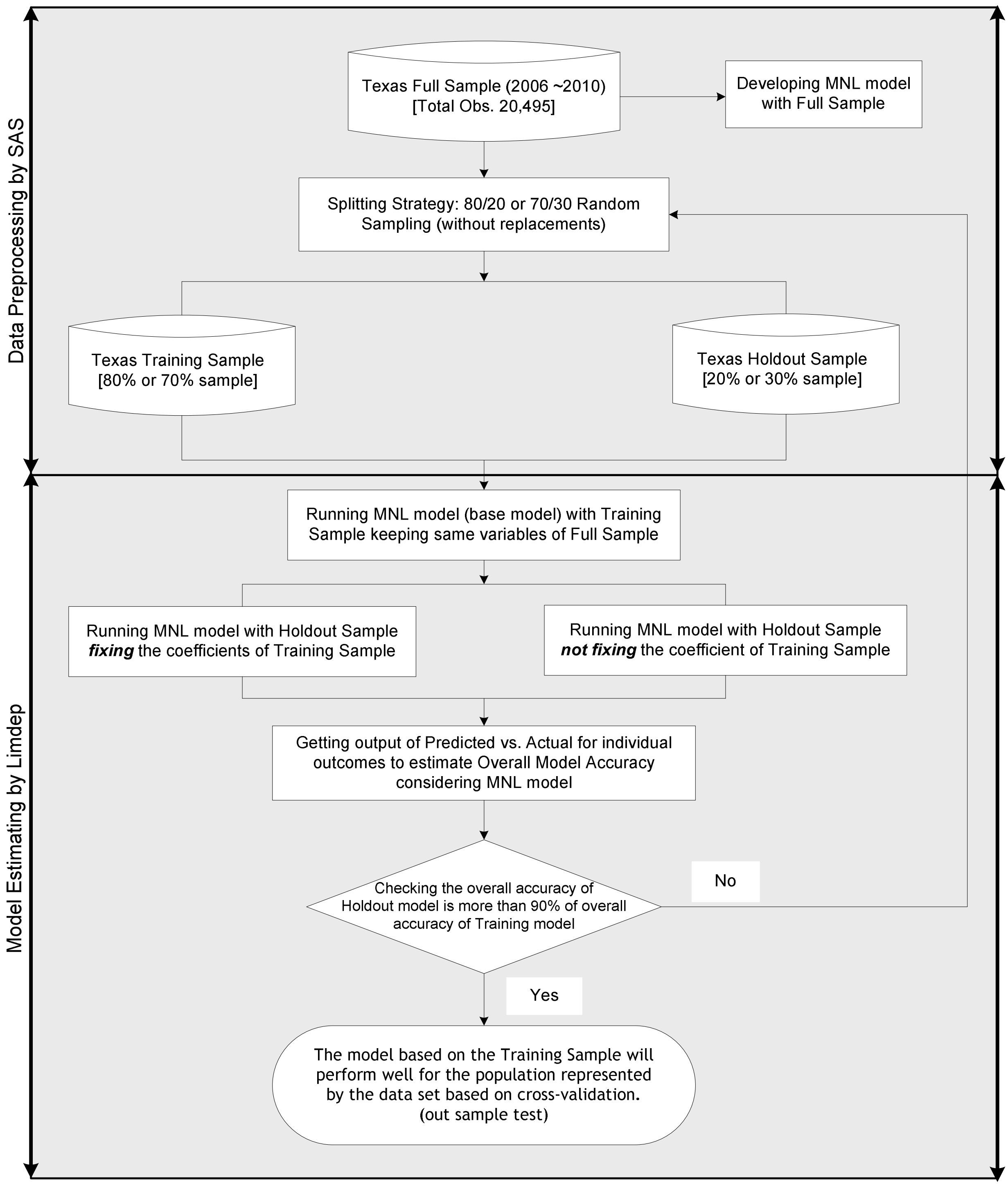

Figure 5.1: Out-of-Sample Model Validation Process

Figure 5.2 shows flow chart for out-of-sample validation process with the multinomial logit model. Here, we randomly sampled the whole Texas crash dataset into two parts (70-30 and 80-20 split) as

training (70 or 80% of sample) and holdout (30% or 20% of sample) sample utilizing SAS seed numbers. Then, we modeled multinomial logit model (i.e., base model) with training sample (70% or 80% of the sample size) using the same variables of multinomial logit models using the full sample in NLOGIT. Having run the multinomial logit with training sample and kept the coefficients of that model, we ran multinomial logit model with holdout sample keeping same variables and the coefficients of training sample. We estimated the predicted and actual probabilities for all the observations for holdout sample and compared the same with training sample. We found that overall model accuracy for holdout sample is more than 90% of model accuracy of training sample for multinomial logit model. Both random splits without replacements in the sampling process such as 70/30 and 80/20 satisfied that the model accuracy of holdout sample is more than 90% of model accuracy of the training sample. The final model validation process is presented in Table 5.4. Although this criterion (i.e., overall model accuracy of holdout sample should be more than 90% of model accuracy of training sample) is more or less is used in this study as rules of thumb for multinomial logit model (Hair et al, 2010). We also ran multinomial model with holdout sample without fixing the coefficients of the training sample. In both cases, this criterion is satisfied as a proof of out-of-sample test. This out-of-sample cross validation test indicates the model based on training sample will perform well for population represented by the dataset.

Table 5.4: Model Cross-Validation (Out of Sample Test) Summary

	70/30 Split		80/20 Split	
	Training Sample	Holdout Sample	Training Sample	Holdout Sample
Base Model – MNL Model	Yes	Yes	Yes	Yes
Same variables as in the full sample	Yes	Yes	Yes	Yes
Overall Model Accuracy	79.9%	80.5%	80.1%	80.2%
90% of Training Sample Model Accuracy	71.9%	-	72.1%	-
Criteria Satisfied?	Yes (80.5% > 71.9%)		Yes (80.2% > 72.1%)	

5.7 Conclusion and future research

This study develops and demonstrated the use of a mixed logit model for studying injury severity due to large-truck involved crashes on the Texas interstate system, and which utilized five distinct injury severity outcomes –fatal, incapacitating, non-incapacitating, possible, and no injury (property damage only), a first. The mixed logit model was developed using crash data over a period of five years from 2006 to 2010. Furthermore, the mixed logit modeling framework is an important approach because it allows us to account and to correct for heterogeneity that can arise from factors such as individuals (i.e., drivers and passengers), vehicle, road-environment, weather, variations in police reporting, temporal and other unobserved factors not captured in the dataset. The data used in this study was the CRIS database for the years of 2006 to 2010, and to the best of our knowledge a first with respect to explicitly modeling large-truck injury severity.

From the model results, temporal characteristics such as time of day – 3 pm to 7 pm, 12 am to 6 am, months of year – summer and fall (i.e., June to August and September to December) and spatial characteristics such as rural and urban settings as well as the traffic conditions expressed as directional AADT per lane significantly affect the injury severity outcomes at different injury levels. Additionally, geometry of highways such as terrain of roadway, shoulder width, and surface condition affect injury severity outcomes. In light of human factors, demographic characteristics such as gender and age of occupants (e.g., 23-35 and 45-55 years age group) also contribute to severity outcomes.

Variables, representing the contributing factors in mixed logit model, influence likelihood of each injury outcome with increasing or decreasing effects. It is evident that higher traffic flow (i.e., vehicles per day per lane) reduces the likelihood of fatalities but increases that of PDO crashes. However, directional traffic flow in a lane in terms of vehicles per day exceeding a threshold of 2000 vehicles per day per lane decreases likelihood of possible injuries. Turning to locational characteristics, crashes occurring in rural settings result in a higher likelihood of fatalities and non-incapacitating

injuries. In contrast, crashes occurring in urban settings result in a lower likelihood of incapacitating and PDO crashes. Then, both of the spatial characteristics are found to be fixed across the observations. Time of day such as 3 pm to 7 pm results in lower likelihood of fatalities, incapacitating and possible injuries. Additionally, time of day such as 12 am to 6 am results in higher likelihood of non-incapacitating injuries. Summer is likely to increase likelihood of incapacitating injuries, whereas fall is likely to decrease likelihood of non-incapacitating injuries. Both time periods – summer and fall are found to be fixed across the observations. Increase in right side shoulder width increases the likelihood of incapacitating injuries. Four lanes in each direction also increase the likelihood of non-incapacitating injuries. The age group of 25 to 35 years is less likely to be involved in possible injuries. Likewise, age group of 45 to 55 years is more likely to be involved in no-injury crashes where this age group is found to be random for possible injury outcome. Occupants being male are more likely to be involved in possible injuries and it is found to be a random parameter which varies across the observations. Dry surface condition reduces the likelihood of possible and PDO crashes and found to be fixed across the observations.

Although this study is exploratory in nature, the modeling approach presented in this paper offers a methodology to analyze large-truck injury severity and at same time account for unobserved factors. From a future work perspective, exploring the statistical differences in regards to rural versus urban settings in Texas.

Chapter 6: Modeling Injury Outcomes of Heavy Vehicle involved Crashes in Rural and Urban Areas in Texas

6.1 Introduction

NHTSA data indicates that rural and urban settings share disproportionate numbers of fatal crashes and fatalities because of the disproportionate number of people living in those areas. According to the 2009 census, rural fatalities accounted for 57% of all traffic fatalities in 2009, despite the fact that only 23% of the US population lived in rural areas. In addition, crash data over a period of 2000 to 2009 indicated that rural fatalities decreased by 22% as opposed to urban fatalities which decreased by only 11%. Moreover, in 2009 NHTSA safety facts on Rural/Urban Comparison show that rural settings accounted for 56% of the fatal crashes, resulting in 57% of the fatalities, as opposed to urban settings which accounted for 43% of the fatal crashes resulting in 42% of the fatalities. Past studies and crash statistics show that fatality rates in rural areas are much higher than in urban areas (Eberhardt et al., 2001). Significant factors behind the higher rates of fatalities directly point to the challenges of: (i) longer notification time, the time it takes emergency medical services to the crash scenes, and (ii) the time spent in delivering the severely injured occupants to the nearby medical facilities with expected quality in rural settings (Clark and Cushing, 1999, 2002, 2004; Svenson et al., 1996; Chen et al., 1995). Also, rural areas experience longer response times and timely medical due to the fact long driving distances compared to urban areas (Svenson et al., 1996; Zlatoper, 1989; Muelleman and Mueller, 1996). All these factors associated with higher fatality rates indicate the need for research that addresses the problem of large-truck involved crashes in rural and urban areas in Texas.

Few studies have investigated the crash characteristics leading to rural crashes with high rates of fatalities. A study by Rakauskas et al. (2009) identified recurrent characteristics of fatal rural crashes to be: (1) more than one fatality per crash; (2) male driver; (3) younger driver; (4) alcohol consumption; (5)

truck involvement; (6) higher speed; (7) vehicle rollover; (8) head-on collision; and (9) ejected person due to seatbelt non-compliance (Blatt and Furman, 1998; NHTSA, 2001, 1996). Additionally, other studies (Lee and Mannering, 2002; Khorashadi et al., 2005; Chen and Chen, 2011) found that there are significant differences in driving characteristics resulting from driving behavior, characteristics of driver populations, and driver behavior with "visual noise" (Khorashadi et al., 2005). More specifically, that differences in the driving patterns in rural and urban areas lies in more vehicle-miles travelled (Weiss et al. 2001; Baker et al., 1987); less safety caution in restraint usage (i.e., seat belts and child restraint) and more alcohol consumption (Weiss et al. 2001; King et al., 1994; Baker et al., 1987; Muelleman and Mueller, 1996); less safe vehicles traveling (Baker et al., 1987; Muelleman and Mueller, 1996); higher speed (Weiss et al. 2001; Muelleman and Mueller, 1996; Muelleman et al. 1993; NHTSA, 2001; Araki and Murata, 1986; NSC, 1985); delayed and poor quality medical response in rural settings than in urban settings (Wylie and Kimball, 1997; Muelleman and Mueller, 1996). Zwerling et al. (2005) also found similar patterns of risks associated with rural and urban settings by analyzing national crash database (such as NASS-GES and FARS). They found that the fatal crash incidence density was more than two times higher in rural than in urban areas with the injury fatality rate more than three times higher in rural areas.

Therefore, this chapter seeks to investigate the contributing factors and the characteristics relating to large-truck involved crashes in rural and urban settings in the state of Texas from 2006 to 2010, utilizing the CRIS database. This study is differentiates itself from other studies, for example, Khorashadi et al. (2005) focusing on large trucks crashes in California, because of the uniqueness of crash characteristics of drivers, vehicular standards, and road-environment in rural and urban areas in Texas. With that in mind, this study is designed to be exploratory in nature, identifying the contributing factors related to injury outcomes in large-truck involved crashes by developing separate mixed logit models for rural and urban regions of Texas rather than a combined model for both. These developed

rural and urban models can be utilized to develop tools for countermeasures to aid transportation safety professionals in mitigating severe injury crashes. Although the mixed logit model has been applied to large-truck crash severity from different modeling perspectives, this research extends the current literature and introduces additional significant variables in regards to large-truck crashes for rural and urban areas, also a first.

6.2 Background of rural and urban crash models

Studies focused specifically on crashes in rural and urban areas is sparse; however the following background presents those works closely related to this topic.

Massie et al. (1995) analyzed crash involvement rates utilizing the National Personal Transportation Survey combined with the Fatality Analysis Reporting System and the General Estimates System. Massie et al. (1995) found that crash distribution based on rural and urban settings had no significant difference between male and female drivers. Still, other studies have found some difference in crash involvement based on rural and urban settings.

Abdel-Aty and Radwan (2000) developed negative binomial models for principal arterial in central Florida and found that urban areas have higher crash frequency than rural areas. This increase in crashes was explained by higher access points (i.e., access density) and higher level of congestions. This study also found that both male and female drivers are also highly involved in the crashes in urban areas. Additionally, this study indicated that young, middle, and old drivers have higher crash involvement in urban areas.

A study by Rakauskas et al. (2009), conducted on ANCOVA analysis for rural and urban self reported risks associated with driving behaviors in fatal crashes and attitudes towards safety interventions based upon a large scale survey conducted in Minnesota. The findings of this study indicated that rural drivers are more likely to take risks due to: lower perception of risk seat-belt usage,

and driving under the influence of alcohol, the types of vehicles driven and also government sponsored traffic safety interventions.

Khorashadi et al. (2005) developed a multinomial logit model based upon California crash data and focusing upon large-truck involved crashes, to identify the risk factors contributing to injury severity outcomes. This study emphasized the difference in risk factors for rural and urban areas where 13 variables are significant in rural areas (which are absent in urban model) and 17 variables are significant in urban areas (which are absent in rural model). These differences in factors resulting in different injury outcomes are attributed to perceptual, cognitive, and response demand for the drivers in rural and urban settings.

6.3 Statistical approach

Many research studies utilized a number of methodological approaches such as the multinomial logit, ordered probit, Bayesian Ordered Probit, nested logit, and mixed logit models. Each of these models have been utilized to model injury severities (Chistoforou et al., 2010; Lemp et al., 2011; Milton et al., 2008; Zhu and Srinivasan, 2011; Duncan et al., 1998; Xie et al., 2009; Khorashadi et al., 2005; Chang and Mannering, 1999). In this study, we estimate a mixed logit model for injury severity of crashes involving large trucks by considering random parameters in the developed model following the similar logical research framework as past research studies (Milton et al., 2008; Gkritza and Mannering, 2008; Chen and Chen, 2011). The level of injury is discrete in nature as coded in the injury scale KABCO (i.e., ‘K’ for Fatal, ‘A’ for Incapacitating injury, ‘B’ for Non-incapacitating Injury, ‘C’ for Possible Injury and ‘O’ for Property-Damage-Only), and a mixed logit model has been widely accepted to model the effects of several contributing factors on the levels of injury severity. Several research studies conducted by Revelt and Train (1997, 1999), Train (1997), Brownstone and Train (1999), McFadden and Train (2000), and Bhat (2001) have clearly pointed to the effectiveness of this methodological approach. Although discrete outcome severity could be modeled by a multinomial logit

model, heterogeneous effects and correlation in unobserved factors could still have potential limitations in the assumption behind utilizing this model for injury severity (Train, 2009). Thus, a mixed logit model overcomes all of these limitations by generalizing multinomial logit structure, allowing for the parameters $\boldsymbol{\beta}_i$ vector to vary across the observation of crashes (Savolainen et al., 2011). The assumption regarding IID (independently and identically distributed errors), IIA (independence of irrelevant alternatives) and unobserved heterogeneity associated with observations of the multinomial logit model are completely relaxed by introducing a mixed logit approach (Jones and Hensher, 2007).

In order to achieve a better understanding of the injury severity of large-truck involved crashes for rural and urban settings, we seek to develop a statistical model that can be used to determine the contributing factors that influence the likelihood of severity outcomes in large-truck involved crashes. To do so, we start with a linear function S_{in} that determines discrete injury severity outcome i (fatality, incapacitating injury, non-incapacitating injury, possible injury, and property damage only) for large-truck involved incident n such that (Washington et al., 2011):

$$S_{in} = \boldsymbol{\beta}_i \mathbf{X}_{in} + \varepsilon_{in} \tag{6.1}$$

where $\mathbf{X}_{in}$ is vector of explanatory variables (driver, vehicle, road, and environment variables), $\boldsymbol{\beta}_i$ is vector of estimable parameters, ε_{in} is the error term. If ε_{in}'s are assumed to be generalized extreme value distributed, McFadden (1981) has shown that the multinomial logit results such that:

$$P_n(i) = \frac{EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]}{\sum_I EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]} \tag{6.2}$$

where $P_n(i)$ is probability of large-truck involved incident n having severity outcome i ($i \in I$ with I denoting all possible injury severity outcomes)

As CRIS crash data is likely to have a significant amount of unobserved heterogeneity because the information regarding any of the factors are not obtained from the in-depth crash investigation or

reconstruction studies (for example, relating to police reporting, roadway, vehicle, and driver factors), we consider the possibility that elements of the parameter vector $\boldsymbol{\beta}_i$ may vary across observations of each large-truck involved crash by using a random-parameters logit model (also known as the mixed logit model). Previous works by McFadden and Rudd (1994), Geweke et al. (1994), Revelt and Train (1997, 1999), Train (1997), Stern (1997), Brownstone and Train (1999), McFadden and Train (2000), and Bhat (2001) have shown that the development and effectiveness of the mixed logit approach can explicitly account for the variations (across crash observations) of the effects that variables have on the severity outcomes (or choices) considered in this study. The mixed logit model is written as (see Train, 2003):

$$P_{in} = \int \frac{EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]}{\sum_I EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]} f(\boldsymbol{\beta}_i|\boldsymbol{\varphi}) d\boldsymbol{\beta}_i \tag{6.3}$$

where, $f(\boldsymbol{\beta}_i|\varphi)$is the density function of $\boldsymbol{\beta}_i$, $\boldsymbol{\varphi}$ is a vector of parameters of the density function (mean and variance), and all other terms are as previously defined. This model can now account for severity outcome specific variations of the effect of $\mathbf{X}$ on large-truck involved crash probabilities, with the density function $f(\boldsymbol{\beta}_i|\boldsymbol{\varphi})$ used to determine $\boldsymbol{\beta}_i$. Mixed logit probabilities are then a weighted average for different values of $\boldsymbol{\beta}_i$ across carriers where some elements of the vector $\boldsymbol{\beta}_i$ may be fixed and some randomly distributed. If the parameters are random, the mixed logit weights are determined by the density function $f(\boldsymbol{\beta}_i|\boldsymbol{\varphi})$ (Milton et al., 2008; Washington et al., 2011).

Maximum likelihood estimation of the mixed logit model shown in Equation (6.3) is undertaken with simulation approaches due to the difficulty in computing probability. The most widely accepted simulation approach is a technique developed by Halton (1960) to generate a systematic non-random sequence of numbers. Halton's technique (known as Halton draws) has been shown to provide a more efficient distribution of the draws for numerical integration than purely random draws (Bhat, 2003; Train, 1999).

A study by Ye and Lord (2011) showed the influence of the sample size on injury severity modeling. Although the analysis by Ye and Lord (2011) regarding the sample size corresponding to each severity model is simulation driven, there are still a few findings that could be generalized in terms of sample size for the three commonly used models. An ordered probit model is the one that requires the least samples (more than 1000), a mixed logit is the most demanding on samples (more than 5000), and a multinomial logit model requirement are located between the ordered probit and mixed logit models (somewhere more than 2000). In our study, the rural crash sample size is 5,484 and the urban crash sample size is 11,560, which are above and over the safe threshold regarding sample size identified by Ye and Lord (2011).

In order to estimate the impact of particular variables on the injury-outcome likelihood, elasticities (or direct-pseudo elasticity) are computed. In the context of the current injury severity model, most of the variables are indicator in nature; direct-pseudo elasticities are estimated to measure the marginal effects of indicator variables when any particular indicator variable switches from 0 to 1 or reverse (Washington et al., 2011). Also, this is translated to percentage change in the likelihood of the indicator variables switching between 0 and 1 or 1 to 0. For binary indicator variables, the direct-pseudo elasticity is estimated as follows (Kim et al., 2010):

$$E_{x_{nk}}^{P_{in}} = \frac{P_{in}[given\ x_{nk} = 1] - P_{in}[given\ x_{nk} = 0]}{P_{in}[given\ x_{nk} = 0]} \tag{6.4}$$

where, P_{in} is given the Equation (6.3) and simulated as shown in Equation (6.5).

x_{nk} = the k-th independent variable associated with injury severity *i* for observation n.

The unconditional probability in Equation (6.3) (Kim et al., 2010) can be estimated with an unbiased and smooth simulator (McFadden and Train, 2000) that is computed as (Walker and Ben-Akiva, 2002):

$$\hat{P}_{in} = \frac{1}{R}\sum_{r=1}^{R} P_{in} = \frac{1}{R}\sum_{r=1}^{R}\int \frac{EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]}{\sum_I EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]} f(\boldsymbol{\beta}_i|\boldsymbol{\varphi})d\boldsymbol{\beta}_i \qquad (6.5)$$

where, R = the total number of draws.

Since the direct pseudo-elasticity is calculated for each observation, it is usually reported as the average direct pseudo-elasticity (taking average over the sample) as a measure of the marginal effect of an indicator variable on the likelihood of a particular injury severity outcome (Kim et al., 2010).

With the simulator in Equation (6.5), Maximum Simulated Likelihood Estimation (MSLE) can be used to estimate parameters and this MSLE estimator is asymptotically normal and consistent (Lee, 1992):

$$max_{\boldsymbol{\beta}_{in}} \sum_{n=1}^{N}\sum_{i=1}^{I} y_{in} ln\hat{P}_{in} \qquad (6.6)$$

where, N = the total number of observations (i.e., crashes in the sample)

y_{in} = 1 if individual n suffers from injury severity i, 0 otherwise.

6.4 Empirical settings

The data used in this study was collected from the Texas Peace Officers' Crash Reports commonly known as Crash Records Information System (CRIS) database. To investigate human, vehicle, and road-environmental factors, a sample of 20,495 data observations were extracted from CRIS database by further focusing on crashes involving large trucks and the highway category over a time period of five years (from 2006 to 2010). Each observation in the sample is a crash, representing the maximum level of injury to the drivers, involving at least one large truck in the interstate system. The crash dataset was linked to the vehicle and person dataset through appropriate linking variables, namely crash number, and the vehicle and person dataset were linked through the vehicle and crash number.

As mentioned earlier, at the initial stage of processing, the Texas crash dataset of 20,495 observations was filtered for large-truck involved crashes on the interstate system. On the second stage, this dataset was further divided into two subsamples – i) urban area having a population of more than 200,000 with a dataset of 11,560 observations (56.4% of processed Texas crash sample), and ii) rural area having population of less than 5,000 with a dataset of 5,484 observations (26.7% of processed Texas crash samples). Crash observations corresponding to locations with populations between 5,000 and 200,000 were not considered in this study. The intent of this analysis is to identify the factors related to rural and urban settings while not considering any factors and error associated with model estimation for the intermediate regions with populations between 5,000 to 200,000 people. The analysis area is on the two opposite scale of the number of residents (less than 5,000 residents in rural areas and more than 200,000 residents in urban areas) in the state of Texas considering the definition of rural and urban areas in the existing literature of CRIS data system.

Although it is likely to have issues of under reporting for minor and no personal injuries in police-reported crash database, the CRIS is a fairly representative sample of the crashes from the police reports for the state of Texas. As such, it is fairly common practice in the modeling approach to assume that sample data selected from the population has equal likelihood of being considered in the sample (Savolainen et al., 2011).

Figure 6.1 shows the distribution of injury severities of these two geographical settings: rural and urban areas in Texas.

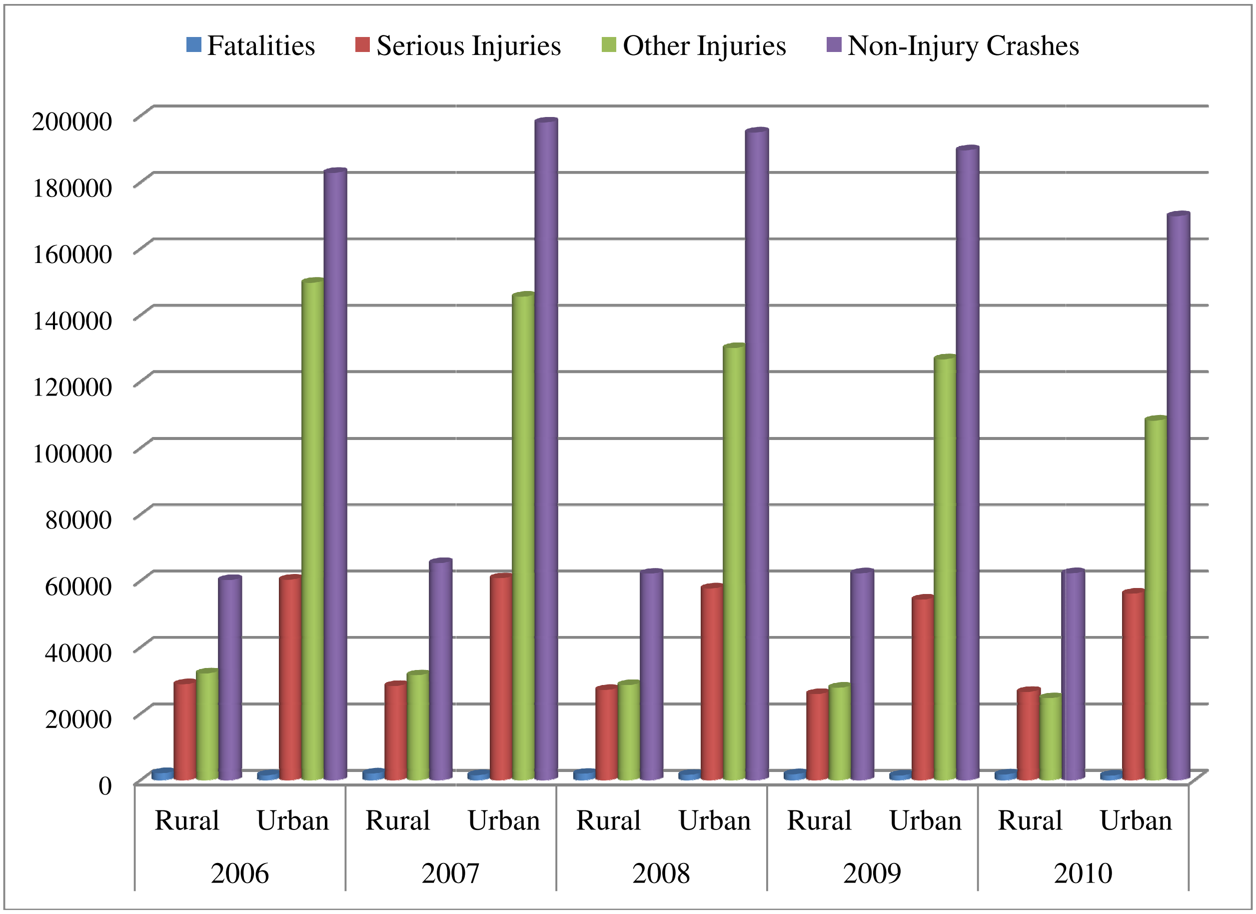

Figure 6.1: Number of Causalities in Road Crashes in Rural and Urban Areas Texas (TxDOT, 2006~10)

Descriptive statistics of the variables used in the models are presented in Table 6.1 and Table 6.2 for rural and urban areas, respectively. Table 6.1 and 6.2 show the mean and standard deviation of the key variables considered in the models.

Table 6.1: Descriptive Statistics of the Key Variables in Rural Severity Model (N = 5,484)

Meaning of Variables in the Model	Mean	Std. Dev.	Severity outcome
Road alignment (1 if level surface, 0 otherwise)	0.740	0.438	Fatal
Crash with fixed objects (1 if roadside fixed object, 0 otherwise)	0.163	0.369	
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	0.429	0.495	
Weather condition at time of crash location (1 if clear sky, 0 otherwise)	0.652	0.476	Incapacitating Injury
Weather condition at time of crash location (1 if clear sky, 0 otherwise)	0.652	0.476	
Months of the year (1 if between June to August, 0 otherwise)	0.251	0.433	
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	0.429	0.495	
Median type (1 if median is unprotected, 0 otherwise)	0.710	0.453	Non-Incapacitating Injury
Time of the day (1 if between 12 am to 6 am, 0 otherwise)	0.222	0.416	
Width of Median (1 if median width is more than 30 feet, 0 otherwise)	0.940	0.236	
Driving speed (1 if unsafe speed determined by the officer, 0 otherwise)	0.111	0.314	
Logarithm of traffic flow (AADT) per lane	8.692	0.618	
Gender of occupant (1 if male, 0 otherwise) (standard error of parameter distribution)	0.949	0.219	Possible Injury
Age group (1 if age between 45 to 55, 0 otherwise)	0.841	0.365	
Time of the day (1 if between 12 am to 6 am, 0 otherwise)	0.257	0.437	
Surface condition at the time of crash (1 if wet surface, 0 otherwise)	0.172	0.377	
Race of driver (1 if white, 0 otherwise)	0.520	0.499	
Light condition at the time of crash (1 if dark and light outside, 0 otherwise)	0.277	0.447	No-injury (PDO)
Surface condition at the time of crash (1 if dry surface, 0 otherwise)	0.738	0.439	
Road alignment (1 if level surface, 0 otherwise)	0.740	0.438	
Percentage of annual average daily trucks	34.3845	12.2201	
Day of the week (1 if Friday, 0 otherwise)	0.164	0.371	

6.4.1 Descriptive statistics for rural severity model

When considering the roadway geometric, level surface and the median having width more than 30 feet account for 74% and 94% of the sample, respectively. Also, unprotected median accounts for 71.1% of the observations. Environmental characteristics such as clear sky condition, dry and wet surface, and dark but lighted roadway condition account for 65.2%, 73.8%, 17.2% and 27.7% of the sample, respectively. When considering the temporal characteristics of the crashes, time period from midnight to 6 am, day of the week such as Friday, and summer months from June to August account for 22.2%, 16.4%, and 25.1% of the sample, respectively. Now considering the driving maneuvers prior to actions, going straight, hitting roadside objects (i.e., running off the road), and unsafe driving speed accounts for 42.9%, 16.3%, and 11.1% of the sample. Traffic condition measured in terms of logarithm of AADT per lane accounts for 8.69 vehicles per lane per day. Also, percentage of annual daily trucks accounts for 34.38% of total traffic flow. Demographics of drivers such as male drivers, drivers of the age group of 45 to 55 years, and white drivers account for 94.9%, 25.7%, and 52.0% of the sample, respectively.

Correlation with all the variables in the model was also run. The matrix does not indicate any serious correlation. However, the maximum correlation found between percentages of truck traffic has a value of correlation of 0.683 with logarithm of AADT per lane. This is particularly true for rural interstate systems, because truck percentage increases with the increase of traffic flow per lane. Also, time period from 12 am to 6 am and dark but lighted condition has the value of 0.561. This indicates that this time period has dark period as well as early morning daylight period.

Table 6.2: Descriptive Statistics of the Key Variables in Urban Severity Model (N = 11,560)

Meaning of Variables in the Model	Mean	Std. Dev.	Severity outcome
Weather condition at time of crash location (1 if clear sky, 0 otherwise)	0.686	0.464	Fatal
Light condition at the time of crash (1 if day light present, 0 otherwise)	0.743	0.437	
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	0.127	0.333	
Weather condition at time of crash location (1 if clear sky, 0 otherwise)	0.686	0.464	Incapacitating Injury
Time of the day (1 if 3 pm to 7 pm, 0 otherwise)	0.216	0.412	
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	0.127	0.333	
Right shoulder width (ft)	19.729	3.926	
Time of the day (1 if between 12 am to 6 am, 0 otherwise)	0.135	0.342	Non-Incapacitating Injury
Median width at crash location (1 if median is less than 30 feet, 0 otherwise)	0.352	0.478	
Time of the day (1 if between 9 am to 12 pm, 0 otherwise)	0.188	0.391	
Light condition at the time of crash (1 if dark and light outside, 0 otherwise)	0.168	0.374	
Road alignment (1 if level surface, 0 otherwise)	0.777	0.416	
Gender of occupant (1 if male, 0 otherwise)	0.941	0.235	Possible Injury
Surface condition at the time of crash (1 if dry surface, 0 otherwise)	0.841	0.365	
Age group (1 if age between 25 to 35, 0 otherwise)	0.179	0.384	
Days of week (1 if weekend (Saturdays and Sundays), 0 otherwise)	0.128	0.334	
Percentage of annual average daily trucks (%)	10.969	5.004	
Surface condition at the time of crash (1 if dry surface, 0 otherwise)	0.686	0.464	No-injury (PDO)
Percentage of annual average daily trucks (%)	10.969	5.004	
Age group (1 if age between 45 to 55, 0 otherwise)	0.229	0.420	
Surface condition at the time of crash (1 if dry surface, 0 otherwise)	0.841	0.365	
Left shoulder width (ft)	15.027	7.317	

6.4.2 Descriptive statistics for urban severity model

Roadway geometrics such as level surface account for 77.7% of observations. Also, average width of right and left side shoulder is about 19.73 and 15.03 feet, respectively. Median width of less than 30 feet accounts for 35.2% of the sample. Environmental characteristics such as clear sky condition, dry surface, daylight and dark but lighted roadway condition account for 68.6%, 84.1%, 74.3%, and 16.8% of the sample, respectively. When considering the temporal characteristics of the crashes, from midnight to 6 am, 9 am to 12 pm, time 3 pm to 7 pm, and weekend (i.e., Saturdays and Sundays) account for 13.5%, 18.8%, 21.6%, and 12.8% of the sample, respectively. Turning to the driving maneuvers prior to crashes, driving straight or lane keeping maneuvers account for 12.7% of the sample. Turning to exposure parameter such as traffic condition, trucks traffic measured in terms of percent of annual average daily trucks is o 10.97% of total traffic flow. Demographics of drivers such as male drivers, drivers of age group of 25 to 35 years and the age group of 45 to 55 years account for 94.1%, 17.9% and 22.9% of the sample, respectively.

Correlation with all the variables in the model was also run. The matrix does not indicate any serious correlation. However, the maximum correlation was found between the time period from 12 am to 6 am and daylight having the value of 0.549. Also, this period (i.e., from 12 am to 6 am) has a value of correlation of 0.44 with dark lighted condition. This is particularly because this time period has dark period as well as early morning daylight period.

6.5 Results with discussions

The constant for fatal outcome in the urban model is found to be normally distributed with a mean of -18.019 and standard deviation of 4.468. With these estimates, this constant is below zero for 99.9% of large-truck involved crashes resulted in fatalities. That is, about 99% of large truck involved crashes on average were less likely to result in fatal outcomes. The factors are broadly categorized in

the following subsection for injury outcomes in rural and urban settings. Table 6.3 and 6.4 presents the model results of severity models for rural and urban areas, respectively.

Table 6.3: Model Estimates for Rural Severity Model (N = 5,484)

Meaning of Variable	Estimate	t-stat	*p*-Value
Fatal outcome			
Constant	-6.161	-12.943	0.000
Road alignment (1 if level surface, 0 otherwise)	-0.512	-1.630	0.103
Crash with fixed objects (1 if roadside fixed object, 0 otherwise)	0.831	2.313	0.021
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	0.753	1.976	0.048
Weather condition at time of crash location (1 if clear sky, 0 otherwise)	0.625	1.642	0.106
Incapacitating Injury Outcome			
Constant	-3.496	-15.638	0.000
Weather condition at time of crash location (1 if clear sky, 0 otherwise)	-0.925	-6.115	0.000
Months of the year (1 if between June to August, 0 otherwise)	0.485	3.249	0.001
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	0.559	3.764	0.000
Median type (1 if median is unprotected, 0 otherwise)	-0.252	-1.593	0.111
Non-incapacitating Injury Outcome			
Constant	-2.558	-2.128	0.033
Time of the day (1 if between 12 am to 6 am, 0 otherwise)	0.380	2.643	0.008
Width of Median (1 if median width is more than 30 feet, 0 otherwise)	1.441	2.449	0.014
Driving speed (1 if 'unsafe' speed determined by the officers, 0 otherwise)	0.922	5.292	0.000
Logarithm of traffic flow (AADT) per lane	-0.301	-2.739	0.006
Possible Injury Outcome			
Constant	-2.199	-7.284	0.000
Gender of occupant (1 if male, 0 otherwise) (standard error of parameter distribution)	-4.262 (3.161)	-1.866 (1.950)	0.062
Age group (1 if age between 45 to 55, 0 otherwise)	-0.533	-1.734	0.083
Time of the day (1 if between 12 am to 6 am, 0 otherwise)	0.815	2.961	0.003
Surface condition at the time of crash (1 if wet surface, 0 otherwise)	0.668	2.586	0.009
Race of driver (1 if white, 0 otherwise)	-0.518	-1.772	0.076
Non-Injury Outcome (Property-Damage-Only)			
Light condition at the time of crash (1 if dark and light outside, 0 otherwise)	-0.272	-2.713	0.006
Surface condition at the time of crash (1 if dry surface, 0 otherwise)	-0.697	-5.677	0.000
Road alignment (1 if level surface, 0 otherwise)	0.423	4.476	0.000
Percentage of annual average daily trucks	-0.009	-2.597	
Day of the week (1 if Friday, 0 otherwise)	-0.302	-2.798	0.005
Number of observations		5,484	
Restricted log-likelihood		-8826.158	
Log-likelihood at convergence		-3267.571	
Chi-squared value		11117.17	
McFadden pseudo R-squared (ρ^2)		0.629	

Table 6.4: Model Estimates for Urban Severity Model (N = 11,560)

Meaning of Variable	Estimate	t-stat	*p*-Value
Fatal outcome			
Constant (standard error of parameter distribution)	-18.019 (4.468)	-3.304 (2.979)	0.001
Weather condition at time of crash location (1 if clear sky, 0 otherwise)	3.935	2.276	0.023
Light condition at the time of crash (1 if day light present, 0 otherwise)	-1.213	-1.675	0.094
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	4.249	3.298	0.001
Incapacitating Injury Outcome			
Constant	-4.725	-10.070	0.000
Weather condition at time of crash location (1 if clear sky, 0 otherwise)	-1.095	-5.907	0.000
Time of the day (1 if between 3 pm to 7 pm, 0 otherwise)	-0.440	-1.962	0.049
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	0.697	3.676	0.000
Right shoulder width (ft)	0.038	1.835	0.067
Non-incapacitating Injury Outcome			
Constant	-3.960	-17.152	0.000
Time of the day (1 if between 12 am to 6 am, 0 otherwise)	0.603	3.184	0.002
Width of Median (1 if median width is less than 30 feet, 0 otherwise)	0.289	2.082	0.037
Time of the day (1 if between 9 am to 12 pm, 0 otherwise)	0.426	2.375	0.018
Light condition of street (1 if the surrounding area is dark but outside is lighted, 0 otherwise)	0.526	2.858	0.004
Road alignment (1 if level surface, 0 otherwise)	-0.874	-6.216	0.000
Possible Injury Outcome			
Constant	-0.384	-2.344	0.019
Gender of the occupants (1 if male, 0 otherwise)	-1.919	-19.092	0.000
Surface condition at the time of crash (1 if dry surface, 0 otherwise)	-0.787	-4.647	0.000
Age group (1 if age between 25 to 35, 0 otherwise)	-0.335	-2.818	0.005
Days of week (1 if weekend (Saturdays and Sundays), 0 otherwise)	0.216	1.924	0.054
Percentage of annual average daily trucks	-0.027	-2.211	0.027
Non-Injury Outcome (Property-Damage-Only)			
Weather condition at the time of crash (1 if clear weather condition, 0 otherwise)	0.466	5.636	0.000
Age group (1 if age between 45 to 55, 0 otherwise)	0.284	3.354	0.001
Surface condition at the time of crash (1 if dry surface, 0 otherwise)	-0.519	-3.568	0.000
Left shoulder width (ft)	0.009	1.979	0.048
Percentage of annual average daily trucks	-0.029	-3.017	0.003
Number of observations		11,560	
Restricted log-likelihood		-18605.10	
Log-likelihood at convergence		-4286.486	
Chi-squared value		28637.23	
McFadden pseudo R-squared (ρ^2)		0.769	

Table 6.5 and 6.6 presents the average direct pseudo elasticities for rural and urban areas, respectively.

Table 6.5: Average Direct Pseudo Elasticities for Rural Severity Model

Variables	**Elasticity (%)**				
	PDO/ No Injury	**Possible Injury**	**Non-incapacitating Injury**	**Incapacitating Injury**	**Fatal**
Road alignment (1 if level surface, 0 otherwise)	0.23	0.14	0.24	0.22	27.08
Crash with fixed objects (1 if roadside fixed object, 0 otherwise)	-0.26	-0.19	-0.38	-0.38	32.45
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	-0.41	-0.31	-0.63	-0.60	50.91
Weather condition at time of crash location (1 if clear sky, 0 otherwise)	-0.41	-0.29	-0.56	-0.42	49.10
Weather condition at time of crash location (1 if clear sky, 0 otherwise)	1.75	1.17	2.10	-43.66	2.87
Months of the year (1 if between June to August, 0 otherwise)	-0.64	-0.43	-0.73	15.94	-0.79
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	-1.20	-0.86	-1.65	30.29	-2.08
Median type (1 if median is unprotected, 0 otherwise)	0.66	0.45	0.86	-16.43	0.85
Time of the day (1 if between 12 am to 6 am, 0 otherwise)	-0.61	-0.62	11.31	-0.84	-1.07
Width of Median (1 if median width is more than 30 feet, 0 otherwise)	-7.32	-5.13	131.88	-8.44	-9.82
Driving speed (1 if ‘unsafe’ speed determined by the officers, 0 otherwise)	-0.90	-0.64	16.41	-1.23	-1.30
Logarithm of traffic flow (AADT) per lane	13.17	9.13	-236.86	15.02	17.34
Gender of occupant (1 if male, 0 otherwise)	-0.50	7.58	-0.03	-0.26	-0.04
Age group (1 if age between 45 to 55, 0 otherwise)	0.39	-6.48	0.43	0.44	0.43
Time of the day (1 if between 12 am to 6 am, 0 otherwise)	-0.85	14.81	-1.45	-1.10	-1.28
Surface condition at the time of crash (1 if wet surface, 0 otherwise)	-0.43	7.02	-0.37	-0.47	-0.85
Race of driver (1 if white, 0 otherwise)	0.84	-14.04	0.91	0.88	0.94
Light condition at the time of crash (1 if dark and light outside, 0 otherwise)	-1.23	4.68	8.28	7.21	7.45
Surface condition at the time of crash (1 if dry surface, 0 otherwise)	-6.99	26.75	44.36	42.99	49.47
Road alignment (1 if level surface, 0 otherwise)	3.79	-15.52	-23.71	-23.73	-19.44
Percentage of annual average daily trucks	-4.77	17.90	31.25	29.24	29.67
Day of the week (1 if Friday, 0 otherwise)	-0.76	2.77	4.87	4.81	4.89

Table 6.6: Average Direct Pseudo Elasticities for Urban Severity Model

Variables	Elasticity (%)				
	PDO/No Injury	Possible Injury	Non-incapacitating Injury	Incapacitating Injury	Fatal
Weather condition at time of crash location (1 if clear sky, 0 otherwise)	-0.40	-0.33	-0.44	-0.31	211.96
Light condition at the time of crash (1 if day light present, 0 otherwise)	0.08	0.06	0.06	0.07	-40.94
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	-0.26	-0.24	-0.34	-0.33	138.76
Weather condition at time of crash location (1 if clear sky, 0 otherwise)	0.56	0.43	0.52	-39.36	0.64
Time of the day (1 if 3 pm to 7 pm, 0 otherwise)	0.09	0.11	0.08	-6.50	0.03
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	-0.23	-0.30	-0.36	16.73	-0.40
Right shoulder width (ft)	-1.07	-1.24	-1.31	77.11	-0.60
Time of the day (1 if between 12 am to 6 am, 0 otherwise)	-0.28	-0.31	14.62	-0.42	-0.32
Median width at crash location (1 if median is less than 30 feet, 0 otherwise)	-0.24	-0.25	12.54	-0.30	-0.16
Time of the day (1 if between 9 am to 12 pm, 0 otherwise)	-0.17	-0.16	8.67	-0.20	-0.09
Light condition (1 if dark and light outside, 0 otherwise)	-0.28	-0.30	14.38	-0.38	-0.29
Road alignment (1 if level surface, 0 otherwise)	0.98	0.96	-50.29	1.10	0.58
Gender of occupant (1 if male, 0 otherwise)	8.16	-134.45	8.55	9.55	4.12
Surface condition at the time of crash (1 if dry surface, 0 otherwise)	3.27	-53.88	3.43	3.55	2.04
Age group (1 if age between 25 to 35, 0 otherwise)	0.26	-4.28	0.29	0.33	0.13
Days of week (1 if weekend (Saturdays and Sundays), 0 otherwise)	-0.19	3.10	-0.23	-0.26	-0.11
Percentage of annual average daily trucks	1.61	-26.66	1.78	2.04	0.96
Weather condition at the time of crash (1 if clear weather condition, 0 otherwise)	2.29	-23.05	-26.14	-15.66	-23.12
Age group (1 if age between 45 to 55, 0 otherwise)	0.48	-4.78	-4.88	-4.78	-3.03
Surface condition at the time of crash (1 if dry surface, 0 otherwise)	-3.59	34.21	39.49	35.78	25.72
Shoulder width (1 if left shoulder width (ft), 0 otherwise)	1.16	-11.35	-11.76	-12.16	-7.24
Percentage of annual average daily trucks	-0.03	27.30	30.11	29.59	18.69

6.5.1 Drivers' demographics

Among the drivers' demographics, gender, age and race of drivers are found to be statistically significant in both models.

In rural areas, male drivers are less likely to be involved in possible injuries by 7.6% as opposed to female occupants. Injury tolerance of the male body is higher than that of the female body. This variable “male” is normally distributed with a mean of -4.26 3and standard deviation of 3.161. With these estimates, this variable is below zero for 91.13% of large-truck involved crashes resulted in possible injuries. This implies that about 91.1% of large-truck involved crashes on average were less likely to result in possible injury outcomes. Like rural areas, male drivers are less likely to be involved in possible injuries by 134.5% as opposed to female occupants in urban areas. A study by Chen and Chen (2011) found a similar result as females are more likely to be involved in fatal or incapacitating injuries in single and multi-vehicle collisions. This evidence is also supported by a study Abdel-Aty and Abdelwahab (2001).

Considering other demographics, drivers from 45 to 55 years are less likely to be involved with possible injuries by 6.5% in rural areas. Likewise, drivers from 45 to 55 years in urban areas are more likely to be involved with non-injury crashes but less likely to be involved in fatalities, incapacitating, non-incapacitating, and possible injuries by 3%, 4.8%, 4.9%, and 4.8%, respectively. This may due to people of this age group being more experienced and professionally responsive in avoiding maneuvers under critical driving conditions on the urban highways than younger age groups. Considering the race of drivers, white drivers in rural settings are less likely to be involved in possible injuries by 14%. They might be more professional considering driving tasks behind wheel of trucks or passenger vehicles. Considering the younger driver age group, drivers from 25 to 35 years in urban areas are less likely to be involved with possible crashes by 4.3%, this may be due to people of this age group being actively responsive in crash avoiding maneuvers under critical driving conditions on the highways than older age groups.

6.5.2 Driving behavior

Manner of collision (i.e., rear-end, running off the road), and traveling speed are found to be significant in rural and urban models.

In rural areas, running off the road and hitting the roadside objects increases the likelihood of fatalities by 32.5%. This may be capturing high speed vehicles going out of highway sections after losing control and hitting roadside infrastructures such as guard rails, utility poles, trees, etc. A study by Chen and Chen (2011) found that trucks running-off the road for single- and multi-vehicle collisions both increase the likelihood of fatal, incapacitating, non-incapacitating injuries and possible injuries.

Considering another manner of collision, large trucks going straight in the lane (i.e., lane keeping) are more likely to result in fatalities by 51% and incapacitating injuries by 30.3% in rural areas. Likewise, large trucks going straight in the lane (i.e., lane keeping) are also more likely to result in fatalities (increase by 139%) and incapacitating injuries (increase by 17%) in urban areas. This is because passenger vehicles hitting the rear-end of heavy vehicles results in higher severities. In contrasts, with trucks going straight, three possible crash scenarios are likely to occur: trucks being hit in the rear, trucks hitting rear of passenger vehicles and passenger vehicles going under ride the large trucks as angled collisions. According to a 1997 Institute study on fatal crashes between large trucks and passenger vehicles, it is estimated that under-ride occurred in half of these crashes. Of the under-ride crashes, 57 percent involved the front of the truck, 22 percent involved the rear, and 20 percent the side. Additionally, in rural areas "unsafe speed" or speed excessive for the condition is likely to increase non-incapacitating injuries by 16.4%. This may indicate that high speed kinematics in controlling the vehicles and also the severity involved with consequences of energy release of the collision partners. A similar result was found by Chang and Mannering (1999) that unsafe driving speed increases likelihood of fatal or incapacitating injuries for truck involved crashes.

6.5.3 Traffic characteristics

Traffic flow and percentage of annual average truck traffic are found to be significant in the rural and urban models.

Traffic flow per lane (logarithm of AADT per lane) on the rural sections, decreases the likelihood of non-incapacitating injuries by 237%, it also increases the likelihood of fatalities, incapacitating injuries, possible injuries and PDO by 17.4%, 15%, 9.2%, and 13.2%, respectively. Considering the truck flow in rural areas, an increase of average annual daily trucks in rural areas decreases the likelihood of no injuries, but increases fatalities, incapacitating, non-incapacitating, and possible injuries by 29.7%, 29.3%, 31.3%, and 18%, respectively.

Likewise, an increase of average annual daily trucks in urban areas decreases the likelihood of possible injuries by 27% as well as no injury crashes, but it increases the likelihood of fatalities, incapacitating, non-incapacitating, and possible injuries by 19%, 30%, 30.1%, and 27.3%, respectively. Because of a high percentage of trucks, passenger vehicle drivers are perhaps more cautious while driving along the same highway system in rural areas. A study by Chen and Chen (2011) found that a low truck percentage (less than 10%) indicator decreases the probability of both possible injury or non-incapacitating injury and incapacitating injury or fatal but increases no injuries for the multi-vehicle collisions. Another study by Milton and Mannering (1998) found that percentage of single-unit truck or percentage of trucks were both likely to reduce frequency of crashes because vehicle overtaking and lane changing events decrease and result in less risk taking driving maneuvers (Miaou, 1994).

6.5.4 Temporal characteristics

Time of the day, days of week, months of year, and lighting condition indirectly indicate the time of day are found to be significant in both models.

Summer months, particularly June to August, are likely to increase incapacitating injuries by 16% in rural settings. This may be capturing traffic interactions relative to other times of the year. Chang and Mannering (1999) found that summer months are likely to increase crashes for four or more occupants in the vehicles. Considering the time of day, the time period from 12 am to 6 am increases the likelihood of non-incapacitating injuries by 11.3% and possible injuries by 14.8%. Likewise, this time period from 12 am to 6 am increases the likelihood of non-incapacitating injuries by 15% in urban areas. This time period is possibly capturing some late night driving conditions which could be highly likely to be influenced by drowsy or fatigue driving. Also, passenger vehicle drivers could have drunk driving cases in the late night. Another time period in the evening peak hours, evening crashes occurring between 3 pm to 7 pm reduce the likelihood of incapacitating injuries by 6.5% in urban areas. This time period captures the afternoon peak while people are going back from work to their residences. Since congestion is very likely during this time period, the speed of vehicles is also relatively slower than other non-peak time. A similar result is also reflected in a study by Doherty et al. (1998) with a fact that accident rates with all severity increases towards evening (2000 – 2359) and late night to early morning (2400 – 0459). Also, light condition, particularly dark surrounded by lights outside, decreases the likelihood of no-injuries, but increases fatalities, incapacitating, non-incapacitating, possible injuries by 7.5%, 7.2%, 8.3%, and 4.7%, respectively in rural settings. This condition clearly indicates that the driving under such conditions in rural areas may not pose risks because the drivers get used to drive after driving for a while along the rural sections particularly for the passenger vehicle drivers. Also, the situation is true for the large truck drivers, who are professionally trained to drive under such conditions. Day of week such as Friday (i.e., the starting of weekend) is likely to reduce no injuries. This infers Friday traffic likely to increase likelihood of serious injuries. This is because of more traffic going out for family or relative visits, scenic visits, or vacation tours for the weekend. Turning to weekends, Saturdays and Sundays increase the likelihood of possible injuries by 3.1% in urban areas. This is

mainly true for the weekend traffic interactions particularly for passenger vehicles in the urban highways due to social interaction with friends and families, weekend visits, and nightlife at urban city centers pose serious risks for the passenger vehicle drivers. A study by Chang and Mannering (1999) found that weekends increase the likelihood of PDO crashes for truck involved crashes.

Considering the off peak period in the morning, the time period between 9 am to 12 pm increases the likelihood of non-incapacitating injuries by 8.7% in urban areas. After the morning peak hours, the average traveling speed increases in the urban interstates, and the likelihood of large-truck involved crashes resulting in non-incapacitating injuries goes up. IIHS 2010 study showed that 17% large truck fatal crashes occurred between 9 am to 12 pm. A similar result was also reflected in a study by Ivan et al. (2000) that daylight conditions between 10:00 and 15:00 hour indicates an increase in multi-vehicle collisions compared to the base time: 06:00 and 10:00 hour.

6.5.5 Roadway geometrics

Terrain, median type and width, and shoulder (i.e., left and right) width are found to be significant in rural and urban models.

Level surface in rural areas reduces the likelihood of fatalities by 27%. Also, it increases likelihood of no injuries, but it decreases fatalities, incapacitating, non-incapacitating, and possible injuries by 19.5%, 23.8%, 23.7%, and 15.6%, respectively. Likewise, driving on a level surface decreases the likelihood of non-incapacitating injuries by 50.3% in urban areas. This is because visibility is not obstructed by any geometrics of longitudinal and vertical elements relating to uphill, downhill or curvature which can pose serious risks to the passenger vehicles rather than large trucks.

Considering the median types, unprotected median on rural highways is likely to decrease incapacitating injuries by 16.5%. This captures the cautious driving behavior of the truck drivers on rural highways section where medians are not properly protected with a barrier (which might be

depressed or raised median). A study by Chen and Chen (2011) found opposite results as found in this study. They found that unprotected median indicators decrease the probability of possible injury/non-incapacitating injury by 24.1%, while also increasing the probability of fatal/incapacitating injury by 37% in multi-vehicle collisions. Next considering the width of median, median width of more than 30 feet increases the likelihood of non-incapacitating injuries by 132%. On rural highways, more than 30 feet of median could pose risks to the drivers because of more safety margins. A study on single- and multi-vehicle large truck crashes on rural highways showed that wider median (more than 60 feet) reduces the likelihood of possible and non-incapacitating injuries. The main reason behind was explained as wide lanes and wide medians provide more physical safety margins for truck drivers although there could be cases of risk compensation. This compensation may be highlighted as “safer” feeling encouraging unsafe driving behavior by the truck drivers.

Considering the shoulder width, right shoulder width increases the likelihood of incapacitating injuries by 77%. The average width of the right shoulder is 19.7 feet. There is a possibility of the presence of roadside infrastructure or utility poles inside the shoulder which is very likely to be found in urban settings in Texas. Because of inattentive driving, increasing right side shoulder width is closely related to running off the road crashes. As a result of running off the road or losing control, hitting the roadside objects could potentially result in incapacitating injuries. These types of crashes are particularly related to human factors – inattentive driving behaviors, distractions and overcorrection. According to data analysis from National Motor Vehicle Crash Causation Survey (NMVCCS), a single vehicle running off the road indicates driver’s errors such as overcompensation, poor directional control, too fast for condition, sleepiness and physical impairment, recognition error, and distraction. However, increase in left shoulder width in urban areas is very likely to increase the likelihood of no-injury crashes and reduce the likelihood of fatalities, incapacitating, non-incapacitating, and possible injuries by 7.3%, 12.2%, 11.8%, and 11.4%, respectively. This is because drivers feel safer with a wider left shoulder,

which provides more cushion in case of emergency or sudden recovery maneuvers. A study by Milton and Mannering (1998) found that narrow left shoulder (less than 1.5 meter wide) is likely to increase crashes on principle arterials. This finding indirectly supports our findings. While considering the median width in urban areas, median width of less than 30 feet increases the likelihood of non-incapacitating injuries by 12.6%. In urban highways less than 30 feet median could pose risks to the drivers of less safety margins.

6.5.6 Environmental characteristics

Clear sky condition, pavement condition, daylight condition, and dark but lighted condition are found to be significant in the rural and urban models.

Crashes in the clear sky condition are likely to result in increases in fatalities by 49.1%, but decreases in incapacitating injuries by 44% in rural areas. This variable captures some of the driving behavior in the clear sky condition. Likewise, crashes in the clear sky condition are likely to result in more fatalities (increase by 212%) but less incapacitating injuries (reduces by 39.4%) in urban areas. This variable captures some of the driving behavior in the clear sky condition. The large truck drivers including other drivers are comfortable to drive faster or taking more risks during the clear sky condition with an advantage of better visibility of other passenger vehicles down the rural or urban highway stretches. A similar result is supported by a study by Edwards (1998), indicating that accident severity decreases significantly in rain compared with fine weather. It is also found that clear weather condition in urban areas triggers some dangers to the passenger vehicle drivers with non-injury outcome by 2.3% but decreases the likelihood of fatalities, incapacitating, non-incapacitating, and possible injuries by 23.2%, 16%, 26.2%, and 23%, respectively because the friendly weather condition makes the drivers more relaxed and risky on the benefit of better visible driving condition along the urban highway sections. A detailed study by Morgan and Mannering (2011) indicated that for female drivers less than

45 years old driving on a wet surface, clear weather condition increases the likelihood of non-injury crashes. Also, this study (Morgan and Mannering, 2011) found that for both male and female drivers under 45 years old driving on snow/icy surface, clear weather condition reduces the likelihood of minor injuries (i.e., combined incapacitating and possible injuries).

Considering the surface condition, crashes occurring on dry pavement in rural areas reduce the likelihood of no injuries, but increases 49.5%, 43%, 44.4%, and 27%, respectively in rural areas. Any evasive actions on dry surface are very effective as opposed to wet surface in terms of effective skidding of tires on dry surface. Likewise, crashes occurring on dry pavement in urban reduce the likelihood of possible injuries by 54% and non-injury crashes. But it increases the likelihood of fatalities, incapacitating, non-incapacitating, and possible injuries by 25.7%, 35.8%, 39.5%, and 34.2%, respectively. Any evasive actions on dry surfaces are very effective as opposed to wet surfaces in terms of effective skidding of tires on dry surface. A more detailed study by Morgan and Mannering (2011) indicated for male drivers less than 45 years old having less likelihood of minor injuries (i.e., combined non-incapacitating and possible injuries) in case of single occupant vehicle, in urban settings and dark road condition. Another study by Chang and Mannering (1999) indicated that dry surface increases the likelihood of fatal, injuries for truck involved crashes. This study (Chang and Mannering, 1999) also found that given the single-occupant non truck involved crashes occurred, the likelihood of PDO increases by 6.7%.

However, crashes occurring on wet pavement in rural areas increase the likelihood of possible injuries by 7%. Any evasive actions on wet surface are ineffective as opposed to dry surface in terms of effective skidding of tires on dry surface.

Considering daylight condition, crashes during the daylight period are more likely to result in fewer fatalities (decrease by 41%). This variable indicates the light condition is crucial to understand relative speed and distance between the vehicles and maintain safe gap in the vehicle streams in the

urban areas. Visibility clearly reduces the likelihood of fatalities in urban areas. A mixed result was found in a study by Golob and Regan (2002) on large-truck involved crashes. Likelihood of crashes decreases in 7 am to 8 am, increases in 8 am to 9 am, 9 am to 2 pm, and continues to decrease from 2 pm to 5 pm on urban freeways. Light condition, particularly dark condition surrounded by lights outside in urban settings, increases the likelihood of non-incapacitating injuries by 14.4%. This condition clearly indicates that driving in such conditions in urban areas poses serious risks particularly for the passenger vehicle drivers, whereas, the large truck drivers are professionally trained to drive through these conditions. A similar result was found in a study by Abdel-Aty that driving fast in a dark environment without street light could be more risky than that with street lights. This condition is very likely to occur in urban interstates where speed of the traveling vehicles is relatively higher which could result in non-incapacitating injuries because of slower reaction and lower perception ability. This reflects the situation of drivers having increased reaction time and better perception ability in environments with good street lighting (Huang et al., 2008). A similar result in another study by Chen and Chen (2011) found that multi-vehicle collisions in the dark night with lighted condition increases the likelihood of fatal, incapacitating, and non-incapacitating injuries.

Considering the left shoulder width, increase in left shoulder width is very likely to increase the likelihood of no-injury crashes and reduce the likelihood of fatalities, incapacitating, non-incapacitating, and possible injuries by 7.3%, 12.2%, 11.8%, and 11.4%, respectively. This is because of drivers feel safer with wider left shoulder which is provides more cushion in case of emergency or sudden recovery maneuvers. A study by Milton and Mannering (1998) found that narrow left shoulder (less than 1.5 meter wide) likely to increase crashes on principle arterials. This finding indirectly supports our findings.

6.6 Model specification test

The likelihood ratio test was conducted to justify the statistical significance of rural and urban injury model separately from the combined model. This is because other studies (Morgan and Mannering, 2011; Geedipally and Lord, 2010) have shown that separate models are statistically significant to capture more factors than a combined model. The null hypothesis is set as the separate models are not statistically and significantly different from the combined model. The following likelihood ratio test was conducted to test the hypothesis (Washington et al., 2011):

$$\chi 2 = -2 * [LL_T(\beta^T) - LL_R(\beta^R) - LL_U(\beta^U)] \quad (6.7)$$

where,

$LL_T(\beta^T)$ *is loglikelihood at convergence of the combined model*

$(-32985.43, degree\ of\ freedom, n_T = 35)$

$LL_R(\beta^R)$ *is loglikelihood at convergence of rural model*

$(-3267.571, degree\ of\ freedon, n_R = 27)$

$LL_U(\beta^U)$ *is loglikelihood at convergence of urban model*

$(-4286.486, degree\ of\ freedom, n_U = 27)$

The test follows χ^2 distribution with degrees of freedom equal to the sum of rural and urban model minus that of total model (combined of rural and urban model). With 19 degrees of freedom and χ^2 value of 50862.746, the test statistic indicates that we reject the null hypothesis. This leads to the fact that separate models (i.e., rural and urban models) are statistically and significantly different from the combined model with *P*-value being less than 0.0001 (two-tailed *P*-value). So, the test result indicates that significance of injury severity should be modeled as rural and urban models separately rather than combined one. Additionally, respective multinomial and mixed logit model specification were

conducted for rural and urban areas. It is found that mixed logit is significantly different than multinomial model with 10% level of significance for rural and urban areas.

6.7 Conclusion and future directions

Five years of Texas CRIS data from 2006 to 2010 on large-truck involved crashes on interstate system were studied. In this study, separate mixed logit models were estimated to analyze the injury severity of truck drivers on rural and urban areas in Texas. Also, the result of the likelihood ratio test indicates that the injury outcomes for rural and urban settings involving truck crashes are clearly distinct. So, it is statistically conclusive to estimate separate models for rural and urban areas with the significant contributing factors. These models clearly demonstrate the risk factors covering driver's demographics, driving behavior, traffic characteristics, temporal characteristics, roadway geometrics, and environmental characteristics. This is one the first of in its attempt to analyze injury outcomes of large-truck involved crashes on interstate system in Texas to identify the risk factors to influence injury severities of large truck drivers.

The findings from rural and urban mixed logit models clearly demonstrate the differences of factors influencing injury outcomes large tuck involved crashes in Texas and also add to the current literature on large-truck involved crashes and injury outcomes. Understanding these risk factors can be utilized in developing effective countermeasures to reduce the likelihood of injury outcomes. The benefitted stakeholders of this study are trucking industry and related agencies, enforcement agencies, transportation safety professionals and above all general public.

Table 6.7: Summary of Variables and their Different Effects in Rural and Urban settings

Significant Variables in Rural Model (not present in Urban model)		Significant Variables in Urban Model (not present in Rural model)	
Hitting with fixed objects Outcome: Fatality	↑	-	
Unprotected median Outcome: Incapacitating injury	↓	Right shoulder width Outcome: Incapacitating injury	↑
Summer months (June to August) Outcome: Incapacitating injury	↑	Time of day – 3 pm to 7 pm Outcome: Incapacitating injury	↓
Median width more than 30 feet Outcome: non-incapacitating injury	↑	Time of day – 9 am to 12 pm Outcome: Non-incapacitating injury	↑
Unsafe driving speed Outcome: Non-incapacitating injury	↑	Median width less than 30 feet Outcome: Non-incapacitating injury	↑
Logarithm of vehicles per lane Outcome: Non-incapacitating injury	↓	-	
Race of driver (i.e., white) Outcome: Possible injury	↓	Weekend – as days of week Outcome: Possible injury	↓
Wet roadway surface Outcome: Possible injury	↑	Driver of age group 25 – 35 years Outcome: Possible injury	↓
Friday – as day of week Outcome: PDO	↓	Left shoulder width Outcome: PDO	↑

Table 6.8: Trend of Variable Effects on Rural and Urban Models

Rural Model		Urban Model	
Road alignment – level surface Outcome: Fatalities	↓	Road alignment – level surface Outcome: Non-incapacitating injuries	↓
Lane keeping – going straight Outcome: Fatalities and Incapacitating injuries	↑	Lane keeping – going straight Outcome: Fatalities and Incapacitating injuries	↑
Clear sky condition Outcome: Fatalities Incapacitating injuries	↑ ↓	Clear sky condition Outcome: Fatalities Incapacitating injuries	↑ ↓
Time period between 12 am to 6 am Outcome: Non-incapacitating and possible injuries	↑	Time period between 12 am to 6 am Outcome: Non-incapacitating injuries	↑
Dark but lighted condition Outcome: Injuries – K,A,B,C	↑	Dark but lighted condition Outcome: Non-incapacitating injuries	↑
Male occupants Outcome: Possible injuries	↓	Male occupants Outcome: Possible injuries	↓
Percentage of daily truck traffic Outcome: Injuries – K,A,B,C	↑	Percentage of daily truck traffic Outcome: Injuries – K,A,B,C	↑
Dry surface condition Outcome: Injuries – K,A,B,C	↑	Dry surface condition Outcome: Injuries – K,A,B,C	↑

Modeling large-truck involved crashes clearly revealed that there are significant differences in factors related to crashes occurring in the rural and urban areas in Texas. Also, there are some contributing factors present both in rural and urban severity models but some are not (see Table 6.7 and Table 6.8). However, the injury outcomes are different because of effective medical response, difference in traveling speed, driving behavior, and roadway geometrics. The variables that are present in rural models but not present in urban model includes (see Table 6.7): running off the road (hitting fixed roadside objects), unprotected median, median larger than 30 feet, unsafe driving, logarithm of AADT per lane, driver's demographics and day of week (i.e., starting of weekend – Friday), and wet surface. Obviously, out of these nine factors, five of these factors directly lead to severe injury outcomes in rural areas. In case of unprotected median, drivers traveling along the rural stretches try to avoid running off the road and drive through depressed and median of different elevations which results in less likelihood of incapacitating injuries. Likewise, higher traffic condition (logarithm of AADT per lane) reduces the likelihood of non-incapacitating injuries. As found in the model, white drivers are safe driving along the rural stretches because the likelihood of white drivers being possibly injured reduces. On the contrary, Friday which is the starting day of the weekend increases the likelihood of other severe injury outcomes. Now turning to urban model for the variables not present in the rural model includes right and left shoulder width, time of day (9 am to 12 pm, 3 pm to 7 pm), median narrower than 30 feet, weekend (i.e., Saturdays and Sundays), and drivers' age group (25-35 years). Out of seven factors, three factors directly lead to severe injuries (See Table 6.7). With that listed variables, right shoulder width, time period between 9 am to 12 pm and median less than 30 feet increases the likelihood of incapacitating, and non-incapacitating injuries. Also, looking at the factors present in both models showed similar trend of effects on injury outcomes, and thus provides some interesting insights regarding spatial distribution of crashes and associated outcomes in Texas. Turning to factors presented in Table 6.8, level surface reduces likelihood of fatalities for rural areas but non-incapacitating injuries for urban areas. Also, Time

period from 12 am to 6 am increases the likelihood of non-incapacitating injuries and possible injuries in rural areas, whereas, this time period increases the likelihood of non-incapacitating injuries in urban areas. Additionally, dark condition surrounded by light in rural areas increases the likelihood of severe injuries (i.e., fatalities, incapacitating, non-incapacitating, and possible injuries), whereas, same condition results in only non-incapacitating injuries in urban areas.

Furthermore, the models results presented in Table 6.8 indicates that roadway characteristics (level surface), driving maneuvers (lane keeping), weather condition (clear sky condition), time of day (12 am to 6 am), dark but lighted condition, demographics of drivers (male drivers), truck traffic condition (percentage of trucks), surface condition (dry condition) have similar effects on injury outcomes, independent of geographical regions – rural and urban areas in Texas.

The limitation of this study is the different types of trucks included in the CRIS database, and single state crash data source. However, this dataset could be further separated into different time of day to uncover a complete understanding on large-truck involved crashes for rural and urban settings in Texas.

Chapter 7: Modeling Injury Outcomes of Crashes involving Large Truck Crashes by Time of Day in Urban Areas in Texas

7.1 Introduction

The measure of traffic safety in terms of exposures such as number of crashes per Annual Average Daily Traffic (AADT), Vehicle-miles Traveled (VMT), and other measures are very aggregate in nature where temporal effects on crash occurrence is not explicitly reflected (Qin et al., 2006). Distribution of daily traffic and weekly traffic is different from each other and this difference is also valid for different geographical locations such as rural and urban areas. The factors associated with crashes vary over time of day particularly for different time periods such as morning peak (AM peak) and evening peak (PM peak) (Jovanis and Delleur, 1983). Furthermore, the driving alertness (or sleeping pattern) and performance vary across time of day following the circadian cycle (Lenne´ et al., 1997; Qin et al., 2006; Garbarino et al., 2000; Langlois et al., 1983). Consequently, crash occurrence and associated injury severities vary between AM, PM, and off peak periods and also vary over different geographical locations.

In general, past studies suggest that drivers react differently to different changing driving environment because of traffic, weather, road surface, light condition which are very naturally subject to vary based on the time of day. More importantly, driving behavior and performance is also influenced by time of day because of the circadian cycle. Thus, there is a complex interaction of time of day and individual driving characteristics independent of professional and non-professional drivers. This is particularly critical in large-truck involved crashes, where a variety of human factors including driving experiences, reactions to different road environmental conditions, driving performance, visual acuity, attentiveness, are very likely to change over the time of day. Concerning the severity outcomes of the

large truck crashes, it is critical and important to investigate the factors associated with different levels of severities varying over the time of day (for instance, over the time periods of AM, PM, and off peak) in large-truck involved crashes. Several past research efforts focusing on injury severity have indicated that time of day or time periods of a day may influence crash outcomes.

A number of studies investigated the crash likelihood or risks associated with crash through variation of time of day in real and simulated driving environments (Qin et al., 2006; Reimer et al., 2007; Lenne´ et al., 1997; Langlois et al., 1985). Current research efforts are limited in the scope of overall severity analysis without considering the time periods on different geographical locations in a state (for instance, rural and urban areas in Texas). These studies analyzed injury severity outcomes through statistical models that very often capture general effects by utilizing indicator variables for the time of day; for example, taking AM peak or PM peak of late night time periods. As a result, these studies are not uncovering the effects of these time periods, and drivers' characteristics (for instance, demographics, driving behavior, alertness over the time of day, etc.) on the injury outcomes where driving patterns change and adjust according to the traffic flow varying over the time of day. An alternative to using indicator variable for different time periods which capture only overall effects, is to split the dataset to several subsets based on AM peak, PM peak, and off peak periods and estimate the models to allow the parameters to vary from one data subset to another (Morgan and Mannering, 2011). The objective of this study to estimate the parameters for three different time periods of a day – AM peak (6 to 9 am), PM peak (4 to 7 pm), and off peak (other than AM and PM peak), and to identify the contributing factors through random parameter logit model for large-truck involved crashes in urban areas on Texas highways. This study also demonstrates the significance of using split dataset for three different time periods (e.g., AM, PM and off peak) rather than a single and combined dataset for these time periods.

7.2 Statistical Approach

Several research studies utilized a number of methodological approaches such as the multinomial logit, ordered probit, Bayesian Ordered Probit, nested logit, and mixed logit models. Each of these models have been utilized to model injury severities (Chistoforou et al., 2010; Lemp et al., 2011; Milton et al., 2008; Zhu and Srinivasan, 2011; Duncan et al., 1998; Xie et al., 2009; Khorashadi et al., 2005; Chang and Mannering, 1999). In this study, we estimate a mixed logit model for injury severity of crashes involving large trucks by considering random parameters in the developed model following the similar logical research framework as of past research studies (Milton et al., 2008; Gkritza and Mannering, 2008; Chen and Chen, 2011). The level of injury is discrete in nature as coded in the injury scale KABCO (i.e., 'K' for Fatal, 'A' for Incapacitating injury, 'B' for Non-incapacitating Injury, 'C' for Possible Injury and 'O' for Property-Damage-Only), and a mixed logit model has been widely accepted to model the effects of several contributing factors on the levels of injury severity. Several research studies conducted by Revelt and Train (1997, 1999), Train (1997), Brownstone and Train (1999), McFadden and Train (2000), and Bhat (2001) have clearly pointed to the effectiveness of this methodological approach. Although discrete outcome severity could be modeled by a multinomial logit model, heterogeneous effects and correlation in unobserved factors could still have potential limitations in the assumption behind utilizing this model for injury severity (Train, 2009). Thus, a mixed logit model overcomes all of these limitations by generalizing multinomial logit structure, allowing for the parameters $\boldsymbol{\beta}_i$ vector to vary across the observation of crashes (Savolainen et al., 2011). The assumption regarding IID (independently and identically distributed errors), IIA (independence of irrelevant alternatives) and unobserved heterogeneity associated with observations of the multinomial logit model are completely relaxed by introducing a mixed logit approach (Jones and Hensher, 2007).

In order to achieve a better understanding of the injury severity of large-truck involved crashes on the US interstate system, we seek to develop a statistical model that can be used to determine the

contributing factors that influence the likelihood of severity outcomes in large-truck involved crashes. To do so, we start with a linear function S_{in} that determines discrete injury severity outcome i (fatality, incapacitating injury, non-incapacitating injury, possible injury, and property damage only) for large-truck involved incident *n* such that (Washington et al., 2011):

$$S_{in} = \boldsymbol{\beta}_i \mathbf{X}_{in} + \varepsilon_{in} \tag{7.1}$$

where $\mathbf{X}_{in}$ is vector of explanatory variables (driver, vehicle, road, and environment variables), $\boldsymbol{\beta}_i$ is vector of estimable parameters, ε_{in} is the error term. If ε_{in}'s are assumed to be generalized extreme value distributed, McFadden (1981) has shown that the multinomial logit results such that:

$$P_n(i) = \frac{EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]}{\sum_I EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]} \tag{7.2}$$

where $P_n(i)$ is probability of large-truck involved incident *n* having severity outcome i ($i \in I$ with I denoting all possible injury severity outcomes)

As CRIS crash data is likely to have a significant amount of unobserved heterogeneity because the information regarding any of the factors are not obtained from the in-depth crash investigation or reconstruction studies (for example, relating to police reporting, roadway, vehicle, and driver factors), we consider the possibility that elements of the parameter vector $\boldsymbol{\beta}_i$ may vary across observations of each large-truck involved crash by using a random-parameters logit model (also known as the mixed logit model). Previous works by McFadden and Rudd (1994), Geweke et al. (1994), Revelt and Train (1997, 1999), Train (1997), Stern (1997), Brownstone and Train (1999), McFadden and Train (2000), and Bhat (2001) have shown that the development and effectiveness of the mixed logit approach can explicitly account for the variations (across crash observations) of the effects that variables have on the severity outcomes (or choices) considered in this study. The mixed logit model is written as (see Train, 2003):

$$P_{in} = \int \frac{EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]}{\sum_I EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]} f(\boldsymbol{\beta}_i|\boldsymbol{\varphi}) d\boldsymbol{\beta}_i \tag{7.3}$$

where, $f(\boldsymbol{\beta}_i|\varphi)$is the density function of $\boldsymbol{\beta}_i$, $\boldsymbol{\varphi}$ is a vector of parameters of the density function (mean and variance), and all other terms are as previously defined. This model can now account for severity outcome specific variations of the effect of $\mathbf{X}$ on large-truck involved crash probabilities, with the density function $f(\boldsymbol{\beta}_i|\boldsymbol{\varphi})$ used to determine $\boldsymbol{\beta}_i$. Mixed logit probabilities are then a weighted average for different values of $\boldsymbol{\beta}_i$ across the crash observations where some elements of the vector $\boldsymbol{\beta}_i$ may be fixed and some randomly distributed. If the parameters are random, the mixed logit weights are determined by the density function $f(\boldsymbol{\beta}_i|\boldsymbol{\varphi})$ (Milton et al., 2008; Washington et al., 2011).

Maximum likelihood estimation of the mixed logit model shown in Equation (7.3) is undertaken with simulation approaches due to the difficulty in computing probability. The most widely accepted simulation approach is a technique developed by Halton (1960) to generate a systematic non-random sequence of numbers. Halton's technique (known as Halton draws) has been shown to provide a more efficient distribution of the draws for numerical integration than purely random draws (Bhat, 2003; Train, 1999).

A study by Ye and Lord (2011) showed the influence of the sample size on injury severity modeling. Although the analysis by Ye and Lord (2011) regarding the sample size corresponding to each severity model is simulation driven, there are still a few findings that could be generalized in terms of sample size for the three commonly used models. An ordered probit model is the one that requires the least samples (more than 1000), a mixed logit is the most demanding on samples (more than 5000), and a multinomial logit model requirement are located between the ordered probit and mixed logit models (somewhere more than 2000). In our study, the rural crash sample size is 5,484 and the urban crash sample size is 11,560, which are above and over the safe threshold regarding sample size identified by Ye and Lord (2011).

In order to estimate the impact of particular variables on the injury-outcome likelihood, elasticities (or direct-pseudo elasticity) are computed. In the context of the current injury severity model, most of the variables are indicator in nature; direct-pseudo elasticities are estimated to measure the marginal effects of indicator variables when any particular indicator variable switches from 0 to 1 or reverse (Washington et al., 2011). Also, this is translated to percentage change in the likelihood of the indicator variables switching between 0 and 1 or 1 to 0. For binary indicator variables, the direct-pseudo elasticity is estimated as follows (Kim et al., 2010):

$$E_{x_{nk}}^{P_{in}} = \frac{P_{in}[given\ x_{nk} = 1] - P_{in}[given\ x_{nk} = 0]}{P_{in}[given\ x_{nk} = 0]} \tag{7.4}$$

where, P_{in} is given the Equation (7.3) and simulated as shown in Equation (7.5).

x_{nk} = the k-th independent variable associated with injury severity *i* for observation n.

The unconditional probability in Equation (7.3) (Kim et al., 2010) can be estimated with an unbiased and smooth simulator (McFadden and Train, 2000) that is computed as (Walker and Ben-Akiva, 2002):

$$\hat{P}_{in} = \frac{1}{R}\sum_{r=1}^{R} P_{in} = \frac{1}{R}\sum_{r=1}^{R} \int \frac{EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]}{\sum_I EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]} f(\boldsymbol{\beta}_i|\boldsymbol{\varphi}) d\boldsymbol{\beta}_i \tag{7.5}$$

where, R = the total number of draws.

Since the direct pseudo-elasticity is calculated for each observation, it is usually reported as the average direct pseudo-elasticity (taking average over the sample) as a measure of the marginal effect of an indicator variable on the likelihood of a particular injury severity outcome (Kim et al., 2010).

With the simulator in Equation (7.5), Maximum Simulated Likelihood Estimation (MSLE) can be used to estimate parameters and this MSLE estimator is asymptotically normal and consistent (Lee, 1992):

$$max_{\boldsymbol{\beta}_{in}} \sum_{n=1}^{N} \sum_{i=1}^{I} y_{in} ln\hat{P}_{in} \tag{7.6}$$

where, N = the total number of observations (i.e., crashes in the sample)

y_{in} = 1 if individual n suffers from injury severity i, 0 otherwise.

7.3 Empirical Settings

The data used in this study were collected from the Texas Peace Officers' Crash Reports, commonly known as the CRIS database. To investigate human, vehicle, and road-environmental factors, a sample of 11,560 data observations were extracted from the CRIS database by filtering crashes involving large trucks on the interstate system over a period of five years from 2006 to 2010. Then, urban and rural areas were further filtered based on the definition population where population more than 200,000 was classified as urban area. In the data processing stage, the vehicle body style (i.e., VEH_BODY_STYL_ID in the main data recording system) was set to truck, truck tractor, semi-trailers and highway facility type was selected for interstates system which was processed in SAS software. Each observation in the sample is a crash representing the maximum level of injury of the drivers, involving at least one large truck in the interstate system. The crash dataset was linked to the vehicle and person datasets through appropriate linking variables, namely crash number, and the vehicle and person datasets were linked through the vehicle and crash number in SAS software. Linking of the three data components (i.e., crash, vehicle, and person) was processed to have a single observation with maximum injury level of the drivers involved the crash with unique Crash ID. Then, urban data sample was further split into three separate subsets based on time of day such as AM peak (6 to 9 AM) having a sample of 2,409 observations; PM peak (4 to 7 PM) having a sample of 2,502 observations; and off peak (other than AM and PM peak periods) having a sample of 6,649 observations.

Then, a discrete choice framework for the data was followed using the Limdep software to set up the mixed logit models. Descriptive statistics of the variables used in the model (i.e., dependent variable – MAX_SEVS and all the independent variables) are presented in Figure 7.1 and Table 7.1. Table 7.1 show the mean and standard deviation of the key variables considered in the model. Figure 7.1 presents dependent variable (i.e., MAX_SEVS) having five levels of injury outcomes such as fatality, incapacitating, non-incapacitating, possible, and no injury (or PDO) account for 0.1% for AM peak, 0.1% for PM peak, and 0.3% for off peak periods; 1.0% for AM peak, 1.0% for PM peak, and 1.7% for off peak periods; 1.0% for AM peak, 1.6% for PM peak, and 2.3% for off peak periods; 5.7% for AM peak, 4.7% for PM peak, and 6.1% for off peak periods; and 92.2% for AM peak, 92.6% for PM peak, and 89.7% for off peak periods of the total observations of AM peak subset (N=2,409), PM peak subset (N=2,502), and off peak subset (N=6.649), respectively. Although percentage of fatality is low (approximately between 0.10 to 0.30% of the entire sample), five distinct injury levels (e.g., disaggregate level of injury severity) provide a clearer picture of contributing factors leading to those different levels rather than combining the two injury outcomes (such as combining fatality and incapacitating injury outcomes).

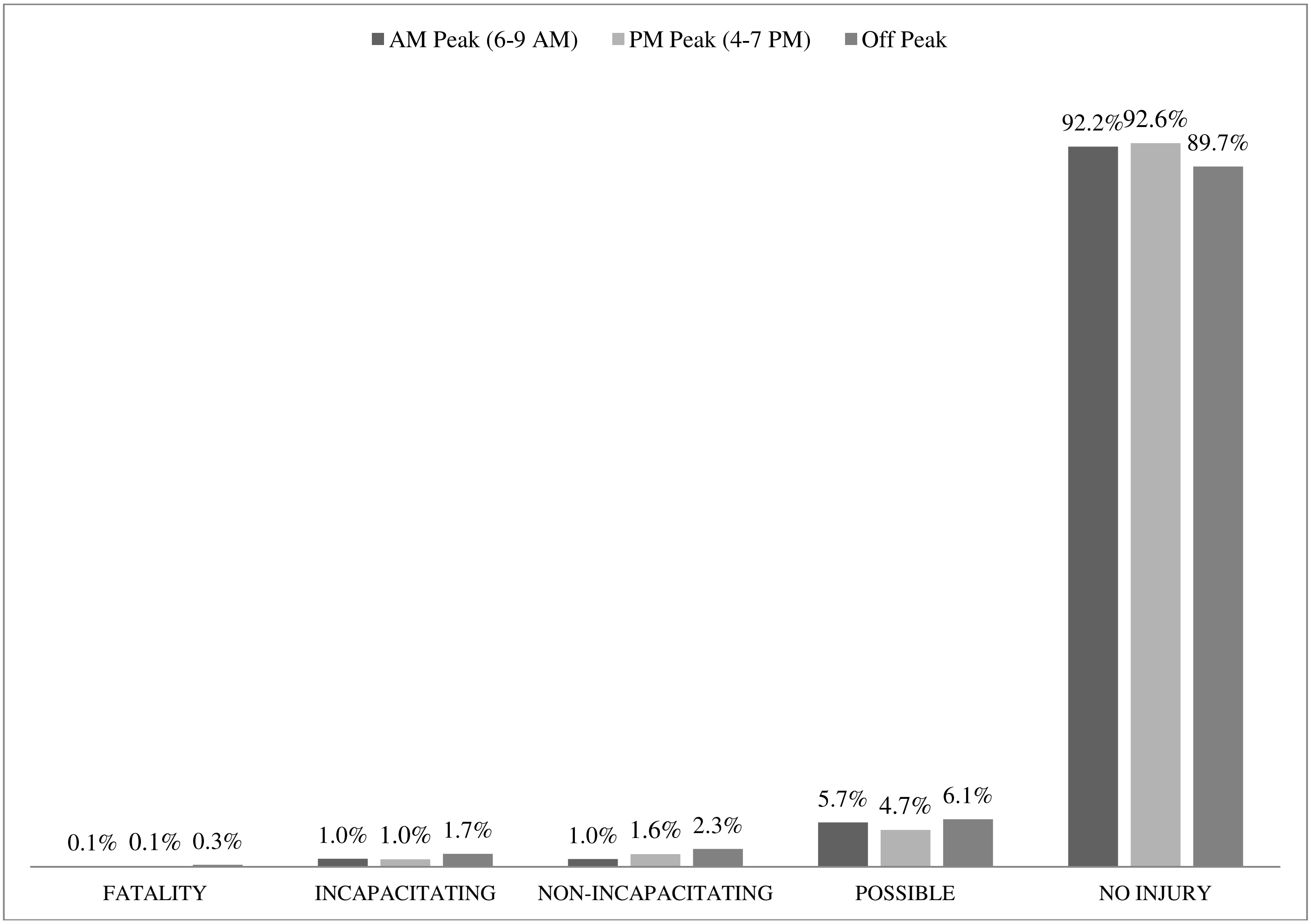

Figure 7.1: Injury Level Break-down for AM, PM, and Off Peak Periods

The descriptive statistics of three samples representing three different time periods of day correspond to their respective crash involvement with in this study. The descriptive statistics are shown in Table 7.1.

Table 7.1: Descriptive Statistics of key Variables in the three Time Periods MNL Models

	AM Peak (6 ~ 9 am) N = 2,409		PM Peak (4 ~ 7 pm) N = 2,502		Off Peak N = 6,649	
Meaning of Variable	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Fatal outcome						
Percentage of truck traffic (% of total traffic flow)	10.809	4.839				
Annual average daily traffic per lane (veh/day/lane)	20541.4	7701.84				
Weather condition at the time of crash location (1 if clear sky, 0 otherwise)					0.686	0.464
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)					0.151	0.358
Time of day (1 if between 12 am to 6 am, 0 otherwise)					0.168	0.374
Percentage of combination truck traffic (% of total traffic flow)			7.557	4.410		
Incapacitating Injury Outcome						
Annual average daily traffic per lane (veh/day/lane)	20541.4	7701.84				
Surface condition at the time of crash (1 if wet surface, 0 otherwise)	0.128	0.335				
Left shoulder width (feet)	15.089	7.388				
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	0.112	0.315	0.079	0.269	0.151	0.358
Percentage of truck traffic (% of total traffic flow)					11.041	5.108
Weather condition at time of crash location (1 if rainy, 0 otherwise)					0.111	0.314
Right shoulder width (feet)					19.674	3.806
Month of the year (1 if summer month, 0 otherwise)			0.273	0.445		
Year of crashes (1 if 2010, 0 otherwise)			0.239	0.426		
Type of vehicle involved at the time of crash (1 if truck, 0 otherwise)			0.109	0.312		
Non-Incapacitating Injury Outcome						
Number of directional lane (1 if 4 lane in both direction)	0.145	0.352				
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	0.112	0.316	0.079	0.269	0.151	0.358
Left shoulder width (feet)	15.089	7.388				
Gender of drivers (1 if male, 0 otherwise)	0.947	0.223				
Surface condition at the time of crash (1 if level surface, 0 otherwise)					0.767	0.422
Number of directional lane (1 if 6 lanes in both directions, 0 otherwise)			0.384	0.486	0.396	0.489
Right shoulder width (feet)					19.674	3.806
Surface condition at the time of crash (1 if surface is dry, 0 otherwise)			0.862	0.345		
Days of week (1 if weekdays, 0 otherwise)			0.885	0.319		
Possible Injury Outcome						
Weather condition at the time of crash location (1 if clear sky, 0 otherwise)	0.698	0.458				
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	0.112	0.316	0.079	0.269	0.151	0.358
Number of vehicle involved in crashes (1 if more than one vehicle involved, 0 otherwise)	0.333	0.471	0.369	0.482		
Gender of drivers (1 if male, 0 otherwise)			0.936	0.245	0.940	0.236
Age group (1 if age between 25 to 35, 0 otherwise)					0.176	0.381
Number of vehicles involved (1 if single vehicle involved, 0 otherwise)					0.636	0.481
Surface condition at the time of crash (1 if level surface, 0 otherwise)					0.767	0.422

	AM Peak (6 ~ 9 am) N = 2,409		PM Peak (4 ~ 7 pm) N = 2,502		Off Peak N = 6,649	
Meaning of Variable	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Surface condition at the time of crash (1 if surface is dry, 0 otherwise)			0.862	0.345		
Non-Injury Outcome (Property-Damage-Only)						
Weather condition at the time of crash location (1 if clear sky, 0 otherwise)	0.698	0.458			0.686	0.464
Age group (1 if age between 45 to 55, 0 otherwise)	0.224	0.417				
Race of drivers (1 if white, 0 otherwise)	0.466	0.498	0.488	0.499		
Year of crashes (1 if 2010, 0 otherwise)	0.217	0.412	0.239	0.426		
Age group (1 if age between 55 to 65, 0 otherwise)					0.117	0.322
Year of crashes (1 if 2009, 0 otherwise)					0.167	0.373
Logarithm of AADT per lane (log of veh/day/ln)					9.853	0.419
Light condition at the time of crash (1 if day light, 0 otherwise)					0.682	0.465
Left shoulder width (feet)			15.211	7.189		

7.3.1 AM peak period sample

Considering the demographics of drivers, age group of 45 to 55 years and white drivers account for 22.4% and 46.6% of the sample size, respectively. Turning to driving behavior, lane keeping accounts for 11.2% of the sample size. Now considering the geometry of the highway, average left shoulder width is 15.1 feet. Moreover, four lanes in both direction account for 14.5% of the sample size. Considering traffic flow in the AM periods, truck traffic is 10.8 percent of total traffic flow and average traffic flow is 20,541 vehicles per lane per day. Turning to weather condition during the AM peak period, clear sky conditions account for 69.8% of the sample size. Considering the number of vehicles involved in the crashes, involvement of multi-vehicles accounts for 33.3 percent of the sample size during the AM peak. AM peak time period during the year 2010 accounts for 21.7% of the sample size.

7.3.2 PM peak period sample

Considering the demographics of drivers, male drivers and white drivers account for 93.6% and 48.8% of the sample size, respectively. Turning to driving behavior, lane keeping accounts for 7.9% of the sample size. Now considering the geometry of the highway, average right shoulder width is 15.2

feet. Moreover, six lanes in both direction account for 38.4% of the sample size. Considering traffic flow in the PM periods, truck traffic is 7.6 percent of total traffic flow. Turning to road-environment during the PM peak period, dry surface accounts for 86.2% of the sample size. Days of week particularly Monday to Friday accounts for 88.5% of the sample size. Considering the number of vehicles involved in the crashes, involvement of multi-vehicles accounts for 36.9 percent of the sample size during the PM peak. PM peak time periods of year 2010 account for 23.9% of the sample size. Summer months (June to August) accounts for 27.3% of the sample size.

7.3.3 Off peak period sample

Considering the demographics of drivers, male drivers and age group from 25 to 35 years and 55 to 65 years drivers account for 94.0%, 17.6%, and 11.7% of the sample size, respectively. Turning to driving behavior, lane keeping accounts for 15.1% of the sample size. Now considering the geometry of the highway, average right shoulder width is 19.7 feet. Moreover, six lanes in both direction and level surface accounts for 39.6% and 76.7% of the sample size, respectively. Considering traffic flow in the off peak periods, logarithm of average traffic flow is 9.8 vehicles per day per lane. Additionally, average truck flow is 11 percent of total traffic flow during off peak periods. Considering weather condition during off peak period, rainy and clear sky condition accounts for 11.1% and 68.6% of the sample, respectively. Time period of the day such as 12 to 6 am and day light condition accounts for 16.8% and 68.2% of the sample size, respectively. Considering the number of vehicles involved in the crashes, involvement of single-vehicles accounts for 63.6 percent of the sample size during the off peak. Off peak period during the year 2009 accounts for 16.7% of the sample size.

7.4 Empirical results and discussions

Maximum likelihood and simulation-based maximum likelihood methods are utilized to estimate parameter vector $\boldsymbol{\beta}_{\mathbf{i}}$ for fixed- and random-parameters logit models, respectively. We considered normal, lognormal, triangular, and uniform distributions for the distribution of the random parameters in our analysis. However, the normal distribution was found to be statistically significant. With regard to random parameters estimation, 200 Halton draws has been empirically shown to produce accurate parameter estimates which was used for the simulation-based maximum likelihood estimation (Milton et al., 2008; Train, 1999).

The discussion of the models results are described in the following subsections along with the elasticities (both in terms of direct pseudo and average direct elasticity) as presented in Table 7.5 and Table 7.6, Table 7.7, respectively.

In estimating mixed logit for three different time periods, we found three variables random and normally distributed. Indicator variable male (specific to possible injury outcome), constant specific to non-incapacitating injury outcome and indicator variable for day light (specific to no injury outcome) were found to be random in AM peak, PM peak, and off peak period models, respectively.

The constant term for non-incapacitating injury outcome is found to be random and normally distributed in PM peak mixed logit model with mean of -5.536 and standard deviation of 2.335. Given these values, this constant is less than zero for 99.11% for large-truck involved crashes which result in non-incapacitating injuries. That is, about 99% of large-truck involved crashes on average were less likely to result in non-incapacitating injury outcomes.

Table 7.2: Mixed Logit Model for AM Peak Time (6 – 9 am) for Urban Texas Areas

Meaning of Variable	Estimate	t-stat	*P*-Value
Fatal outcome			
Constant	-14.631	-3.178	0.002
Percentage of daily truck traffic (% of total traffic flow)	0.334	2.172	0.029
Annual average daily traffic per lane (veh/day/lane)	0.157x10^(-03)	1.755	0.079
Incapacitating Injury Outcome			
Constant	-2.335	-2.717	0.007
Annual average daily traffic per lane (veh/day/lane)	-0.469x10^(-04)	-1.682	0.093
Surface condition at the time of crash (1 if wet surface, 0 otherwise)	-1.309	-1.683	0.092
Left shoulder width (feet)	-0.049	-1.682	0.093
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	1.297	2.748	0.006
Non-incapacitating Injury Outcome			
Constant	-3.771	-4.954	0.000
Number of directional lane (1 if 4 lanes in both directions, 0 otherwise)	1.447	2.899	0.004
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	1.676	3.873	0.000
Left shoulder width (feet)	0.055	2.117	0.034
Gender of drivers (1 if male, 0 otherwise)	-1.932	-3.551	0.000
Possible Injury Outcome			
Constant	-1.918	-3.983	0.000
Gender of drivers (1 if male, 0 otherwise) (standard error of parameter distribution)	-3.511 (1.862)	-3.182 (2.082)	0.002
Weather condition at time of crash location (1 if clear sky, 0 otherwise)	1.847	4.752	0.000
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	1.812	4.515	0.000
Number of vehicle involved in crashes (1 if more than one vehicle involved, 0 otherwise)	0.843	3.192	0.001
Non-Injury Outcome (Property-Damage-Only)			
Weather condition at time of crash location (1 if clear sky, 0 otherwise)	1.022	2.336	0.019
Age group (1 if age between 45 to 55, 0 otherwise)	0.526	2.059	0.039
Race of drivers (1 if white, 0 otherwise)	0.507	2.716	0.007
Year of crashes (1 if 2010, 0 otherwise)	-0.695	-1.831	0.067
Number of observations		2,409	
Restricted log-likelihood		-3877.136	
Log-likelihood at convergence		-708.848	
Chi-squared value		6336.577	
McFadden pseudo R-squared (ρ^2)		0.817	

Table 7.3: Mixed Logit Model for PM Peak Time (4 – 7 pm) for Urban Texas Areas

Meaning of Variable	Estimate	t-stat	*P*-Value
Fatal outcome			
Constant	-8.745	-5.378	0.000
Percentage of combination truck traffic (% of total traffic flow)	0.199	1.933	0.053
Incapacitating Injury Outcome			
Constant	-7.568	-3.750	0.000
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	1.288	1.642	0.100
Month of the yea (1 if summer month, 0 otherwise)	-1.788	-1.964	0.049
Year of crashes (1 if 2010, 0 otherwise)	2.671	2.962	0.003
Type of vehicle involved at the time of crash (1 if truck, 0 otherwise)	-1.347	-1.694	0.090
Non-incapacitating Injury Outcome			
Constant (standard error of the parameter)	-5.536 (2.335)	-5.040 (2.370)	0.000
Number of directional lane (1 if 6 lanes in both directions, 0 otherwise)	0.751	2.342	0.019
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	1.397	3.396	0.000
Surface condition at the time of crash (1 if surface is dry, 0 otherwise)	2.010	1.969	0.048
Days of week (1 if weekdays, 0 otherwise)	-0.659	-1.695	0.090
Possible Injury Outcome			
Constant	-0.621	-1.731	0.083
Gender of the drivers (1 if male, 0 otherwise)	-2.527	-10.857	0.000
Surface condition at the time of crash (1 if surface is dry, 0 otherwise)	-0.531	-2.169	0.030
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	2.025	8.029	0.000
Number of vehicles involved in the crashes (1 if more than one vehicle, 0 otherwise)	0.491	2.254	0.024
Non-Injury Outcome (Property-Damage-Only)			
Left shoulder width (feet)	0.020	1.768	0.077
Race of drivers (1 if white, 0 otherwise)	0.434	2.538	0.011
Year of crashes (1 if 2010, 0 otherwise)	-0.548	-2.972	0.003
Number of observations		2,502	
Restricted log-likelihood		-4026.814	
Log-likelihood at convergence		-710.207	
Chi-squared value		6633.212	
McFadden pseudo R-squared (ρ^2)		0.824	

Table 7.4: Mixed Logit Model for Off peak Time (other than AM and PM peak) for Urban Texas

Meaning of Variable	Estimate	t-stat	*P*-Value
Fatal outcome			
Constant	-10.096	-7.242	0.000
Weather condition at time of crash location (1 if clear sky, 0 otherwise)	2.301	3.017	0.003
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	2.667	5.157	0.000
Time of day (1 if between 12 am to 6 am, 0 otherwise)	0.993	2.012	0.044
Incapacitating Injury Outcome			
Constant	-6.773	-5.313	0.000
Percentage of truck traffic (% of total traffic flow)	0.042	2.457	0.014
Right shoulder width (feet)	0.047	1.878	0.060
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	0.912	3.873	0.000
Weather condition at time of crash location (1 if rainy, 0 otherwise)	-0.629	-1.903	0.057
Non-incapacitating Injury Outcome			
Constant	-5.776	-4.743	0.000
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	1.649	8.892	0.000
Surface condition at the time of crash (1 if level surface, 0 otherwise)	-0.801	-4.635	0.000
Number of directional lane (1 if 6 lanes in both directions, 0 otherwise)	0.403	2.366	0.018
Right shoulder width (feet)	0.037	1.627	0.104
Possible Injury Outcome			
Constant	-2.077	-1.856	0.063
Gender of drivers (1 if male, 0 otherwise)	-1.827	-11.908	0.000
Age group (1 if age between 25 to 35, 0 otherwise)	-0.338	-2.135	0.032
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	1.209	8.342	0.000
Number of vehicles involved (1 if single vehicle involved, 0 otherwise)	-0.398	-3.413	0.001
Surface condition at the time of crash (1 if level surface, 0 otherwise)	-0.332	-2.724	0.007
Non-Injury Outcome (Property-Damage-Only)			
Weather condition at time of crash location (1 if clear sky, 0 otherwise)	0.538	5.591	0.000
Age group (1 if age between 55 to 65, 0 otherwise)	0.386	2.482	0.013
Year of crashes (1 if 2009, 0 otherwise)	0.270	2.035	0.042
Logarithm of AADT per lane (log of veh/day/ln)	-0.181	-1.635	0.102
Light condition at the time of crash (1 if day light, 0 otherwise) (standard error of parameter distribution)	0.456 (0.906)	1.690 (2.027)	0.090
Number of observations		6,649	
Restricted log-likelihood		-10701.15	
Log-likelihood at convergence		-2676.073	
Chi-squared value		16050.16	
McFadden pseudo R-squared (ρ^2)		0.749	

Table 7.5: Average Pseudo Elasticities for AM peak Time (6-9 am) for Urban Texas

Variables	**Elasticity (%)**				
	PDO/ No Injury	**Possible Injury**	**Non-incapacitating Injury**	**Incapacitating Injury**	**Fatal**
Percentage of daily truck traffic	-0.49	-0.38	-1.16	-0.88	88.2
Annual average daily traffic per lane (veh/day/lane)	-0.30	-0.22	-0.55	-0.41	56.7
Annual average daily traffic per lane (veh/day/lane)	0.88	0.88	1.75	-85.07	1.51
Surface condition at the time of crash (1 if wet surface, 0 otherwise)	0.10	0.12	0.43	-10.37	0.37
Left shoulder width (feet)	0.66	0.67	1.58	-64.34	1.12
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	-0.32	-0.67	-1.82	34.14	-0.68
Number of directional lane (1 if 4 lanes in both directions, 0 otherwise)	-0.36	-0.80	39.07	-0.86	1.55
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	-0.54	-1.95	63.59	-2.26	-1.76
Left shoulder width (feet)	-0.84	-1.68	89.06	-1.66	-1.53
Gender of drivers (1 if male, 0 otherwise)	1.47	1. 34	-147.88	3.33	3.13
Gender of drivers (1 if male, 0 otherwise)	4.38	-73.47	6.36	6.71	44.42
Weather condition at time of crash location (1 if clear sky, 0 otherwise)	-5.62	93.16	-6.31	-4.16	-2.84
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	-1.63	29.59	-12.05	-5.12	-2.39
Number of vehicle involved in crashes(1 if more than one vehicle involved, 0 otherwise)	-1.51	25.42	-3.20	-1.71	-1.19
Weather condition at time of crash location (1 if clear sky, 0 otherwise)	3.90	-50.54	-31.27	-37.19	35.88
Age group (1 if age between 45 to 55, 0 otherwise)	0.52	-5.93	-6.48	-7.27	7.28
Race of drivers (1 if white, 0 otherwise)	1.22	-14.01	-15.48	-15.69	14.89
Year of crashes (1 if 2010, 0 otherwise)	-1.32	10.38	26.79	32.31	0.16

Table 7.6: Average Pseudo Elasticities for PM Peak Time (4-7 pm) for Urban Texas

Variables	Elasticity (%)				
	PDO/No Injury	**Possible Injury**	**Non-incapacitating Injury**	**Incapacitating Injury**	**Fatal**
Percentage of combination truck traffic (% of total traffic flow)	-0.22	-0.22	-0.22	-0.17	279.6
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	-0.12	-0.54	-0.53	15.22	-0.14
Month of the yea (1 if summer month, 0 otherwise)	0.12	0.19	0.17	-13.27	0.10
Year of crashes (1 if 2010, 0 otherwise)	-1.25	-1.94	-2.28	136.78	-1.45
Type of vehicle involved at the time of crash (1 if truck, 0 otherwise)	0.16	0.30	0.26	-17.73	0.18
Number of directional lane (1 if 6 lanes in both directions, 0 otherwise)	-0.67	-0.77	40.40	-0.81	-0.79
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	-0.37	-1.34	25.24	-1.01	-0.42
Surface condition at the time of crash (1 if surface is dry, 0 otherwise)	-3.13	-3.64	189.82	-3.74	-3.30
Days of week (1 if weekdays, 0 otherwise)	0.83	0.96	-50.18	0.94	0.90
Gender of the drivers (1 if male, 0 otherwise)	7.38	-150.65	10.76	9.79	8.23
Surface condition at the time of crash (1 if surface is dry, 0 otherwise)	1.64	-33.60	2.77	1.94	1.68
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	-2.02	42.49	-5.59	-4.29	-2.54
Number of vehicles involved in the crashes (1 if more than one vehicle, 0 otherwise)	-0.80	16.14	-0.86	-0.84	-0.78
Left shoulder width (feet)	1.93	-23.89	-27.08	-20.35	-27.83
Race of drivers (1 if white, 0 otherwise)	1.16	-14.67	-15.36	-12.26	-15.77
Year of crashes (1 if 2010, 0 otherwise)	-1.30	14.73	15.79	25.23	11.46

Table 7.7: Average Pseudo Elasticities for Off peak Time (other than AM and PM peak) for Urban Texas

Variables	**Elasticity (%)**				
	PDO/No Injury	**Possible Injury**	**Non-incapacitating Injury**	**Incapacitating Injury**	**Fatal**
Weather condition at time of crash location (1 if clear sky, 0 otherwise)	-0.50	-0.87	-1.20	-0.97	200.6
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	-0.40	-0.98	-1.52	-0.99	173.4
Time of day (1 if between 12 am to 6 am, 0 otherwise)	-0.10	-0.20	-0.30	-0.20	42.89
Percentage of truck traffic (% of total traffic flow)	-0.79	-1.25	-1.46	49.23	-1.61
Right shoulder width (feet)	-1.48	-2.38	-2.77	92.60	-2.81
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	-0.31	-0.96	-1.51	22.88	-2.09
Weather condition at time of crash location (1 if rainy, 0 otherwise)	0.09	0.22	0.33	-6.11	0.08
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	-1.41	-4.71	70.21	-3.81	-8.07
Surface condition at the time of crash (1 if level surface, 0 otherwise)	0.92	1.63	-41.26	1.80	2.31
Number of directional lane (1 if 6 lanes in both directions, 0 otherwise)	-0.40	-0.88	18.47	-0.78	-1.09
Right shoulder width (feet)	-1.51	-3.25	69.32	-3.07	-4.10
Gender of drivers (1 if male, 0 otherwise)	7.88	-126.46	17.88	13.88	15.82
Age group (1 if age between 25 to 35, 0 otherwise)	0.25	-4.00	0.57	0.46	0.51
Driving maneuver at the time of crash (1 if going straight – lane keeping, 0 otherwise)	-1.74	30.60	-9.08	-4.70	-10.08
Number of vehicles involved (1 if single vehicle involved, 0 otherwise)	1.23	-19.93	3.24	2.30	3.02
Surface condition at the time of crash (1 if level surface, 0 otherwise)	1.25	-19.64	1.78	2.14	2.30
Weather condition at time of crash location (1 if clear sky, 0 otherwise)	2.79	-22.97	-22.03	-28.17	-38.54
Age group (1 if age between 55 to 65, 0 otherwise)	0.31	-2.66	-2.71	-2.86	-2.82
Year of crashes (1 if 2009, 0 otherwise)	0.32	-2.65	-2.81	-2.88	-3.19
Logarithm of AADT per lane (log of veh/day/ln)	-16.36	138.97	140.17	148.91	141.66
Light condition at the time of crash (1 if day light, 0 otherwise)	-0.47	3.47	3.46	6.93	2.88

7.4.1 Drivers' demographics

Male drivers during AM peak period are 147.9% and 73.5% less likely to be involved in non-incapacitating and possible injuries, respectively. Also, this parameter was found to be random and normally distributed with a mean of -3.511 and a standard deviation of 1.862 for possible injuries. Given these estimates, this parameter is below zero for 97.0% of large-truck involved crashes. This suggests that about 97.0% of large-truck involved crashes involving male drivers during AM peak period on average were less likely to result in possible injuries. Additionally, male drivers are 150.6% and 126.5% less likely to be involved in possible injuries during PM peak period and off peak period, respectively. Chen and Chen (2011) found a similar result as females are more likely to be involved in fatal or incapacitating injuries in single and multi-vehicle collisions. Additionally, this evidence is also supported by Abdel-Aty and Abdelwahab (2004). Turning to age groups, age group of 45 to 55 years during AM peak period is more likely to be involved in no injury crashes. A supporting study by Stamatiadis and Deacon (1995) indicated that age group of 25 to 35 years were on average more likely to be involved in the crashes compared to age group of 45 to 55 years (25-35 years having relative accident involvement ratio of 1.672 compared to 1.386 for 45-55 years age group, averaging over 1978 to 1988 years crash data). Younger drivers perform relatively poorly because of limited driving experience, risk-taking behaviors, and attitudinal factors particularly at night which lead to a significant finding that middle age drivers are safer than younger group (Stamatiadis and Deacon, 1995). Age group of 25 to 35 years is less likely to be involved in possible injuries during off peak period. However, age group of 55 to 65 years is more likely to be involved in no injury crashes during off peak period. Considering the race of the drivers, white drivers during AM and PM peak period are more likely to be involved in no injury crashes and less likely to be involved in severe injury crashes.

7.4.2 Driving behavior

Lane keeping or driving straight during AM peak period increases the likelihood of incapacitating, non-incapacitating, and possible injuries by 34.1%, 63.6%, and 29.6%, respectively. Similarly, lane keeping or driving straight during PM peak period increases the likelihood of incapacitating, non-incapacitating, and possible injuries by 15.2%, 25.2%, and 42.5%, respectively. Additionally, lane keeping or driving straight during off peak period increases the likelihood of fatality, incapacitating, non-incapacitating, and possible injuries by 173.4%, 22.9%, 70.2%, and 30.6% respectively. This driving behavior indirectly indicates the rear-end collisions for the passenger vehicles. This driving behavior indirectly indicates the rear-end collisions for the passenger vehicles.

7.4.3 Roadway geometrics

Wider left shoulder during AM peak periods decreases incapacitating injuries by 64.3%, but increases non-incapacitating injuries by 89.1%. Likewise, wider left should during PM peak periods decreases fatalities, incapacitating, non-incapacitating, and possible injuries by 27.8%, 20.4%, 27.1%, and 23.9%, respectively. Previous study supports the fact that increase in shoulder width (up to 7.5 meters) decreases the crash rates (Choueiri et al., 1994, in Ben-Bassat and Shinar, 2011). On the contrary, increase in shoulder width can also trigger of crashes and increases the likelihood severe injuries. Wider right shoulder during off peak periods increases incapacitating and non-incapacitating injuries by 92.6% and 69.3%, respectively. A previous study indicated that approximately ten percent of fatal freeway crashes are related to vehicles stopped on shoulders (Hauer and Lovelly, 1984, in Hauer, 2000).

Number of directional lanes also influences driving behavior and impacts safety in different time of the day. Four lanes during AM peak period, six lanes during PM peak period and six lanes during off peak periods increase the likelihood of non-incapacitating injuries by 39.1%, 40.4%, and 18.5%,

respectively. AM and PM peak hour clearly indicates higher traffic (average daily traffic – ADT) flow along the freeways. A previous study by Lundy (1965) supported the results indicating that the four lane freeways have higher accident rates than the six lane; whereas, six lane freeways have higher rate than the eight lane. With the increase of traffic flow, the difference in accident rates increases. Level roadway surface during off peak period decreases non-incapacitating and possible injuries by 41.3% and 19.6%, respectively.

7.4.4 Traffic characteristics

Percentage of daily truck traffic during AM peak periods increases the likelihood of fatalities by 588.2%. Similarly, percentage of combination truck during PM peak period increases the likelihood of fatalities by 279.6%. Additionally, percentage of daily truck traffic during off peak period increases the likelihood of incapacitating injuries by 47.2%.

Traffic flow also has impact on the severities. Average annual daily traffic (AADT) per lane during AM peak also increases the likelihood of fatalities by 356.7%. However, AADT per lane during AM peak period decreases the likelihood of incapacitating injuries by 85.1%. Also, annual average daily traffic (AADT) per lane (taking the logarithm of AADT per lane) during off peak period increase fatality, incapacitating, non-incapacitating, and possible injuries by 141.7%, 148.9%, 140.2%, and 138.9%, respectively. A France study by Martin (2002) found that given heavier traffic flows, crash rates increase steadily as traffic increases on 2- and 3-lane motorways. This incidence highly affects 2-lane motorways when traffic increases to a level of 3000 vehicles per hour.

7.4.5 Temporal characteristics

AM peak periods during the year 2010 increases fatality, incapacitating, non-incapacitating, and possible injuries by 30.2%, 32.3%, 26.8%, and 10.4%, respectively. Weekdays during PM peak period

decreases the likelihood of non-incapacitating injuries by 50.2%. This capture the slow moving traffic (e.g., congested condition). Months from June to August during PM peak period decreases the likelihood of incapacitating injuries by 13.3%. Also, PM peak period during the year 2010 increases the likelihood of incapacitating injuries by 136.8%. Also, this time period decreases likelihood of no injuries.

Off peak periods such as 12 to 6 am increases the likelihood of fatalities by 42.9%. This severity outcome is very likely to be associated with single-vehicle collisions. Also, off peak periods during the year 2009 increases the likelihood of no injury crashes.

Time of day is highly associated with the traffic volume and types of crash occurrence (e.g., single and multi-vehicle collisions). A detailed study by Qin et al. (2006) found that a negative relationship (e.g., convex downward) between single-vehicle collisions and hourly traffic volume at some duration of day time and concave upward during other times. However, this is not the case for multivehicle collisions where either concave or convex relationship exists between multi-vehicle collisions and hourly volume. Overall, this study found that there is "U" shape relationship for all vehicle collisions with hourly volume given convex downward for single and convex upward for multi-vehicle collisions.

7.4.6 Weather characteristics

Wet surface during AM peak period decreases the likelihood of incapacitating injuries by 10.4%. In a study by Morgan and Mannering (2011) found that female of less than 45 older are less likely to be involved in the minor injuries on wet surfaces during AM peak hours. On the other hand, the clear sky condition during AM peak periods increases the likelihood of possible injuries by 93.2% and no injury crashes but reduces the likelihood of severe injuries. A similar finding was indicated in a study by Morgan and Mannering (2011) that male of 45 and older are less likely to be involved in minor injuries in clear weather. Dry surface during PM peak period increases the likelihood of incapacitating injuries

by 189.8% but decreases the likelihood of possible injuries by 33.6%. A study by Morgan and Mannering (2011) found that male drivers less than 45 years old are less likely to be involved with minor injuries on dry surface during AM or PM peak periods. Clear sky condition during off peak periods increases fatalities by 200.1%. However, this condition during the off peak decreases the likelihood of incapacitating, non-incapacitating, possible injuries.

Day light condition during the off peak period was found to be random and normally distributed with a mean of 0.456 and a standard deviation of 0.906 for no injury crashes. Given these estimates, this parameter is above zero for 30.7% of large-truck involved crashes during off peak period. This suggests that about 30.7% of large-truck involved crashes during day light time during off peak period on average were more likely to result in no injury crashes. A similar result was found in a study by Morgan and Mannering (2011) that male drivers of 45 and more than 45 years are more likely to be involved in no injuries on dry surface.

7.4.5 Crash dynamics

Number of vehicles involved in the crashes, depending on the traffic flow, highly influences the outcomes of the injuries. The model result showed that multiple vehicles involved in the crashes during AM and PM peak period increases the likelihood of possible injuries by 25.4% and 16.1%, respectively. Single vehicle crashes during off peak period decreases possible injuries by 19.9%. Previous studies indicated that a convex relationship between single-vehicle accidents and traffic flow, but a concave relationship for the multi-vehicle accident (Persaud and Mucsi, 1995). Ivan et al. (2000) also found a nonlinear relationship between single-vehicle crashes and the hourly volume to capacity ratio on two-lane rural road segments (Ivan et al.,2000).

7.5 Model specification test

The likelihood ratio test was conducted to justify the statistical significance of rural and urban injury model separately from the combined model. This is because other studies (Morgan and Mannering, 2011; Geedipally and Lord, 2010) have shown that separate models are statistically significant to capture more factors than a combined model. The null hypothesis is set as the separate models are not statistically and significantly different from the combined model. The following likelihood ratio test was conducted to test the hypothesis (Washington et al., 2011):

$$\chi 2 = -2 * [LL_U(\beta^U) - LL_{AM}(\beta^{AM}) - LL_{PM}(\beta^{PM}) - LL_{PM}(\beta^{PM})] \quad (7.7)$$

where,

$$LL_U(\beta^U)\ is\ loglikelihood\ at\ convergence\ of\ the\ combined\ model$$
$$(-18605.10,\ degree\ of\ freedom, n_U = 27)$$
$$LL_{AM}(\beta^{AM})\ is\ loglikelihood\ at\ convergence\ of\ AM - peak\ model$$
$$(-708.848,\ degree\ of\ freedon, n_{AM} = 23)$$
$$LL_{PM}(\beta^{PM})\ is\ loglikelihood\ at\ convergence\ of\ PM - peak\ model$$
$$(-710.207,\ degree\ of\ freedom, n_{PM} = 20)$$
$$LL_{OFF}(\beta^{OFF})\ is\ loglikelihood\ at\ convergence\ of\ OFF - peak\ model$$
$$(-2676.073,\ degree\ of\ freedom, n_{OFF} = 25)$$

The test follows χ^2 distribution with degrees of freedom equal to the sum of AM, PM and Off peak model minus that of total model (Urban model). With 41 degrees of freedom and χ^2 value of 29019.944, the test statistic indicates that we reject the null hypothesis. This leads to the fact that separate models (i.e., AM, PM, and off peak models) are statistically and significantly different from the combined model with *P*-value being less than 0.0001 (one-tailed *P*-value). So, the test result indicates

that significance of injury severity should be modeled as AM, PM, and off peak models separately rather than combined one.

The estimated variables in both models were found to be statistically significant within a 95% confidence level. A likelihood ratio test comparing the multinomial logit (i.e., fixed parameter model) and mixed logit (i.e., random parameter model) was performed to test the null hypothesis that the fixed parameter model is statistically equivalent to the random parameters model and the procedure is as follows (Washington et al., 2011):

$$\chi^2 = -2[LL_{MNL}(\beta^{MNL}) - LL_{ML}(\beta^{ML})] \quad (7.8)$$

where:

$LL_{MNL}(\beta^{MNL})$:is the log-likelihood at convergence of the fixed parameters model

$LL_{ML}(\beta^{ML})$:is the log-likelihood at convergence of the random parameters model

The Chi-square statistic for the likelihood ratio test with one degree of freedom gave value greater than the 80.0% ($\chi^2 = 1.716$), 87.0% ($\chi^2 = 2.308$), and 77.0% ($\chi^2 = 1.458$) confidence interval for AM peak, PM peak and Off Peak model, indicating that the mixed logit model (i.e., random parameter model) is statistically superior to the corresponding multinomial model (i.e., fixed parameter model). This means that the null hypothesis that the random parameters estimated model is no better than the fixed model comparison model is rejected. However, this log likelihood test does not clearly shows a strong statistical significance in favor of mixed logit model compared to the base model (i.e., multinomial logit model).

7.6 Conclusion and future research

In this study, we utilized Texas police reported crash data commonly known as CRIS over a period of five years from 2006 to 2010 for urban areas (population greater than 200,000) in Texas. The severity levels considered in this analysis includes fatal, incapacitating, non-incapacitating, possible injury and no injury (Property Damage Only). Separate mixed logit models were estimated for AM peak (6 to 9 am), PM peak (4 to 7 pm) and off peak (other than AM and PM peak). The model results clearly indicated the difference in driving patterns as well as crashes in different time periods of day. As traffic flow is highly dependent on the movement of people for various activities throughout the day, the likelihood of crashes varies accordingly as well as associated injury severities. Driving alertness varies over the span of day and complex traffic interactions results in near misses to severe crashes. Overall the model results indicate there is need to research more in-depth on the traffic interaction and severity levels of large-truck involved crashes in urban areas in Texas. This is evidenced from the wide variation of sign and magnitude of individual parameter estimates in these models. The contributing factors obtained from model estimation were discussed under drivers' demographics, driving behavior, roadway geometrics, traffic characteristics, weather characteristics, temporal characteristics, and crash dynamics. Driving patterns as reflected through these factors widely vary depending on the time of day and so do level of injury severities in large truck crashes in urban areas.

Although the modeling approach such as random parameter logit model helps to overcome the assumptions regarding limitation of information in the data reported from the investigating police officers, there is still need of detailed information on driver's reaction or evasive maneuvers, visibility, familiarity of road network, and so on. The detailed information on these contributing factors will provide greater insights into the problem in urban areas and this research framework will lead with a wider avenue for future research. This study provides a basis of analyzing different subsets of crash database in the modeling framework of injury analysis which has been sparsely explored for large-truck

involved. In future, we plan to do more detailed modeling utilizing rural areas in Texas based on time of day.

Past research studies (Qin et al. 2006; Geedipally and Lord, 2010; Morgan and Mannering, 2011) provides a strong basis of splitting the dataset, based on manner of collisions (single and multi-vehicle collisions, etc.) and time of day, age/gender/road surface condition, and different model estimation procedures to unveil the complexity hidden inside the data captured through the log likelihood ratio tests and differences of estimated model parameters (Morgan and Mannering, 2011). These approaches clearly enhance our understanding level about the individual parameters and the contributing factors in crashes. This improved understanding will lead to have a strong knowledgebase to use as tool for countermeasures for the safety professionals, policy makers, truck operating managers, and overall general public.

Chapter 8: The Spatial and Temporal Transferability of Severity Outcome Models: An Application to Texas Crash Data

8.1 Introduction

This chapter focuses on the estimation of severity models for large truck crashes and evaluation of the spatial and temporal transferability of severity models by demonstrating a conceptual framework utilizing Texas crash database (CRIS). Other studies have investigated the temporal transferability of severity models by focusing on transferring and updating of the accident prediction models (Hadayeghi et al., 2006), the evacuation system for hurricanes (Hasan et al., 2012), and mode choice behavior (Habib et al., 2012) in the context of medium and long range planning. We set two objectives in this study. First, we develop severity models from a unified crash database (CRIS) with reasonable data and second, we develop a conceptual framework on spatial (between two major cities in Texas) and temporal (between two time periods) transferability and its evaluation process. The following sections describe the methodology.

8.2 Methodology

Many research studies utilized a number of methodological approaches such as the multinomial logit, ordered probit, Bayesian Ordered Probit, nested logit, and mixed logit models. Each of these models have been utilized to model injury severities (Chistoforou et al., 2010; Lemp et al., 2011; Milton et al., 2008; Zhu and Srinivasan, 2011; Duncan et al., 1998; Xie et al., 2009; Khorashadi et al., 2005; Chang and Mannering, 1999). In this study, we estimate a mixed logit model for injury severity of crashes involving large trucks by considering random parameters in the developed model following the similar logical research framework as of past research studies (Milton et al., 2008; Gkritza and Mannering, 2008; Chen and Chen, 2011). The level of injury is discrete in nature as coded in the injury

scale KABCO (i.e., ‘K’ for Fatal, ‘A’ for Incapacitating injury, ‘B’ for Non-incapacitating Injury, ‘C’ for Possible Injury and ‘O’ for Property-Damage-Only), and a mixed logit model has been widely accepted to model the effects of several contributing factors on the levels of injury severity. Several research studies conducted by Revelt and Train (1997, 1999), Train (1997), Brownstone and Train (1999), McFadden and Train (2000), and Bhat (2001) have clearly pointed to the effectiveness of this methodological approach. Although discrete outcome severity could be modeled by a multinomial logit model, heterogeneous effects and correlation in unobserved factors could still have potential limitations in the assumption behind utilizing this model for injury severity (Train, 2009). Thus, a mixed logit model overcomes all of these limitations by generalizing multinomial logit structure, allowing for the parameters $\boldsymbol{\beta}_i$ vector to vary across the observation of crashes (Savolainen et al., 2011). The assumption regarding IID (independently and identically distributed errors), IIA (independence of irrelevant alternatives) and unobserved heterogeneity associated with observations of the multinomial logit model are completely relaxed by introducing a mixed logit approach (Jones and Hensher, 2007).

In order to achieve a better understanding of the injury severity of large-truck involved crashes on the US interstate system, we seek to develop a statistical model that can be used to determine the contributing factors that influence the likelihood of severity outcomes in large-truck involved crashes. To do so, we start with a linear function S_{in} that determines discrete injury severity outcome i (fatality, incapacitating injury, non-incapacitating injury, possible injury, and property damage only) for large-truck involved incident *n* such that (Washington et al., 2011):

$$S_{in} = \boldsymbol{\beta}_i \mathbf{X}_{in} + \varepsilon_{in} \tag{8.1}$$

where $\mathbf{X}_{in}$ is vector of explanatory variables (driver, vehicle, road, and environment variables), $\boldsymbol{\beta}_i$ is vector of estimable parameters, ε_{in} is the error term. If ε_{in}'s are assumed to be generalized extreme value distributed, McFadden (1981) has shown that the multinomial logit results such that:

$$P_n(i) = \frac{EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]}{\sum_I EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]} \tag{8.2}$$

where $P_n(i)$ is probability of large-truck involved incident n having severity outcome i ($i \in I$ with I denoting all possible injury severity outcomes)

As CRIS crash data is likely to have a significant amount of unobserved heterogeneity because the information regarding any of the factors are not obtained from the in-depth crash investigation or reconstruction studies (for example, relating to police reporting, roadway, vehicle, and driver factors), we consider the possibility that elements of the parameter vector $\boldsymbol{\beta}_i$ may vary across observations of each large-truck involved crash by using a random-parameters logit model (also known as the mixed logit model). Previous works by McFadden and Rudd (1994), Geweke et al. (1994), Revelt and Train (1997, 1999), Train (1997), Stern (1997), Brownstone and Train (1999), McFadden and Train (2000), and Bhat (2001) have shown that the development and effectiveness of the mixed logit approach can explicitly account for the variations (across crash observations) of the effects that variables have on the severity outcomes (or choices) considered in this study. The mixed logit model is written as (see Train, 2003):

$$P_{in} = \int \frac{EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]}{\sum_I EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]} f(\boldsymbol{\beta}_i|\boldsymbol{\varphi}) d\boldsymbol{\beta}_i \tag{8.3}$$

where, $f(\boldsymbol{\beta}_i|\varphi)$is the density function of $\boldsymbol{\beta}_i$, $\boldsymbol{\varphi}$ is a vector of parameters of the density function (mean and variance), and all other terms are as previously defined. This model can now account for severity outcome specific variations of the effect of $\mathbf{X}$ on large-truck involved crash probabilities, with the density function $f(\boldsymbol{\beta}_i|\boldsymbol{\varphi})$ used to determine $\boldsymbol{\beta}_i$. Mixed logit probabilities are then a weighted average for different values of $\boldsymbol{\beta}_i$ across the crash observations where some elements of the vector $\boldsymbol{\beta}_i$ may be fixed and some randomly distributed. If the parameters are random, the mixed logit weights are determined by the density function $f(\boldsymbol{\beta}_i|\boldsymbol{\varphi})$ (Milton et al., 2008; Washington et al., 2011).

Maximum likelihood estimation of the mixed logit model shown in Equation (8.3) is undertaken with simulation approaches due to the difficulty in computing probability. The most widely accepted simulation approach is a technique developed by Halton (1960) to generate a systematic non-random sequence of numbers. Halton's technique (known as Halton draws) has been shown to provide a more efficient distribution of the draws for numerical integration than purely random draws (Bhat, 2003; Train, 1999).

A study by Ye and Lord (2011) showed the influence of the sample size on injury severity modeling. Although the analysis by Ye and Lord (2011) regarding the sample size corresponding to each severity model is simulation driven, there are still a few findings that could be generalized in terms of sample size for the three commonly used models. An ordered probit model is the one that requires the least samples (more than 1000), a mixed logit is the most demanding on samples (more than 5000), and a multinomial logit model requirement are located between the ordered probit and mixed logit models (somewhere more than 2000). In our study, the rural crash sample size is 5,484 and the urban crash sample size is 11,560, which are above and over the safe threshold regarding sample size identified by Ye and Lord (2011).

In order to estimate the impact of particular variables on the injury-outcome likelihood, elasticities (or direct-pseudo elasticity) are computed. In the context of the current injury severity model, most of the variables are indicator in nature; direct-pseudo elasticities are estimated to measure the marginal effects of indicator variables when any particular indicator variable switches from 0 to 1 or reverse (Washington et al., 2011). Also, this is translated to percentage change in the likelihood of the indicator variables switching between 0 and 1 or 1 to 0. For binary indicator variables, the direct-pseudo elasticity is estimated as follows (Kim et al., 2010):

$$E_{x_{nk}}^{P_{in}} = \frac{P_{in}[given\ x_{nk} = 1] - P_{in}[given\ x_{nk} = 0]}{P_{in}[given\ x_{nk} = 0]} \tag{8.4}$$

where, P_{in} is given the Equation (8.3) and simulated as shown in Equation (8.5).

x_{nk} = the k-th independent variable associated with injury severity i for observation n.

The unconditional probability in Equation (8.3) (Kim et al., 2010) can be estimated with an unbiased and smooth simulator (McFadden and Train, 2000) that is computed as (Walker and Ben-Akiva, 2002):

$$\hat{P}_{in} = \frac{1}{R}\sum_{r=1}^{R} P_{in} = \frac{1}{R}\sum_{r=1}^{R}\int \frac{EXP[\boldsymbol{\beta}_i\mathbf{X}_{in}]}{\sum_I EXP[\boldsymbol{\beta}_i\mathbf{X}_{in}]} f(\boldsymbol{\beta}_i|\boldsymbol{\varphi})d\boldsymbol{\beta}_i \tag{8.5}$$

where, R = the total number of draws.

Since the direct pseudo-elasticity is calculated for each observation, it is usually reported as the average direct pseudo-elasticity (taking average over the sample) as a measure of the marginal effect of an indicator variable on the likelihood of a particular injury severity outcome (Kim et al., 2010).

With the simulator in Equation (8.5), Maximum Simulated Likelihood Estimation (MSLE) can be used to estimate parameters and this MSLE estimator is asymptotically normal and consistent (Lee, 1992):

$$max_{\boldsymbol{\beta}_{in}} \sum_{n=1}^{N}\sum_{i=1}^{I} y_{in} ln\hat{P}_{in} \tag{8.6}$$

where, N = the total number of observations (i.e., crashes in the sample)

y_{in} = 1 if individual n suffers from injury severity i, 0 otherwise.

Then, a major focus of this study is to address temporal and spatial transferability of severity models. We used CRIS data spatial transferability between two major cities in Texas such as Dallas Fort Worth and Houston. Since the data collection process for the state of Texas is uniform across CRIS database, the spatial transferability evaluation is much valid since there are no major grounds of anomaly or non-uniformity in data collection procedures. The estimated parameters of the multinomial logit models are evaluated with statistical test whether the estimated parameters are spatially or

temporarily transferable. This transferability greatly impacts the scope of data collection and model estimations. Usually, log-likelihood ratio test is conducted to evaluate the spatial and temporal transferability of the estimated parameters in the models (Washington et al., 2011):

$$\chi^2 = -2 * [LL(\beta_{ba}) - LL(\beta_a)] \tag{8.7}$$

where $LL(\beta_{ba})$ is the log-likelihood of at convergence of a model using the converged parameters from region "b" which is Dallas-Fort Worth in this study (using only the data of Dallas Fort Worth) on the data of region "a" which is Houston (restricting the parameters of Dallas Fort Worth severity model). $LL(\beta_a)$ is log likelihood at convergence of the model using region "a" which is Houston data.

The temporal transferability is also conducted in a similar fashion by performing log likelihood ratio test as shown in Equation 8.3 with the notations. In case of temporal transferability test, $LL(\beta_{ba})$ is the log-likelihood of at convergence of a model using the converged parameters from time period "b" which is time period from 2006 to 2008 in this study (using only the data of 2006 to 2008) on the data of time period "a" which is time period from 2009 to 2010 (restricting the parameters of 2006 to 2008 severity model). $LL(\beta_a)$ is log likelihood at convergence of the model using time period "a" which is time period from 2009 to 2010.

The statistics is χ^2 distributed with degree of freedom equal to the number of estimated parameters in $\beta_{ba.}$ If the resulting χ^2 is greater than the critical value of χ^2 then we reject the null hypothesis of single model (i.e., estimated parameters are transferable) representing both regions, and accept the alternative hypothesis that separate models are warranted. In the statistical evaluation process, if the model is affected by the omitted variables and specification errors, then transferability is erroneously rejected (Washington et al., 2011).

There are other methodologies of conducting transferability in case of travel demand, route choice, accident prediction models (Bekhor and Prato, 2009; Ben-Akiva and Morikawa, 1990; Hadayeghi et al., 2006; Hasan et al., 2012). These methodologies include informal parameter

comparisons to formal likelihood ratio test (Hasan et al., 2012). Informal test indicates the ratio comparison of parameters of each single data set (Louviere et al., 2000) and thus showing the potential parameters to be equal despite the differences in scale parameters in the respective data sets. On the other hand, formal test indicates the model equality test with the joint model estimation combining different datasets (Koppleman and Wilmot 1982; Ben-Akiva and Morikawa, 1990; Hadayeghi et al., 2006).

8.3 Data description

The crash data was mainly extracted from the Texas Crash Records Information System (CRIS) over a period of five years from 2006 to 2010. The main focus of this study is the spatial and temporal transferability of severity models of large-truck involved crashes on Texas interstate systems. Each observation in the samples is the maximum injury outcomes of drivers in large-truck involved crashes. These samples are further processed and filtered into a) the major two cities in Texas such as Dallas-Fort Worth and Houston, and b) two different samples of time periods consisting of 2006 to 2008 data and 2009 to 2010 data. Regarding the sample size, Horowitz (1980) showed that an average sample size of 2,000 observations is required for simple multinomial logit and multinomial probit model estimation. In all of the split samples, we have more than 2,000 observations in estimated multinomial logit models.

8.3.1 Spatial crash sample

We split the CRIS sample into major cities by using City IDs in of Houston and Dallas-Fort Worth from 2006 to 2010. Number of observations of Houston and Dallas Fort Worth crash sample was found to be 2,439 and 3,114, respectively. The break-down of five levels of injury outcomes for Houston and Dallas-Fort Worth is shown in Figure 8.1. Descriptive statistics of the key variables in the models are described in Table 8.1.

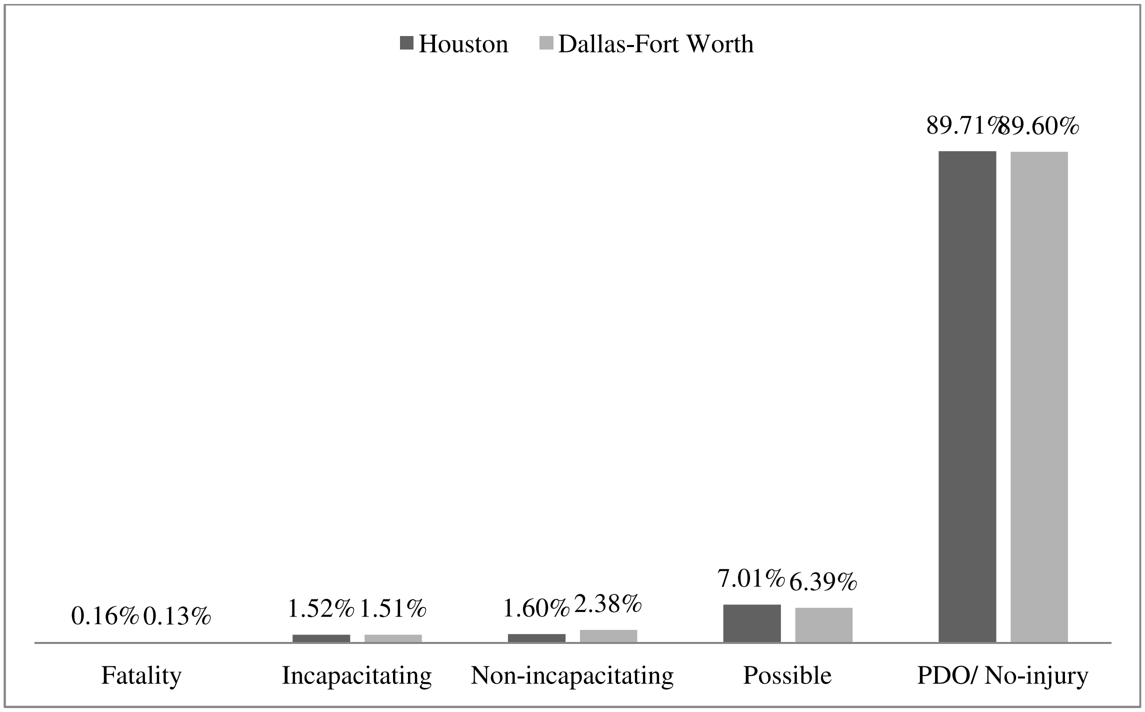

Figure 8.1: Severity Distribution of Houston and Dallas-Fort Worth (2006 ~ 10)

Although there are only few observations in fatality outcomes for both of the cities, we kept as five separate outcomes in the modeling framework to capture the contributing factors associated with each individual outcome. Considering descriptive statistics of Dallas Fort Worth sample, drivers' demographics, driving maneuver, roadway geometry, temporal characteristics, and weather conditions are found statistically significant in the model. Regarding drivers' demographics, there are 93.4% of male drivers, 56.9% white drivers in large truck crashes. Turing to driving maneuvers, going straight or lane keeping accounts for 13.2% of the sample size. Restraint usage reflected by seat belt or not being ejected from the vehicles accounts for 91.5% of the sample size. Considering the geometry, median width greater than 30 feet and left should between 10 to 14 feet accounts for 49.1% and 17.1% of sample size, respectively. The average width of the right shoulder is 10.0 feet. Six lanes in both directions and level roadway surface account for 39.5% and 75.1% of the sample size, respectively. Turning to temporal characteristics, 12 am to 6 am and 9 am to 12 pm account for 16.2% and 17.7% of

sample size, respectively. Then, considering weather condition, wet surface and clear weather condition account for 13.5% and 70.9% of the sample size, respectively.

Considering descriptive statistics of Houston sample, drivers' demographics, driving maneuver, roadway geometry, temporal characteristics, and weather conditions are found statistically significant in the model. Regarding the drivers' demographics, male drivers, white drivers account for 93.9%, and 39.4% of the sample size, respectively. Additionally, drivers of age groups of 25 to 35 and 45 to 55 years account for 18.8%, and 22.2% of the sample size, respectively. Turning to driving maneuvers during the pre-crash stages, going straight or lane keeping accounts for 9.8% of the sample size. Considering the road geometry, median less 30 feet, left shoulder width between 22 to 26 feet, accounts for 12.8% and 12.6% of the sample size, respectively. Also, average number of lanes is 8.3 in both directions and level roadway surface accounts for 81.9% of the sample size. Turning to temporal characteristics, 12 am to 6 am and 9 am to 12 pm timer periods account 12.8% and 18.2% of the sample size, respectively. In addition, fall season of the year such as months from August to December accounts for 28% of the sample size. Then, considering weather condition, clear weather condition and rainy weather accounts for 56.7%, 10.7% of the sample size, respectively. Lighting condition such dark but lighted outside condition accounts for 16.4% of the sample size.

Table 8.1: Descriptive Statistics for Dallas-Fort Worth and Houston

	Dallas-Fort Worth N = 3,114		Houston N = 2,439	
Meaning of Variable	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Fatal outcome				
Ejection of drivers (1 if not ejected outside, 0 otherwise)	0.915	0.279		
Driving maneuver (1 if going straight or lane keeping, 0 otherwise)			0.098	0.297
Incapacitating Injury Outcome				
Surface condition at the time of crash (1 if wet surface, 0 otherwise)	0.135	0.343		
Weather condition at the time of crash (1 if clear weather condition, 0 otherwise)	0.709	0.454	0.567	0.495
Gender of the drivers (1 if male, 0 otherwise)	0.938	0.241		
Time of the day (1 if between 12 am to 6 am, 0 otherwise)	0.162	0.368	0.128	0.334
Width of median (1 if width of median is less than 30 feet, 0 otherwise)			0.128	0.334
Number of direction lanes			8.266	1.508
Months of year (1 if fall months (August to December), 0 otherwise)			0.280	0.449
Non-Incapacitating Injury Outcome				
Race of the drivers (1 if white, 0 otherwise)	0.569	0.495		
Driving maneuver (1 if going straight or lane keeping, 0 otherwise)	0.132	0.339	0.098	0.297
Time of the day (1 if between 9 am to 12 pm, 0 otherwise)	0.177	0.382		
Number of lanes on highways (1 if 6 lanes in both direction, otherwise)	0.395	0.489		
Left shoulder width (1 if width is between 22 to 26 ft, 0 otherwise)			0.126	0.333
Weather condition at the time of crash (1 if clear weather condition, 0 otherwise)			0.567	0.495
Possible Injury Outcome				
Terrain of roadway (1 if level roadway surface, 0 otherwise)	0.751	0.432	0.819	0.385
Race of the drivers (1 if white, 0 otherwise)	0.569	0.495	0.394	0.489
Driving maneuver (1 if going straight or lane keeping, 0 otherwise)	0.132	0.339	0.098	0.297
Ejection of drivers (1 if not ejected outside, 0 otherwise)	0.915	0.279		
Light condition of street (1 if the surrounding area is dark but outside is lighted, 0 otherwise)			0.164	0.371
Age group (1 if age between 45 to 55, 0 otherwise)			0.222	0.415
Non-Injury Outcome (Property-Damage-Only)				
Age group (1 if age between 25 to 35, 0 otherwise)			0.188	0.391
Light condition (1 if the surrounding area is dark but outside is lighted, 0 otherwise)			0.164	0.371
Weather condition at the time of crash (1 if rainy weather condition, 0 otherwise)			0.107	0.309
Time of the day (1 if between 9 am to 12 pm, 0 otherwise)			0.182	0.386
Gender of the drivers (1 if male, 0 otherwise)	0.938	0.241	0.939	0.238
Width of median (1 if width of median is more than 30 ft, 0 otherwise)	0.491	0.499		
Terrain of roadway (1 if level roadway surface, 0 otherwise)	0.751	0.432		
Right shoulder width (ft)	10.022	1.178		
Left shoulder width (1 if width is between 10 to 14 ft, 0 otherwise)	0.171	0.377		

8.3.2 Temporal crash sample

We split the CRIS sample into two time periods such as 2006 to 2008 and 2009 to 2010. Number of observations for 2006 to 2008 and 2009 to 2010 crash sample was found to be 13,032 and 7,463, respectively. The break-down of five levels of injury outcomes for these two time periods is shown in Figure 8.2. Descriptive statistics of the key variables in the models are described in Table 8.2.

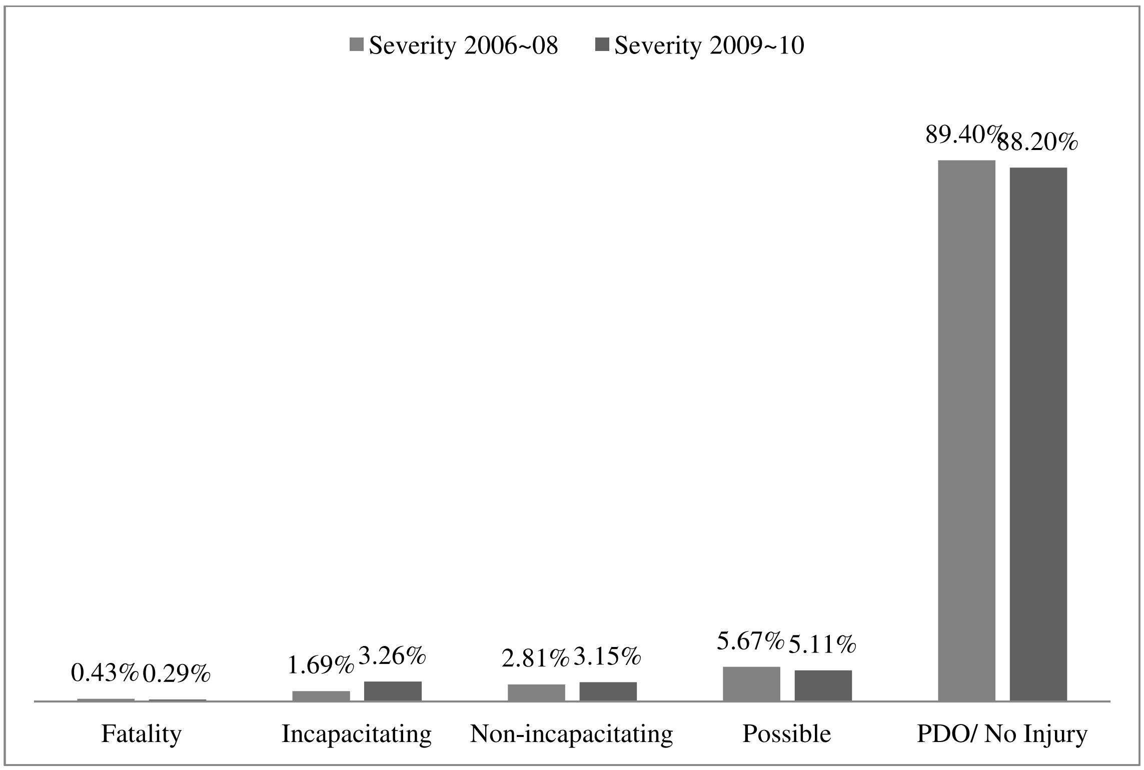

Figure 8.2: Severity Distribution of Grouped Data 2006~08 and 2009~2010

Considering the descriptive statistics of the time period from 2006 to 2008, drivers' demographics, driving maneuver, roadway geometry, temporal characteristics, weather, lighting, and exposure conditions are found statistically significant in the model. Regarding the drivers' demographics, male driver accounts for 94.7% of sample size. Age groups of 25 to 35 years and 45 to 55 years account for 18.5% and 23.4% of the sample, respectively. With driving maneuvers, going straight or lane keeping accounts for 21.9% of the sample size. Turning to roadway geometry, average width of right shoulder is

about 19.7 feet. Six lanes in both direction and level road surface account for 28.6% and 77.0% of the sample size, respectively. Turning to temporal characteristics, 12 am to 6 am and 3 pm to 7 pm time periods account for 16.0% and 19.6% of sample size, respectively. Summer season of the year from June to August accounts for 25.4% of the sample size. Considering weather condition, clear weather condition and dry road surface accounts for 84.8% and 81.5% of the sample size. Land use pattern such as rural and urban settings account for 27.7% and 54.9% of the sample size, respectively. Lighting condition such dark but lighted outside condition accounts for 12.3% of the sample size. Turning to traffic condition, average annual daily traffic is 15,083 vehicles per lane per day.

Considering the descriptive statistics of the time period from 2009 to 2010, drivers' demographics, driving maneuver, roadway geometry, temporal characteristics, weather and lighting, conditions are found statistically significant in the model. Regarding the drivers' demographics, male driver accounts for 93.9% of sample size. With driving maneuvers, going straight or lane keeping accounts for 22.7% of the sample size. Turning to roadway geometrics, median width more than 30 feet accounts for 65.2% of the sample size. Level road surface accounts for 75.2% of the sample size, respectively. Turning to temporal characteristics, 12 am to 6 am and 3 pm to 7 pm time periods account for 15.9% and 21.6% of sample size, respectively. Considering weather condition, clear weather condition and dry road surface accounts for 39.2% and 80.1% of the sample size. Land use pattern such as rural and urban settings account for 25.1% and 58.8% of the sample size, respectively. Lighting condition such dark but lighted outside condition accounts for 11.3% of the sample size.

Table 8.2: Descriptive Statistics for TX 2006~08

Meaning of Variable	Year: 2006 ~ 08 N = 13,032 Mean	Std. Dev.	Year: 2009 ~ 10 N = 7,463 Mean	Std. Dev.
Fatal outcome				
Terrain of roadway (1 if level roadway surface, 0 otherwise)	0.770	0.421		
Time of day (1 if between 3 pm to 7 pm, 0 otherwise)	0.196	0.397		
Time of day (1 if between 12 am to 6 pm, 0 otherwise)	0.160	0.366		
Traffic flow at the time of crash (ADT per lane in each direction – veh/day/ln)	15,082.5	8,786.67		
Surface condition (1 if dry surface, 0 otherwise)	0.815	0.387		
Land-use pattern at crash location (1 if rural area, 0 otherwise)			0.251	0.433
Weather condition at the time of crash (1 if clear weather condition, 0 otherwise)			0.392	0.488
Driving maneuver (1 if going straight or lane keeping, 0 otherwise)			0.227	0.419
Weather condition at the time of crash (1 if clear weather condition, 0 otherwise)			0.391	0.488
Median width (1 if median width more than 30 ft, 0 otherwise)			0.652	0.476
Incapacitating Injury Outcome				
Month of the year (1 if summer months (June to August), 0 otherwise)	0.254	0.435		
Driving maneuver (1 if going straight or lane keeping, 0 otherwise)	0.219	0.414	0.227	0.419
Shoulder width (right shoulder width (feet))	19.672	3.628		
Time of the day (1 if between 3 pm to 7 pm, 0 otherwise)	0.196	0.397		
Land-use pattern at crash location (1 if urban area, 0 otherwise)	0.549	0.497	0.588	0.492
Non-Incapacitating Injury Outcome				
Driving maneuver (1 if going straight or lane keeping, 0 otherwise)	0.219	0.414		
Terrain of roadway (1 if level roadway surface, 0 otherwise)	0.770	0.421		
Time of the day (1 if between 12 am to 6 am, 0 otherwise)	0.160	0.366	0.159	0.365
Months of year (1 if summer months (June to August), 0 otherwise)	0.254	0.435		
Land-use pattern at crash location (1 if rural area, 0 otherwise)	0.277	0.448		
Driving maneuver (1 if going straight or lane keeping, 0 otherwise)			0.227	0.419
Possible Injury Outcome				
Gender of the drivers (1 if male, 0 otherwise)	0.947	0.224	0.939	0.238
Time of the day (1 if between 3 pm to 7 pm, 0 otherwise)	0.196	0.397	0.216	0.411
Land-use pattern at crash location (1 if urban area, 0 otherwise)	0.549	0.497		
Age group (1 if age between 25 to 35, 0 otherwise)	0.185	0.388		
Number of lanes (1 if 6 lanes in both directions, otherwise)	0.286	0.452		
Surface condition at the time of crash (1 if dry surface, 0 otherwise)			0.801	0.399
Terrain of roadway (1 if level roadway surface, 0 otherwise)			0.752	0.432
Non-Injury Outcome (Property-Damage-Only)				
Weather condition at the time of crash (1 if clear weather condition, 0 otherwise)	0.848	0.358	0.392	0.488
Light condition of street (1 if the surrounding area is dark but outside is lighted, 0 otherwise)	0.123	0.328	0.113	0.317
Land-use pattern at crash location (1 if urban area, 0 otherwise)	0.549	0.497	0.588	0.492
Age group (1 if age between 45 to 55, 0 otherwise)	0.234	0.424		
Surface condition at the time of crash (1 if dry surface, 0 otherwise)			0.801	0.399
Traffic flow at the time of crash (ADT per lane in each direction – veh/day/ln)	15,082.5	8,786.67		

8.4 Empirical Results

Model results on injury severities for Dallas-Fort Worth (Table 8.3) and Houston (Table 8.4) as well as for time periods 2006~08 (Table 8.5) and 2009~10 (Table 8.6) are described in the following subsections.

8.4.1 Background of regional crash statistics

Comparing with other cities in Texas, the Houston region was found to have the worst safety problem, accounting for about 26% of all serious crashes in Texas (Levine, 2006). According to more recent 2011 crash statistics in the state of Texas, Houston accounts for 13.4% and 10.2% of all fatal and serious injury crashes, respectively (TxDOT, 2011). On the similar grounds, Dallas and Fort Worth combinedly accounts for 12.7% and 11.9% of all fatal and serious injury crashes, respectively in Texas (TxDOT, 2011). Another motivation to analyze injury severities of large-truck involved crashes in big metropolitan cities in Texas is that Houston and Dallas-Fort Worth accounts for 13.1% and 10.9% of total crashes in the state of Texas (TxDOT, 2011).

8.4.2 Model constants

Constant specific to non-incapacitating injury outcome in Houston model was found to be normally distributed random parameter. With mean of -10.242 and standard deviation of 5.218, 97.47% observations of the sample is below zero, which indicates that overall the drivers are less likely to be involved in non-incapacitating injury outcome for 97.47% of the cases in Houston area. Also, the constant specific to incapacitating injury outcome for time period of 2006~08 and non-incapacitating injury outcome for time period of 2009~10 was found to be normally distributed random parameter. With mean of -4.212 and standard deviation of 4.658, 81.71% observations of the sample is below zero,

which indicates the overall drivers are less likely to be involved in incapacitating injury outcome for 81.71% cases in the time period of 2006~2008. Likewise, with mean of -10.895 and standard deviation of 4.328, 99.41% observations of the sample is below zero, which indicates the overall drivers are less likely to be involved in non-incapacitating injury outcome for 99.41% cases in the time period of 2009~2010.

8.4.3 Drivers' characteristics

Keeping the vehicles inside the lane or lane keep maneuvers increases the likelihood of fatalities and incapacitating injuries in Houston and non-incapacitating and possible injuries in Dallas-Fort Worth regions. This maneuver indirectly indicates the rear-end collision with other passenger vehicles. Keeping the vehicles inside the lane or lane keep maneuvers increases the likelihood of incapacitating and non-incapacitating injuries for 2006~08 model and fatalities, incapacitating, and non-incapacitating injuries for 2009~10 model. This maneuver indirectly indicates the rear-end collision with other passenger vehicles. In case of rear-end collision, three possible crash scenarios are likely to occur: trucks being hit in the rear, trucks hitting rear of passenger vehicles and passenger vehicles going under ride the large trucks as angled collisions. Also, passenger vehicles hitting the rear-end of heavy vehicles results in higher severities. According to a 1997 Insurance Institute for Highway Safety study on fatal crashes between large trucks and passenger vehicles, it is estimated that under-ride occurred in half of these crashes. Of the under-ride crashes, 57 percent involved the front of the truck, 22 percent involved the rear, and 20 percent the side.

Turning to demographics of the drivers, white drivers are less likely to be involved in possible injuries in Houston and non-incapacitating and possible injuries in Dallas-Fort Worth region. For Dallas-Fort Worth model, indicator variable "white driver" was found to be normally distributed random parameter. With mean of -2.716 and standard deviation of 2.080, 90.42% of the sample is below zero,

which indicates white drivers are less likely to be involved with non-incapacitating injuries in 90.42% of cases. Considering the age group of drivers, 25 to 35 years old drivers are more likely to be involved in the non-injury crashes; whereas, 45 to 55 years old drivers are less likely to be involved in possible injuries in Houston region. Male drivers are more likely to be involved in non-injury crashes in both regions. However, male drivers are found to be more involved in incapacitating injuries in Dallas-Fort Worth region. In the sample, male drivers are over represented as found to be more than 90% of the sample size. The age group of 45 to 55 years old is more likely to be involved with non-injury crashes for 2006~08 model. This is because people of this age group are more experienced in driving but might be slower to respond with avoiding maneuvers in critical driving condition on the highways than a young age group (for instance, age group of 25 to 35 years old). However, the age group of 25 to 35 years old is less likely to be involved with possible crashes for 2006~08 model. This is because people of this age group, though less experienced in critical driving conditions, are faster to respond in crash avoiding maneuvers in critical driving condition on the highways than the older age group.

Male occupants are less likely to be involved in possible injuries as opposed to female occupants. In supporting the fact, a study by Chen and Chen (2011) found that females are more likely to be involved in fatal or incapacitating injuries in single and multi-vehicle collisions. Additionally, this evidence is also supported by a study Abdel-Aty and Abdelwahab (2001). Furthermore, drivers, being belted properly inside the vehicles and not ejected outside of the vehicles during the crashes, are less likely to be involved in fatality and possible injuries in Dallas-Fort Worth region. This result is supported indirectly by Abdel-Aty (2003) that not wearing seatbelt would increase the likelihood of severe injuries for the drivers.

8.4.4 Road geometric characteristics

In terms of number of directional lanes where crashes occurred in these two cities, Houston on average has eight lanes; whereas, Dallas-Fort Worth has six lanes in both directions. Also, the average annual daily traffic in Houston is higher than that in Dallas-Fort Worth region.

Median width of less than 30 feet increases the likelihood of incapacitating injuries in Houston; whereas, median width of more than 30 feet increases the likelihood of non-injury crashes in Dallas-Fort Worth. Median width more than 30 feet increases the likelihood of incapacitating injuries for 2009~10 model. Left shoulder width between 22 to 26 feet increases the likelihood of non-incapacitating injuries in Houston; whereas, left shoulder width between 10 to 14 feet increases likelihood of non-injury crashes. Also, as the right shoulder width increases, there is less likelihood of non-injury crashes in Dallas-Fort Worth. Wider right shoulder increased likelihood of incapacitating injuries for 2006~08 model. This is because the drivers feel comfortable in having the advantage of wider right shoulder being part of their risk compensating behavior.

As the number of lanes increase, there is more likelihood of incapacitating injuries in Houston. Similarly, six lanes in both directions are more likely to increase non-incapacitating injuries in Dallas-Fort Worth. This geometric characteristic indicates lane changing maneuvers of the drivers. Level terrain in Houston decreases the likelihood of possible injuries. However, level terrain in Dallas-Fort Worth increases the likelihood of possible and non-injury crashes. Crashes occurring on six lanes in both directions of the highways were less likely to result in possible injuries for 2006~08 model and four lanes more likely to result in non-incapacitating injuries for 2009~10 model. Mostly the large trucks travel on the rightmost lane of the highways, although four lanes provide more ease in lane changing behavior of fast moving passenger vehicles. This ease in lane changing flexibility adds risk of crashes resulting in non-incapacitating injuries. A possible explanation may be that level surfaces may instill increased driver awareness due to favorable driving visibility.

Crashes occurring in the vicinity of a rural area (with a population of less than 5,000) resulted in non-incapacitating injuries for 2006~08 model and fatalities for 2009~10 model. This indicates that the drivers drive faster in rural sections of the highway because the speed limit for rural settings is higher than that for urban settings. Moreover, the driving pattern (e.g., reckless driving patterns) is more likely to be apparent in rural sections because of low enforcement level or perception of lower enforcement level among the drivers driving along rural settings. A more detailed study by Khorashadi et al. (2005) found similar findings for rural settings where the probability of drivers' injuries in crashes involving excessive speeds, improper lane passing and single vehicle collisions increased likelihood of severe injuries or fatalities. Another study by Khattak et al. (2002) also found similar findings that injuries to older drivers increase to more severe if the crash occurred in rural settings. Urban settings experience less severity.

Crashes occurring in an urban area having the population over than 200,000 reduced likelihood of incapacitating injuries for 2006~08 and 2009~10 models and increased likelihood of possible and no-injuries for 2006~08 model and no injuries for 2009~10 models. This indicates that the drivers drive slower in urban sections of the highway because the speed limit for urban is lower than that in rural area. Additionally, the cautious driving pattern is more apparent in urban sections because of high enforcement level. A similar finding by Khorashadi et al. (2005) indicated that in an urban setting there in more likelihood of non-injuries crashes because of improper lane passing and multi-vehicle collisions

Crashes on the level surface are less likely to result in fatalities, and non-incapacitating injuries for 2006~08 model and possible injuries for 2009~10 model. This variable captures some of the driving behavior on the level surface. The large truck drivers are cautious while driving on the level surface, whereas the passenger vehicles could be reckless and the visibility is unobstructed down the highway stretches.

8.4.5 Temporal characteristics

Time of day such as 12 am to 6 am decreases the likelihood of incapacitating injuries in Houston; whereas, this same time period increases the likelihood of incapacitating injuries in Dallas-Fort Worth. A study by Otmani et al. (2005) found that physiological decline in alertness occurs at two time period of the day: afternoon post-lunch dip (between 1 pm and 4 pm) and during the early morning (between 4 am and 6 am). Also, time of day from 9 am to 12 pm decreases the likelihood of non-injury crashes. However, same time period increases the likelihood of non-incapacitating injuries in Dallas-Fort Worth. In addition, months of the year from August to December decreases the likelihood of incapacitating injuries in Houston. The time period from 12 am to 6 am increased likelihood of fatalities and non-incapacitating injuries for 2006~08 model and increased the likelihood of non-incapacitating injuries for 2009~10 model. This time period captures some late night driving behavior which could be influenced by drowsy driving or fatigue. Also, passenger vehicle drivers could have drunk driving cases in the late night. Time period from 3 pm to 7 pm reduced the likelihood of fatalities, incapacitating, and possible injuries for 2006~08 model and possible injuries for 2009~10 model. A similar result is also reflected in a study by Doherty et al. (1998) with a fact that crash rates with all severity increases towards evening (2000 – 2359) and late night to early morning (2400 – 0459).

Summer months from June to August resulted in more likelihood of incapacitating injuries for 2006~08 model as there is more traffic interactions in the highways. Because of sunny weather, there is more exposure of people with their vehicles on the highways. A similar result was also found in terms of increased number and rates and deaths by Brown and Baass (1997) during the months of June to August. This finding is also further evidenced by Ulfarsson and Mannering (2004) that summer months increase incapacitating injuries for female drivers in Sport Utility Vehicle/minivan for single vehicle collisions.

8.4.6 Weather and environmental characteristics

Clear weather condition decreases the likelihood of incapacitating and non-incapacitating injuries in Houston. Similarly, clear weather condition decreases the likelihood of incapacitating injuries in Dallas-Fort Worth. A possible explanation may stem from drivers being more relaxed due to better visible driving conditions. Turning to weather condition, clear weather condition increases likelihood of no injuries for 2006~08 model. Also, it increase likelihood of fatalities and no injuries but reduces the likelihood of incapacitating injuries for 2009~10 model. On the other hand, rainy condition increases the likelihood of non-injury crashes in Houston. However, wet surface condition decreases the likelihood of incapacitating injuries in Dallas-Fort Worth. This is because the friendly weather condition makes the drivers more relaxed and invokes risk-taking behaviors on the benefit of better visible driving condition along the highway sections. A similar result is supported by Edwards (1998) as indicated that accident severity decreases significantly in rainy condition compared with fine weather. On the contrary, clear weather condition reduced 43.5% likelihood of incapacitating injuries because such weather condition makes the drivers benefit from better visible driving condition along the highway sections. Highway geometry has also significant effects on injury outcomes.

Lighting condition also influences driving performance. Dark surrounding but lighted outside decreases likelihood of possible and non-injury crashes. Crashes occurring in the dark highway section (but the in lighted condition) reduced likelihood of no injuries for 2006 ~ 08 and 2009 ~ 10 models. Similar results were found on that fact that darkness in the street or highway increases severe injuries by Morgan and Mannering (2011) for single vehicle crashes for female under 45 years on dry surface and male under 45 years on wet surface, Malyshkina and Mannering (2009) for single vehicle collisions on Indiana interstates, Anastasopoulos and Mannering (2001) for rural interstates in Indiana.

Crashes occurring on dry pavement increase likelihood of fatalities but reduce the likelihood of no injuries for 2006~08 model. Also, it reduces the likelihood of possible injuries and no injuries for

2009~10 model. On dry surface any evasive actions are very effective as opposed to wet surface in terms of effective skidding of tires on dry surface. A more detailed study by Morgan and Mannering (2011) indicated that male drivers less than 45 years old have less likelihood of minor injuries (i.e., combined non-incapacitating and possible injuries) in case of single occupant vehicle, in urban settings and dark road condition.

8.4.7 Exposure characteristics

Higher traffic flow in terms of average daily traffic per lane in each direction reduced the likelihood of fatalities and increased the likelihood of no injuries for 2006~08 model. This is because higher traffic volume indicates the congested traffic condition which slows down the traffic flow, and severity of crashes goes down with speed reduction.

Table 8.3: Mixed Logit Model for Dallas-Fort Worth (2006 ~ 10)

Meaning of Variable	**Estimate**	**t-stat**	***P*-value**
Fatal outcome			
Constant	-3.565	-4.020	0.000
Ejection of drivers (1 if not ejected outside, 0 otherwise)	-3.619	-3.122	0.002
Incapacitating Injury Outcome			
Constant	-4.110	-4.191	0.000
Surface condition at the time of crash (1 if wet surface, 0 otherwise)	-1.010	-2.195	0.028
Weather condition at the time of crash (1 if clear weather condition, 0 otherwise)	-1.567	-4.899	0.000
Gender of the drivers (1 if male, 0 otherwise)	1.681	2.242	0.025
Time of the day (1 if between 12 am to 6 am, 0 otherwise)	0.645	1.925	0.054
Non-incapacitating Injury Outcome			
Constant	-3.438	-4.702	0.000
Race of the drivers (1 if white, 0 otherwise)	-2.716	-2.017	0.044
(Standard error of parameter distribution)	(2.080)	(2.425)	
Driving maneuver (1 if going straight or lane keeping, 0 otherwise)	2.036	7.270	0.000
Time of the day (1 if between 9 am to 12 pm, 0 otherwise)	0.522	1.739	0.082
Number of lanes on highways (1 if 6 lanes in both directions, otherwise)	0.428	1.641	0.101
Possible Injury Outcome			
Constant	-1.524	-2.175	0.022
Terrain of roadway (1 if level roadway surface, 0 otherwise)	0.622	2.519	0.012
Race of the drivers (1 if white, 0 otherwise)	-0.504	-3.256	0.001
Driving maneuver (1 if going straight or lane keeping, 0 otherwise)	1.549	9.303	0.000
Ejection of drivers (1 if not ejected outside, 0 otherwise)	-1.066	-5.128	0.000
Non-Injury Outcome (Property-Damage-Only)			
Gender of the drivers (1 if male, 0 otherwise)	1.418	7.034	0.000
Width of median (1 if width of median is more than 30 ft, 0 otherwise)	0.319	2.423	0.015
Terrain of roadway (1 if level roadway surface, 0 otherwise)	0.916	4.599	0.000
Right shoulder width (ft)	-0.140	-2.203	0.027
Left shoulder width (1 if width is between 10 to 14 ft, 0 otherwise)	0.422	2.193	0.028
Number of observations		3,114	
Restricted log-likelihood		-5011.790	
Log-likelihood at convergence		-1184.367	
Chi-squared value		7654.844	
McFadden pseudo- R squared		0.764	

Table 8.4: Mixed Logit Model for Houston (2006 ~ 10)

Meaning of Variable	**Estimate**	**t-stat**	***P*-value**
Fatal outcome			
Constant	-6.371	-6.217	0.000
Driving maneuver (1 if going straight or lane keeping, 0 otherwise)	3.353	2.885	0.004
Incapacitating Injury Outcome			
Constant	-5.105	-4.397	0.000
Weather condition at the time of crash (1 if clear weather condition, 0 otherwise)	-1.648	-4.031	0.000
Time of the day (1 if between 12 am to 6 am, 0 otherwise)	-1.747	-2.301	0.021
Width of median (1 if width of median is less than 30 ft, 0 otherwise)	1.423	3.505	0.001
Number of directional lanes	0.355	2.862	0.004
Months of year (1 if fall months (August to December), 0 otherwise)	-1.014	-1.874	0.061
Non-incapacitating Injury Outcome			
Constant	-10.241	-1.725	0.085
(Standard error of parameter distribution)	(5.218)	(1.773)	
Driving maneuver (1 if going straight or lane keeping, 0 otherwise)	2.602	1.669	0.095
Left shoulder width (1 if width is between 22 to 26 ft, 0 otherwise)	2.034	1.511	0.131
Weather condition at the time of crash (1 if clear weather condition, 0 otherwise)	-1.621	-1.638	0.101
Possible Injury Outcome			
Constant	-0.487	-1.792	0.073
Terrain of roadway (1 if level roadway surface, 0 otherwise)	-0.472	-2.462	0.014
Race of the drivers (1 if white, 0 otherwise)	-0.639	-3.442	0.001
Driving maneuver (1 if going straight or lane keeping, 0 otherwise)	1.159	5.535	0.000
Light condition (1 if the surrounding area is dark but outside is lighted, 0 otherwise)	-1.114	-2.835	0.000
Age group (1 if age between 45 to 55, 0 otherwise)	-0.558	-2.382	0.017
Non-Injury Outcome (Property-Damage-Only)			
Age group (1 if age between 25 to 35, 0 otherwise)	0.449	2.180	0.029
Light condition (1 if the surrounding area is dark but outside is lighted, 0 otherwise)	-1.439	-4.095	0.000
Weather condition at the time of crash (1 if rainy weather condition, 0 otherwise)	0.573	2.161	0.031
Time of the day (1 if between 9 am to 12 pm, 0 otherwise)	-0.294	-1.519	0.128
Gender of the drivers (1 if male, 0 otherwise)	1.727	8.292	0.000
Number of observations		2,439	
Restricted log-likelihood		-3925.419	
Log-likelihood at convergence		-932.7538	
Chi-squared value		5985.331	
McFadden pseudo R-squared		0.762	

Table 8.5: Mixed Logit Model for Texas (2006 ~ 08)

Meaning of Variable	Estimate	t-stat	*P*-value
Fatal outcome			
Constant	-4.567	-9.812	0.000
Terrain of roadway (1 if level roadway surface, 0 otherwise)	-1.075	-3.803	0.000
Time of day (1 if between 3 pm to 7 pm, 0 otherwise)	-0.941	-1.717	0.086
Time of day (1 if between 12 am to 6 pm, 0 otherwise)	0.631	2.119	0.034
Traffic flow at the time of crash (ADT per lane in each direction – veh/day/ln)	-0.475x 10^(-04)	-2.291	0.022
Surface condition (1 if dry surface, 0 otherwise)	1.101	2.313	0.021
Incapacitating Injury Outcome			
Constant	-4.212	-0.931	0.000
(Standard error of parameter distribution)	(4.658)	(4.875)	
Month of the year (1 if summer months (June to August), 0 otherwise)	0.250	1.638	0.097
Driving maneuver (1 if going straight or lane keeping, 0 otherwise)	0.670	4.684	0.000
Shoulder width (right shoulder width (feet))	0.032	1.527	0.127
Time of the day (1 if between 3 pm to 7 pm, 0 otherwise)	-0.501	-2.354	0.018
Land-use pattern at crash location (1 if urban area, 0 otherwise)	-0.392	-1.811	0.070
Non-incapacitating Injury Outcome			
Constant	-3.527	-18.244	0.000
Driving maneuver (1 if going straight or lane keeping, 0 otherwise)	1.338	11.484	0.000
Terrain of roadway (1 if level roadway surface, 0 otherwise)	-0.401	-3.480	0.000
Time of the day (1 if between 12 am to 6 am, 0 otherwise)	0.355	2.847	0.004
Months of year (1 if summer months (June to August), 0 otherwise)	0.210	1.740	0.082
Land-use pattern at crash location (1 if rural area, 0 otherwise)	0.302	2.016	0.044
Possible Injury Outcome			
Constant	-0.527	-4.154	0.000
Gender of the occupants (1 if male, 0 otherwise)	-2.129	-21.805	0.000
Time of the day (1 if between 3 pm to 7 pm, 0 otherwise)	-0.318	-2.944	0.003
Land-use pattern at crash location (1 if urban area, 0 otherwise)	0.524	2.988	0.002
Age group (1 if age between 25 to 35, 0 otherwise)	-0.221	-2.031	0.042
Number of lanes on highways (1 if 6 lanes in both direction, otherwise)	-0.209	-2.170	0.030
Non-Injury Outcome (Property-Damage-Only)			
Weather condition at the time of crash (1 if clear weather condition, 0 otherwise)	0.182	2.398	0.016
Light condition (1 if the surrounding area is dark but outside is lighted, 0 otherwise)	-0.274	-3.320	0.001
Land-use pattern at crash location (1 if urban area, 0 otherwise)	0.367	2.265	0.024
Age group (1 if age between 45 to 55, 0 otherwise)	0.158	2.204	0.027
Surface condition at the time of crash (1 if dry surface, 0 otherwise)	-0.416	-4.944	0.000
Traffic flow at the time of crash (ADT per lane in each direction – veh/day/ln)	0.969x 10^(-05)	1.937	0.053
Number of observations		13,032	
Restricted log-likelihood		-20974.19	
Log-likelihood at convergence		-5481.496	
Chi-squared value		30985.40	
McFadden pseudo R-squared		0.739	

Table 8.6: Mixed Logit Model for Texas (2009 ~ 10)

Meaning of Variable	**Estimate**	**t-stat**	***P*-value**
Fatal outcome			
Constant	-8.719	-14.782	0.000
Land-use pattern at crash location (1 if rural area, 0 otherwise)	1.866	3.614	0.003
Weather condition at the time of crash (1 if clear weather condition, 0 otherwise)	1.416	3.098	0.002
Driving maneuver (1 if going straight or lane keeping, 0 otherwise)	1.556	3.290	0.001
Incapacitating Injury Outcome			
Constant	-5.797	-3.409	0.001
(Standard error of parameter distribution)	(2.344)	(2.058)	
Weather condition at the time of crash (1 if clear weather condition, 0 otherwise)	-2.046	-3.142	0.002
Driving maneuver (1 if going straight or lane keeping, 0 otherwise)	1.059	3.081	0.001
Median width (1 if median width more than 30 feet, 0 otherwise)	0.494	1.934	0.053
Land-use pattern at crash location (1 if urban area, 0 otherwise)	-1.198	-3.633	0.000
Non-incapacitating Injury Outcome			
Constant	-10.895	-2.919	0.004
(Standard error of parameter distribution)	(4.328)	(2.358)	
Driving maneuver (1 if going straight or lane keeping, 0 otherwise)	3.018	2.800	0.005
Number of lanes on highways (1 if 4 lanes in both direction, otherwise)	0.591	1.803	0.071
Time of the day (1 if between 12 am to 6 am, 0 otherwise)	0.715	1.578	0.115
Possible Injury Outcome			
Constant	-1.625	-7.965	0.000
Gender of the occupants (1 if male, 0 otherwise)	-0.973	-6.074	0.000
Time of the day (1 if between 3 pm to 7 pm, 0 otherwise)	-0.225	-1.638	0.101
Surface condition at the time of crash (1 if dry surface, 0 otherwise)	-0.961	-3.325	0.001
Terrain of roadway (1 if level roadway surface, 0 otherwise)	-0.549	-4.862	0.000
Non-Injury Outcome (Property-Damage-Only)			
Weather condition at the time of crash (1 if clear weather condition, 0 otherwise)	0.506	4.373	0.000
Light condition (1 if the surrounding area is dark but outside is lighted, 0 otherwise)	-0.435	-3.230	0.001
Land-use pattern at crash location (1 if urban area, 0 otherwise)	-0.373	-3.306	0.001
Surface condition at the time of crash (1 if dry surface, 0 otherwise)	-0.928	-3.475	0.001
Number of observations		7,463	
Restricted log-likelihood		-12011.24	
Log-likelihood at convergence		-3502.443	
Chi-squared value		17017.58	
McFadden pseudo R-squared		0.708	

8.5 Transferability evaluation

This study focuses on the spatial and temporal transferability of the severity models. The CRIS database from 2006 to 2010 is split into two major cities such as Houston and Dallas Fort Worth and two time periods such as 2006 to 2008 and 2009 to 2010. We performed log likelihood ratio test between two converged ML models as shown in the Figure 8.3 and detailed estimation process is described in the

following subsections. This test is based on the null hypothesis is that single model represents the two regions and time periods and the estimated parameters are transferable. Obviously, the alternative hypothesis would be separate models are warranted. Since we used the unified crash database, the noise in the parameters is minimized as opposed to other studies where different database or multiple data sources were utilized to conduct this log likelihood ration test (Hasan et al., 2012).

When log likelihood ratio test favors for the null hypothesis indicating the single model represents two regions or two time periods (i.e., estimated parameters are transferable), the two samples should have similar characteristics in the variables. This similarity in characteristics for regions depends on the proximity of the regions or cities as well as the population, city area, socio-economic characteristics, traffic characteristics, driving characteristics, land use patterns, economic productivities of the cities, road connectivity, and above all crash history of the cities. In Figure 8.3, its shows the acceptance of null hypothesis (estimated chi-square is less then critical chi-square value) indicates close proximity of regions or time periods in terms of variables. In this study, this close proximity identifies the cities which are close geographically and traffic condition, road geometry, driving characteristics including drivers' demographics and crash history are also very similar in nature. This is decision step in Figure 8.3, where the analyst can decide on the City IDs which cities should be considered under this analysis.

The following subsection describes the spatial and temporal transferability based on log likelihood ratio test.

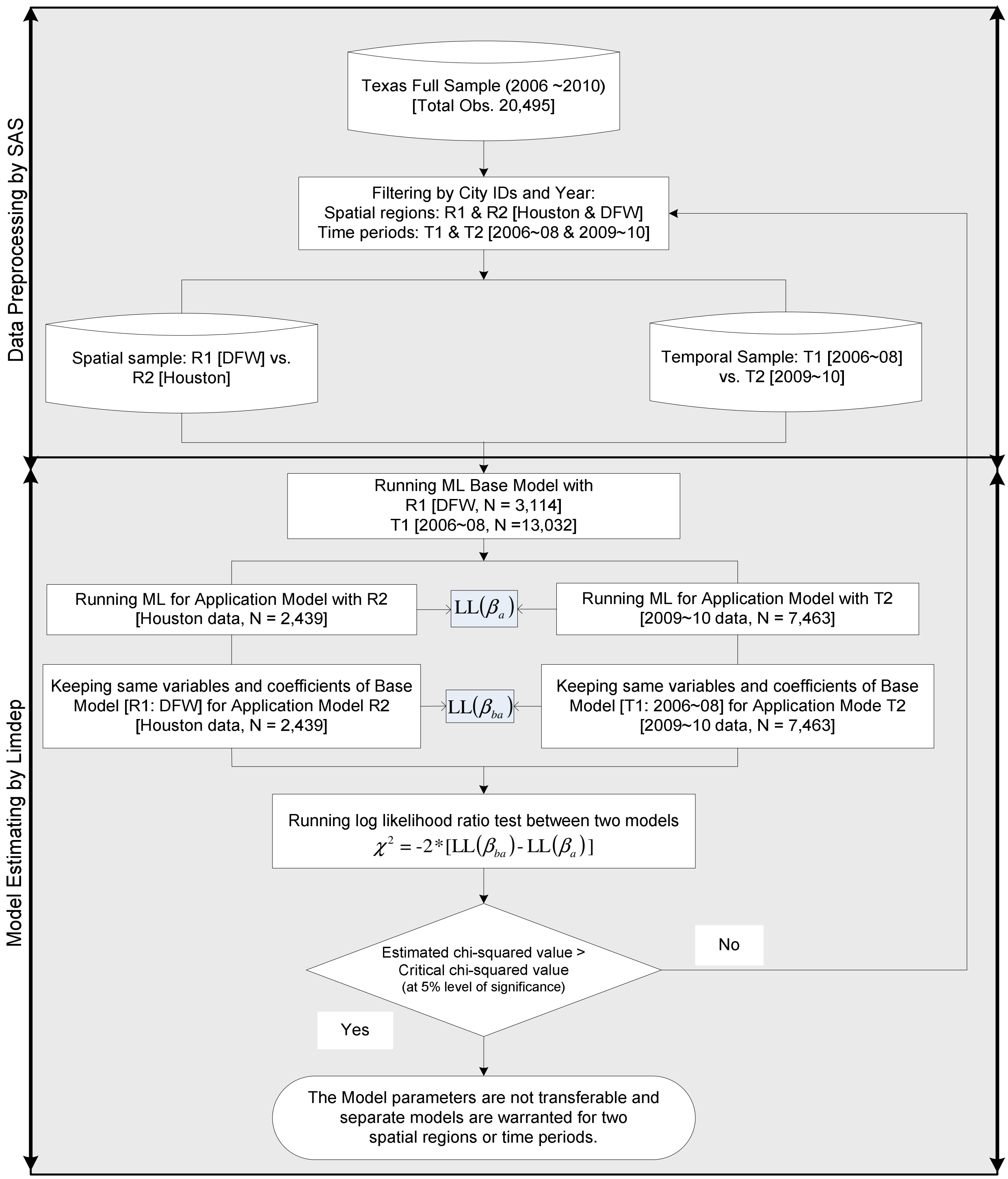

Figure 8.3: Framework for Transferability Evaluation

8.5.1 Spatial transferability test

In running the spatial transferability, the null and alternate hypothesis is set up in the following:

Null hypothesis, H_o: There is no difference in the parameters and single model represents base/ context (i.e., Dallas-Fort Worth) and application (i.e., Houston). (i.e., parameters are stable over the space/ locations indicating spatial transferability).

Alternative hypothesis, H_A: There is difference in the parameters and separate models are warranted (i.e., parameters are not stable over the space and separate models are required for separate locations/regions indicating space in-transferability).

$$\chi^2_{spatial} = -2 * [LL(\beta_{ba}) - LL(\beta_a)] \tag{8.4}$$

where $LL(\beta_{ba}) = LL(\beta_{DFW|HOUS}) = -985.0627$, Degrees of freedom, $\mathrm{n_{DFW}} = 23$

$$LL(\beta_a) = LL(\beta_{HOUS}) = -932.7538, \quad \text{Degrees of freedom, } \mathrm{n_{HOUS}} = 24$$

$$\chi^2 = -2 * [-985.0627 + 932.7538] = 104.6178, \text{ Degrees of Freedom, } n = 23$$

With chi-square value of 104.6178 and 23 degrees of freedom, P-value for right tail test is 0.000. Again, with 0.0001% level of significance or 99.9999% confidence limit, estimated chi-square of 104.6178 is higher than critical chi-square of 55.5246. Finally, the decision of the test indicates that null hypothesis of single model (i.e., parameters are stable over space) is rejected. This clearly provides the evidence that the separate models are warranted for different space or regions (i.e., separate model for Dallas-Fort Worth and Houston). Furthermore, we ran mixed logit for combined dataset (i.e., Dallas-Fort Worth and Houston) and ran split mixed models for Dallas-Fort Worth and Houston. Also, log likelihood ratio test was also conducted in the following manner:

$$\chi^2 = -2[LL_{DFW-HOUS}(\beta^{DFW-HOUST}) - LL_{DFW}(\beta^{DFW}) - LL_{HOUS}(\beta^{HOUS})] \qquad (8.5)$$

where $LL_{DFW-HOUS}(\beta^{DFW-HOUST}) = -2212.501$, Degrees of freedom, $n_{DFW-HOUS} = 24$

$LL_{DFW}(\beta^{DFW}) = -1184.367$, Degrees of freedom, $n_{DFW} = 23$

$LL_{HOUS}(\beta^{HOUS}) = -932.7538$, Degrees of freedom, $n_{HOUS} = 24$

$$\chi^2 = -2 * [-2212.501 + 1184.367 + 932.7538] = 190.7604$$

$$Degrees\ of\ Freedom, n = (23 + 24) - 24 = 23$$

With chi-square value of 190.7604 and 23 degrees of freedom, P-value for right tail test is 0.000. Again, with 0.0001% level of significance or 99.9999% confidence limit, estimated chi-square of 104. 7604 is higher than critical chi-square of 58.6130. Finally, the decision of the test indicates that null hypothesis of single model (i.e., parameters are stable over space) is rejected. This clearly provides the evidence that the separate models are warranted for different space or regions (i.e., separate model for Dallas-Fort Worth and Houston).

8.5.2 Temporal transferability test

In running the temporal transferability, the null and alternate hypothesis is set up in the following:

Null hypothesis, H_o : There are no difference in the parameters and single model represents context (i.e., 2006~08) and application (i.e., 2009~10). (i.e., parameters are stable over the entire time period indicating temporal transferability).

Alternative hypothesis, H_A : There are difference in the parameters and separate models are warranted (i.e., parameters are not stable over the entire time and separate models are required for separate time periods indicating temporal in-transferability).

$$\chi^2_{temporal} = -2 * [LL(\beta_{ba}) - LL(\beta_a)] \quad (8.6)$$

where $LL(\beta_{ba}) = LL\left(\beta_{2006\sim08|2009\sim10}\right) = -3640.652$, Degrees of freedom, $\mathrm{n}_{2006\sim08} = 30$

$LL(\beta_a) = LL(\beta_{2009\sim10}) = -3502.443$, Degrees of freedom, $\mathrm{n}_{2009\sim10} = 24$

$\chi^2 = -2 * [-3640.652 + 3502.433] = 276.438$, Degrees of Freedom, $n = 30$

With chi-square value of 276.39 and 30 degree of freedom, P-value for right tail test is 0.000. Again, with 0.01% level of significance or 99.99% confidence limit, estimated chi-square of 270.39 is higher than critical chi-square of 59.703. Finally, the decision of the test indicates that null hypothesis of single model (i.e., parameters are stable over space) is rejected. This clearly provides the evidence that the separate models are warranted for different time periods (i.e., separate model for 2006 to 2008 and 2009 to 2010). Furthermore, we ran ML for combined dataset (i.e., time period of 2006 to 2010) and ran split ML models for 2006~08 and 2009~10. Also, log likelihood ratio test was also conducted in the following manner:

$$\chi^2 = -2[LL_{2006\sim10}(\beta^{2006\sim10}) - LL_{2006\sim08}(\beta^{2006\sim08}) - LL_{2009\sim10}(\beta^{2009\sim10})] \quad (8.7)$$

where $LL_{2006\sim10}(\beta^{2006\sim10}) = -9231.729$, Degrees of freedom, $\mathrm{n}_{2006\sim10} = 35$

$LL_{2006\sim08}(\beta^{2006\sim08}) = -5481.496$, Degrees of freedom, $\mathrm{n}_{2006\sim08} = 30$

$LL_{2009\sim10}(\beta^{2009\sim10}) = -3502.443$, Degrees of freedom, $\mathrm{n}_{2009\sim10} = 24$

$$\chi^2 = -2 * [-9231.729 + 5481.496 + 3502.443] = 495.58$$

$$Degrees\ of\ Freedom, n = (30 + 24) - 35 = 19$$

With chi-square value of 495.58 and 19 degrees of freedom, P-value for right tail test is 0.000. Again, with 0.0001% level of significance or 99.9999% confidence limit, estimated chi-square of 495.58 is higher than critical chi-square of 50.7955. Finally, the decision of the test indicates that null hypothesis of single model (i.e., parameters are stable over time) is rejected. This clearly provides the evidence that the separate models are warranted for different time periods (i.e., separate model for time periods of 2006~08 and 2009~10).

8.6 Conclusion and summary

In this study a methodological procedure is demonstrated to examine spatial and temporal transferability (i.e., between two major cities and two time periods) utilizing Texas crash data. Through Filtering two regions or two time periods from the main data sample as shown in Figure 8.3, this study also extends the possibility of conducting transferability tests between any cities and adjacent time periods. Figure 8.3 provides the framework of such possibility which could be applied in transportation safety planning projects.

Turning to the study, the estimated coefficients for Houston and Dallas Fort Worth are not transferable between these cities. The number of observations for Dallas Fort Worth (N = 3,114) is higher compared to Houston (N = 2,439) and statistically significant variables for Dallas Fort Worth model were different compared with those for the Houston model. However, crash involvement in clear weather has more pronounced effects in Dallas Fort Worth (70.9% of sample) than in Houston (56.7% of sample). In addition, drivers' demographics such as white drivers are more involved in Dallas Fort Worth (56.9% of sample) than in Houston (39.4% of sample). There are some variables found to be statistically significant in Dallas Fort Worth model such as ejection of the drivers, wet surface condition, median width (more than 30 feet), right shoulder width and left shoulder width (10 to 14 feet), six lanes in both directions, male drivers, are not present in Houston model. Likewise, there are some variables found to be statistically significant in Houston model such as median width (less than 30 feet), number

of lanes in both directions, fall season of the year (August to December), left should width (22 to 26 feet), age of the drivers (25 to 35 years and 45 to 55 years), lighting condition (dark but lighted outside), and rainy weather condition, are not found in Dallas Fort Worth model. These variables represent the contributing factors of injury outcomes in these two major cities. It is clearly evident that big metropolitan cities in Texas like Houston and Dallas Fort Worth, their road geometry, traffic, weather condition, and above all the driving behaviors of the trucks drivers are different.

Additionally, this study estimated coefficients for 2006 to 2008 model and 2009 and 2010 model which were found to not be transferable between these time periods. The number of observations for 2006 to 2008 dataset (N = 13,032) is higher compared to 2009 to 2010 dataset (N = 7,463) and statistically significant variables for 2006 to 2008 model were different compared with those for 2009 to 2010 model. There are some variables found to be statistically significant in 2006 to 2008 model such as time of day (3 pm to 7 pm), summer season of the year (June to August), traffic flow, right shoulder width, six lanes in both directions, drivers' age (25 to 35 years and 45 to 55 years), are not present in 2009 to 2010 model. Surprisingly, there are not many variables found to be statistically significant and different in 2009 to 2010 model except median width (more than 30 feet). Except that particular variable, the rest of the variables are common in both time periods. Since 2009 to 2010 dataset is about half of 2006 to 2008 dataset in terms of number of observations, naturally the number of variables are found to be less in 2009 to 2010 model compared to 2006 to 2008 model.

A study by Hadayeghi et al. (2006) demonstrated the transferability focused to temporal aspect of accident prediction model (i.e., accident frequency models) and updating procedures of these prediction models. Furthermore, this study (Hadayeghi et al., 2006) also indicated future research plan on spatial and temporal transferability identifying the conditions making the "study areas" as close proximity. In our study, we performed both spatial and temporal transferability although the conditions of close proximity between regions were not considered. Cities like Dallas and Fort Worth separately

and San Marcos and San Antonio or any similar size of cities in terms of population, area, road networks, traffic characteristics, and economic productivity should be considered in defining the regions where the models should be spatially transferable. However, this study estimated severity models for Dallas Fort Worth and Houston model and also two different time periods (2006 to 2008 and 2009 to 2010).

Chapter 9: Concluding Comments

This chapter presents concluding comments on research performed, proposed framework, highlights the significance of modeling framework, and suggests innovative directions for future research. Section 10.1 summarizes the research findings and discusses related insights from the perspective of safety professionals. Section 10.2 highlights the significance of the research, and Section 10.3 discusses possible extensions and directions for future work.

9.1 Summary and conclusion

Considering other safety studies dealing with large truck crashes, this study proposes an analytical framework for the analysis injury severity of large-truck involved crashes on highways. The framework utilizes random parameters models for the injury severity analysis to overcome the limitations of the data and to address the existence of unobserved heterogeneity with regards to traditional crash data collection and police reporting systems. We utilize national and state level detailed crash databases to understand the contributing factors affecting five levels of injury outcomes (fatality, incapacitating, non-incapacitating, possible injury, and no injury) of large-truck involved crashes. To model these five outcomes we utilize a mixed logit framework which provides a clearer and fuller understanding of the contributing factors affecting injury levels through greater methodological flexibility over random parameters ordered probit and regular multinomial logit models. In addition, we estimate a random parameter tobit regression model on the maximum injury level – fatality from an exposure based risk and productivity perspective.

The research objectives as well as the research hypothesis were adequately addressed through this dissertation. The objectives of all chapters under the scope of this dissertation were properly addressed and met. More importantly, the research hypothesis with regards to unobserved heterogeneity being present in the crash database (i.e., FARS, NASS-GES, and CRIS) was addressed through the use

of the random parameters models rather than fixed parameter models. Additionally, severity models with split datasets were statistically significant as opposed to combined model as evidenced from chapter 4, chapter 6, and chapter 7 where splitting subset of dataset were addressed from single- and multi-vehicle collision models, rural and urban models, and different time periods models, respectively. The following paragraphs highlight the research accomplishments chapter by chapter.

Chapter 3 deals with the research starting with a framework of maximum injury severity – fatality and taking the rate as fatalities per million truck-miles and fatalities per ton-miles of freight as dependent variables brought up some significant variables into the framework of injury severity analysis. The significant factors associated maximum injury outcomes, utilizing the national crash database namely Fatality Analysis Reporting Systems (FARS) over a period of four years (2005 to 2008), provides useful insights about the large trucks on US interstates systems. Under the scope of this research, random parameter tobit regression was demonstrated as a viable modeling framework for research. These two exposure based measures used in this framework namely fatalities per million truck-miles and ton-miles of freight highlights the amount of fatal risks with travel based exposure and fatal risks from economic productivity role associated with large trucks.

Considering the methodological advantage of mixed logit (random parameter logit model), we analyzed injury outcomes in five separate utility functions in Chapter 4. Additionally, we split the sample into two subsets – single and multi-vehicle collisions to estimate the parameters which were found accurate and specific to single and multi-vehicle collisions. Statistical test namely the log likelihood ratio test indicated the statistical significance of separate model estimation. Variables such as negotiating curve, running off the roadway, day of week (i.e., Friday) indicates single-vehicle collisions. While variables such as manner of collisions (i.e., angled), pre-crash driving maneuvers (i.e., lane changing, lane keeping), time of day (4 to 7 pm) clearly indicates multi-vehicle collisions. Additionally, the state indicator variable for Texas was found to be statistically significant.

Chapter 5 presents the severity model for the Texas crash data. A mixed logit modeling approach was found to provide more insights of the contributing factors in the State of Texas. With the detailed CRIS data for the period of five years (2006 to 2010), we estimated a mixed logit model for five levels of drivers' injury in large-truck involved crashes on Texas interstates. The model results indicated drivers' demographics (i.e., male, age group of 25 to 35 years and 45 to 55 years old), land use characteristics (i.e., rural and urban areas), temporal characteristics (i.e., 12 to 6 am, 3 to 7 pm, June to August, September to December), traffic characteristics (i.e., AADT per lane more than 2000 vehicle per day per lane), weather characteristics (i.e., clear sky), road geometrics (i.e., shoulder width, level surface), and lighting characteristics (i.e., dark condition) strong influence on injury outcomes in large truck crashes in Texas. The model results also found that constant specific to fatality, non-incapacitating, possible injury, male drivers, drivers of age group from 45 to 55 years, and traffic flow were normally distributed random parameters for Texas during the time frame of 2006 to 2010.

Chapter 6 illustrated the severity analysis for rural (population less than 5,000) and urban (population more than 200,000) areas for the State of Texas with a log likelihood ratio test indicating the statistical significance of split subsets. These models clearly indicated the significance of variables specific to rural and urban areas through split model estimation. There is a wide variation of parameter estimates and different variables statistically significant for different models. The parameters that are very distinct in rural and urban model include the driver's demographics, driving characteristics, road geometrics, types of collision, and temporal characteristics. Furthermore, some variables in both models indicate the similar injury outcomes (i.e., sign of same direction). However, some variables clearly indicated severe injury outcome in rural areas because of remote distance and low medical response, and severity of the crash dynamics (for instance, speeding, hitting fixed objects).

Chapter 7 dealt with the time of day in urban areas in Texas which provided some interesting insights with regards to the different time periods consider of AM peak (6 to 9 am), PM peak (4 to 7

pm), and off peak (other than AM and PM peak). The analysis demonstrated the significance of using subset of datasets corresponding to three different time periods (e.g., AM, PM, and off peak) rather than a combined dataset to estimate the parameters for three different time periods through mixed logit models for large-truck involved crashes in urban areas on Texas highways. The analysis provided evidence of wide variations of sign and magnitude of individual parameter estimates in these models. The contributing factors obtained from model estimation were drivers' demographics, driving behavior, roadway geometrics, traffic characteristics, weather characteristics, temporal characteristics, and crash dynamics. The results showed that driving patterns as reflected through these factors widely vary depending on the time of day and so do level of injury severities in large truck crashes in urban areas.

Chapter 8 demonstrated the application of transferability of severity models across the geographical locations and time periods. We estimated severity outcome models for two major cities in Texas over a time period of five years from 2006 to 2010. Then, an appropriate statistical test is performed to evaluate the spatial transferability, indicating that severity outcome model developed for one city could be applied to another. Similarly, temporal transferability, meaning that severity model developed with one time period could be applied to another time period, is evaluated through a statistical test for two different time periods. A unified crash data reporting system in Texas commonly known as CRIS is utilized to develop mixed logit models to estimate the parameters and perform spatial and temporal transferability. The results from the statistical test suggest that the parameters of the severity model are not transferable over two different major cities and two different time periods. Regarding the geographical locations, the basis of comparing these two locations requires to further research to define or set the scale as indicated from past studies. For instance, cities like Dallas and Fort Worth (separately) and San Marcos and San Antonio or any similar size of cities in terms of population, area, road networks, traffic characteristics, and economic productivity should be considered in defining the regions where the models should be spatially transferable. These findings provide important implications and

insights for the policy makers, safety planners, freight managers, and the overall stakeholders. This is because proper countermeasures are usually devised and planned through empirical severity outcome models in freight transportation. From the perspective of safety planning, spatially and temporally transferable models provide a strong basis of flexibility, effectiveness, and efficiency with limited time and resources in freight safety management.

In summary, in regards to the national crash datasets (i.e., FARS, NASS-GES), tobit and mixed logit models indicated that state of Texas was more likely to experience severe injury outcomes with regards to large-truck involved crashes. Also, both of mixed logit models indicated that crashes in curved segments most likely to result in severe injury outcomes. However, single trailing unit trucks more likely to result in no injury outcomes. Likewise, rollover crashes involving large trucks result in more likelihood less severe injuries.

Considering state specific crash dataset – CRIS, we estimated aggregate model as well as disaggregate models with subset of crash datasets. Texas and rural and urban models indicated that male drivers are less likely to be involved in severe crashes. Also, level and dry surface was less likely to increase severe injury outcomes. However, the time period between 12 am to 6 am increased the likelihood of severe injury outcomes. Likewise, clear weather condition also increased the likelihood of severe injury outcomes. Based on the time of day in Texas, mixed logit models for AM, PM, and off peak periods indicated that percent of truck traffic increased the likelihood of severe injury outcomes. Likewise, going straight (i.e., lane keeping) also indirectly relates to rear-end collisions indicated severe injury outcomes. Turning to drivers' demographics, male drivers and white drivers are less to be involved in more severe injury crashes.

Furthermore, the variables considered in the modeling approaches were approximately 10% of the sample size. However, there are some variables which indicated strong correlation with the injury outcomes found in the previous studies, were taken in the modeling estimation process even though they

could be than 10% of the sample. Also, the correlation matrix was also considered in the model estimation process. None of the variables were found to be more than 0.7 with another variable in the correlation matrix.

9.2 Research contributions

This research provides a conceptual framework of estimating discrete outcome severity models to explore the contributing factors using national to state level data for: split subsets based on number of vehicles, locational characteristics and temporal characteristics; between geographic locations and time periods in large-truck involved crashes. The research performed in this dissertation contributes to the existing literature of severity analysis by considering data fusion concept among crash, vehicle, and drivers' datasets and econometric models ranging from regression to discrete outcome models considering random parameters in the estimation process in the context of large-truck involved crashes for five injury categories, a first. In addition, this research performed here provides a framework for safety planners, transportation professionals, and the freight industry to analyze and asses the impacts of injury severity due to large-truck involved crashes.

9.3 Future Research Direction

In future research, the framework presented in this dissertation can be extended to apply spatial econometrics to uncover the interactions of spatial correlation of crash and injury outcomes in specific locations of interest. Also, extending the framework to study the potential impacts of the proximity to trade borders.

References

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

Abdel-Aty, M, and E. Radwan. 2000. "Modeling traffic accident occurrence and involvement." *Accident Analysis and Prevention* 32(5): 633-642.

Abdel-Aty, M., and H. T. Abdelwahab. 2001. "Development of artificial neural network models to predict driver injury severity in traffic accidents at signalized intersections." *Transportation Research Record: Journal of the Transportation Research Board* 1745:6–13.

Abdel-Aty, M., and H. T. Abdelwahad. 2004. "Analysis and prediction of traffic fatalities resulting from angle collisions including the effect of vehicles' configuration and compatibility." *Accident Analysis and Prevention* 36(3): 457–469.

Amemiya, T. 1973. "Regression-analysis when dependent variable is truncated normal." Econometrica 41(6): 997–1016.

Amemiya, T. 1985. *Advanced Econometrics*. Cambridge, MA: Harvard University Press.

Anastasopoulos, P. Ch., and F. L. Mannering, 2009. "A note on modeling vehicle-accident frequencies with random parameter count models." *Accident Analysis and Prevention* 41(1): 153-159.

Anastasopoulos, P..Ch., S. Labi, M. G. Karlaftis, and F. L. Mannering. 2009a. "Exploring relationships between aggregate pavement performance, and surface geology, climate, and expenditure: an econometric analysis of state-level data." 88th Annual Meeting of Transportation Research Board, Washington, DC.

Anastasopoulos, P. Ch., S. Labi, and B. McCullouch. 2009b. "Analyzing duration and prolongation of performance-based contracts using hazard-based duration and zero-inflated random parameters poisson models." *Transportation Research Record: Journal of the Transportation Research Board* 2136: 11-19.

Anastasopoulos, P. Ch., A. P. Tarko, and F. L. Mannering. 2008. "Tobit analysis of vehicle accident rates on interstate highways." *Accident Analysis and Prevention* 40(2): 768–775.

Anastasopoulos, P. Ch., F. L. Mannering, V. N. Shanker, and J. E. Haddock. 2012. "A study of factors affecting highway accident rates using the random-parameters tobit model." *Accident Analysis and Prevention* 45(1): 628–633.

Anastasopoulos, P. Ch., and F. L. Mannering. 2009. "A Note on modeling vehicle accident frequencies with random-parameters count models." *Accident Analysis and Prevention* 41(1): 153–159.

Anastasopoulos, P. Ch., and F. L. Mannering. 2001. "An empirical assessment of fixed and random parameter logit models using crash- and non-crash-specific injury data." *Accident Analysis and Prevention* 43(3): 1140–1147.

Araki S., and K.Murata. 1986. “Social risk factors for the mortality from motor vehicle accident.” *The Tohoku Journal of Experimental Medicine* 149(4): 379–387.

Atherton, T. J., and M. E. Ben-Akiva, 1976. “Transferability and updating of disaggregation travel demand models.” *Transportation Research Record: Journal of the Transportation Research* 610: 12–18.

Badoe, D.A., and E. J. Miller. 1995. “Comparison of alternative methods for updating disaggregated logit mode choice models.” *Transportation Research Record: Journal of the Transportation Research* 1493: 90–100.

Baker, C. J. 1991. “Ground vehicles in high cross winds part III: The interaction of aerodynamic forces and the vehicle system.” *Journal of Fluids Structure* 5(2): 221–241.

Baker, S.P., R. A. Whitfield, and B. O’Neill. 1987. “Geographic variations in mortality from motor vehicle crashes.” *The New England Journal of Medicine* 316(22): 1384–1387.

Bekhor, S., and C. G. Prato. 2009. “Methodological transferability in route choice modeling.” *Transportation Research Part B* 43(4): 422–437.

Ben-Akiva, M. E., and T. Morikawa. 1990. Estimation of travel demand models from multiple data sources, Proceedings of 11th International Symposium on Transportation and Traffic Theory, M. Koshi, ed., Elsevier, NY: 461–476.

Ben-Akiva, M. E., and D. Bolduc. 1987. “Approaches to model transferability and updating: the combined transfer estimator.” *Transportation Research Record: Journal of the Transportation Research* 1139: 1–7.

Ben-Akiva, M.E. Issues in transferring and updating travel-behaviour models. In: Stopher, P.R., et al. (Eds.) 1981. New Horizons in Travel Behaviour. Lexington Books, Lexington, MA: 665–686.

Ben-Bassat, T., and D. Shinar. 2011. “Effect of shoulder width, guardrail and roadway geometry on driver perception and behavior.” *Accident Analysis and Prevention* 43(6): 2142– 2152.

Bhat, C. 2001. “Quasi–random maximum simulated likelihood estimation of the mixed multinomial logit model.” *Transportation Research Part B* 17(1): 677-693.

Bhat, C. 2003. “Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences.” *Transportation Research Part B* 37(1): 837-855.

Blatt, J., S.M. Furman. 1998. “Residence location of drivers involved in fatal crashes.” *Accident Analysis and Prevention* 30(6): 705–711.

Blower, D. F., and L. P. Kostyniuk. 2007. Strategies to Reduce CMV-involved Crashes, Fatalities, and Injuries in Michigan, University of Michigan Transportation Research Institute. http://www.michigan.gov/documents/msp/CMV_Crashes_in_Michigan-FINAL_206130_7.pdf. (Accessed on April 25, 2011).

Bonneson, J. A., L. Dominique, K. Zimmerman, K. Fitzpatrick, and M. P. Pratt. 2007. Development of tools for evaluating the safety implications of highway design decisions. *In: TTI Report FHWA/TX-07/0-4703-4*. Texas Transportation Institute, College Station, TX.

Boufous, S., C. Finch, A. Hayen, and A. Williamson. 2008. "The Impact of environmental, vehicle and driver characteristics on injury severity in older drivers hospitalized as a result of a traffic crash." *Journal of Safety Research* 39(1): 65–72.

Brown, B., and K. Baass. 1997. "Seasonal variation in frequencies and rates of highway Accidents as function of severity." *Transportation Research Record: Journal of the Transportation Research* 1581: 59–65.

Brownstone, D., and K. Train. (1999). "Forecasting new product penetration with flexible substitution patterns." *Journal of Econometrics* 89(1-2): 109-129.

Bureau of Transportation Statistics (BTS), Research and Innovative Technology Administration (RITA), Special Tabulation. U.S. Department of Transportation, http://www.bts.gov. (accessed on October 10, 2010).

Carson, J., and F. L. Mannering. 2001. "The effect of ice warning signs on accident frequencies and severities." *Accidient Analysis and Prevention* 33(1): 99–109.

Cate, M. A., and S. H. Richards. 2001. "An evaluation of large truck rollover crashes on Tennessee interstate highways." 80th Annual Meeting of Transportation Research Board, Washington, DC.

Chakravarthy, A., K. Song, and E. Feron. 2009. "Preventing automotive pileup crashes in mixed-communications environments." *IEEE Transaction on Intelligent Transportation System* 10(2): 211–225.

Chang, Li-Y., F..L. Mannering. 1998. "Predicting Vehicle Occupancies from Accident Data: An Accident Severity Approach." *Transportation Research Record: Journal of the Transportation Research Board* 1635: 93-104.

Chang, Li-Y., and F. L Mannering. 1999. "Analysis of injury severity and vehicle occupancy in truck- and non-truck-involved accidents." *Accident Analysis and Prevention* 31(5): 579–592.

Chen, F., and S. Chen. 2011. "Injury severities of truck drivers in single- and multi-vehicle accidents on rural highways." *Accident Analysis and Prevention* 43(5): 1677-1688.

Chimba, D., and T. Sando. 2009. "Neuromorphic prediction of highway injury severity." *Advances in Transportation Studies* 19:17–26.

Chin, H. C., and M. A. Quddus. 2003. "Applying the Random Effect Negative Binomial Model to Examine Traffic Accident Occurrence at Signalized Intersections." *Accident Analysis and Prevention* 35(2): 253–259.

Chistoforou, Z., S. Cohen, and M. G. Karlaftis. 2010. "Vehicle occupant injury severity on highways: an empirical investigation." *Accident Analysis and Prevention* 42(6): 1606 –1620.

Clark, D. E., B. M. Cushing. 1999. "Predicting regional variations in mortality from motor vehicle crashes." *Academic Emergency Medicine* 6(2): 125–130.

Clark, D. E., and B. M Cushing. 2002. "Predicted effect of automatic crash notification on traffic mortality." *Accident Analysis and Prevention* 34(4): 103–109.

Clark, D. E., and B. M Cushing. 2004. "Rural and urban traffic fatalities, vehicle miles, and population density." *Accident Analysis and Prevention* 36(6): 967–972.

Craft, R. 2010. 2009: Historic truck crash declines, webinar, Federal Motor Carrier Safety Administration, United States Department of Transportation, Washington, D.C.

Choueiri, E.M., R. Lamm, J.H. Kloeckner, and T. Mailaender. 1994. "Safety aspects of individual design elements and their interactions on two-lane highways: international perspective." *Transportation Research Record: Journal of the Transportation Research Board* 1445: 34–46.

Doherty, S. T., J. C. Andrey, and C. MacGrego. 1998. "The situational risks of young drivers: The influence of passengers, time of day and day of week on accident rates." *Accident Analysis and Prevention* 30(1): 45-52.

Douglas, J.G. 2003. Strategies for Managing Increasing Truck Traffic, NCHRP Report 314, TRB, National Research Council, Washington, D.C.

Driscoll, T., S. Marsh, B. McNoe, J. Langley, N. Stout, A-M. Feyer and A. Williamson. 2005. "Comparison of fatalities from work related motor vehicle traffic incidents in Australian, New Zealand, and the United States." *Journal of Injury Prevention* 11(5): 294-299.

Duncan, C. S., A. J. Khattak, and F. M. Council 1998. "Applying the Ordered Probit Model to Injury Severity in Truck–Passenger Car Rear-End Collisions." *Transportation Research Record: Journal of the Transportation Research Board* 1635: 63-71.

Eberhardt, M. S., D.D. Ingram, D.M. Makuc. 2001. Urban and Rural Health Chartbook. Health, United States, 2001. National Center for Health Statistics, Hyattsville, Maryland.

Edwards, J. B. 1998. "The relationship between road accident severity and recorded weather." *Journal of Safety Research* 29(4): 249-262.

Eluru, N., C. R. Bhat, and D. A. Hensher, 2007. "A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes." *Accident Analysis and Prevention* 40(3):1033-1054.

FARS Analytic Reference Guide1975 to 2009, 2010. National Highway Safety Administration, DOT HS 811 352. http://www-nrd.nhtsa.dot.gov/Pubs/811352.pdf. (Accessed on August 28, 2010).

Federal Highway Administration. 2012. Highway statistics, 2010. U.S. Department of Transportation. Washington, DC.

Garbarino, S., L. Nobili, M. Beelke, F. D. Carli, and F. Ferrillo. 2001. "The Contributing role of sleepiness in highway vehicle accidents." Journal of Daytime Sleepiness 24(2): 203–206 (accepted for publication).

Geedipally, S. R., and D. Lord. 2010. "Investigating the effect of modeling single-vehicle and multi-vehicle accidents separately on confidence intervals of Poisson–gamma models." *Accident Analysis and Prevention* 42(4): 1273–1282.

Geweke, J., M. Keane, and D. Runkle. 1994. "Alternative computational approaches to inference in the multinomial probit model." *The Review of Economics and Statistics* 76(4): 609-632.

GHSA, 2012. Governors Highway Safety Association, 2012, http://www.ghsa.org/html/stateinfo/bystate/tx.html. (accessed on April 26, 2012).

Gkritza, K., and F. L. Mannering. 2008. "Mixed logit analysis of safety-belt use in single- and multi-occupant vehicles." *Accident Analysis and Prevention* 40(2): 443-451.

Golob, T., W. Recker, and J. D. Leonard. 1987. "An analysis of the severity and incident duration of truck-involved freeway accidents." *Accident Analysis and Prevention* 19(5): 375–396.

Golob, T. F., A. C. Regan. 2003. Truck-involved Crashes and Traffic Levels on Urban Freeways. 84th Annual Meeting of Transportation Research Board, Washington DC.

Governors Highway Safety Association, 2012, http://www.ghsa.org/html/stateinfo/bystate/tx.html (accessed on March 26, 2012).

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

Greene, W. 1997. Econometric Analysis, 3rd ed. Macmillan, NY.

Greene, W. 1999. "Marginal effects in the censored regression model." *Economic Letters* 64, 43–49.

Greene, W. 2007. Limdep, Version 9.0. Econometric Software, Inc., Plainview, NY.

Griffith, M. S. 1999. "Safety evaluation of rolled-in continuous shoulder rumble strips installed on freeways." *Transportation Research Record: Journal of the Transportation Research Board* 1665: 28–35.

Gunn, H. F. 2001. "Spatial and temporal transferability of relationships between travel demand, trip cost and travel time." *Transportation Research Part E: Logistiscs Transportation Revivew* 37(2-3): 163–189.

Gunn, H. F., M. E. Ben-Akiva, and M. A. Bradley,. 1985. "Tests of the scaling approach to transferring disaggregate travel demand models." *Transportation Research Record: Transportation Research Record* 1037: 21–30.

Habib, K. N., J. Swait, and S. Salem. 2012. Investigating Structural Changes in Commuting Mode Choice Preferences with Repeated Cross-Sectional Travel Survey Data: The Contexts of Greater Toronto and Hamilton (GTHA) Area, 13th International Conference on Travel Behaviour Research, Toronto.

Hadayeghi, A., A. S. Shalaby, and B. Persaud, Carl Cheung. 2006. "Temporal transferability and updating of zonal level accident prediction models." *Accident Analysis and Prevention* 38(3): 579-589.

Halton, J. 1960. "On the efficiency of evaluating certain quasi-random sequences of points in evaluating multi-dimensional integrals." *Numerische Mathematik* 2: 84–90.

Hair, J. F., Jr; W. C. Black, B. J. Babin, and R. E Anderson. 2010. *Multivariate Data Analysis*, Seventh Edition. Upper Saddle River, New Jersey: Prentice Hall.

Harwood, D. W., K. M. Bauer, K. R Richard, D. K. Gilmore, J. L. Graham, I. B. Potts, D. J. Torbic, and E. Hauer. 2007. Methodology to predict the safety performance of urban and suburban arterials. *In: NCHRP Web Document No. 129: Phases I and II*. Transportation Research Board of the National Academies, Washington, D.C.

Hasan, S., R. Mesa-Arango, S. Ukkusuri, and P. Murray-Tuite. 2012. "Transferability of hurricane evacuation choice model: joint model estimation combining multiple data sources." *Journal of Transportation Engineering* 138(5): 548-556.

Hauer, E. 1997. Observational Before-After Studies in Road Safety. Pergamon, NY.

Hauer, E. 2000. Shoulder width, shoulder paving, and safety. http://www.roadsafetyresearch.com (unpublished work).

Helai, H., H. C. Chin, and M. M. Haque. 2008. "Severity of driver injury and vehicle damage in traffic crashes at intersections: a Bayesian hierarchical analysis." *Accident Analysis and Prevention* 40(1): 45–54.

Horowitz, J. 1980. "The accuracy of the multinomial logit model as an approximation to the multinomial probit model of travel demand." *Transportation Research Part B* 14(4) :331–341.

Federal Motor carrier Safety Administration, 2009 Large Truck Crash Overview http://www.fmcsa.dot.gov/facts-research/LTCO2009/2009LargeTruckCrashOverview.aspx, (accessed on April 25, 2011).

Huang, H., H. C. Chin, and A. H. H. Heng, 2006. "Effect of red light camera on accident risk at intersections." *Transportation Research Record: Journal of the Transportation Research* 1969:18-26.

Insurance Institute for Highway Safety, 2009. Fatality Facts 2009 – Large Trucks. http://www.iihs.org/research/fatality_facts_2009/largetrucks.html#sec3. (accessed January 05, 2012).

Insurance Institute for Highway Safety. Fatality Facts 2010, Large Trucks.

http://www.iihs.org/research/fatality.aspx?topicName=Largetrucks. (accessed January 05, 2012).

Islam, M. B., and S. Hernandez. An Empirical Analysis of Fatality Rates for Large-truck involved Crashes on Interstate Highways. In 3rd International Road Safety and Simulation, CD-ROM. Indianapolis, 2011.

Islam, M. B., and S. Hernandez. Large-truck involved Crashes: An Exploratory Injury Severity Analysis. In 91st Transportation Research Board, CD-ROM. Washington D.C., 2012.

Islam, S., and F. L. Mannering. 2006. "Driver aging and its effects on male and female single-vehicle accident injuries: some additional evidence." *Journal of Safety Research*, 37(3): 267–276.

Ivan, J. N., C. Wang, and N. R. Bernardo. 2000. "Explaining two-lane highway crash rates using land use and hourly exposure." *Accident Analysis and Prevention* 32(6): 787–795.

Ivan, J. N., R. K. Pasupathy, and P. J. Ossenbruggen. 1999. "Differences in causality factors for single and multi-vehicle crashes on two-lane roads." *Accident Analysis and Prevention* 31(6): 695–704.

Ivan, J. N. 2004. "New approach for including traffic volumes in crash rate analysis and forecasting." *Transportation Research Record: Journal of the Transportation Research Board* 1897: 134–141.

Jones, S, and D. A. Hensher. 2007. "Evaluating the behavioural performance of alternative logit models: an application to corporate takeovers research." *Journal of Business Finance Accunting* 34(7-8): 1193–1220.

Jonsson, T., J. N. Ivan, and C. Zhang. 2007. "Crash prediction models for intersections on rural multilane highways: differences by collision type." *Transportation Research Record: Journal of the Transportation Research Board* 2019: 91–98.

Jovanis, P. P., and H.-Li Chang. 1986. "Modeling the relationship of accidents to miles traveled." *Transportation Research Record: Journal of the Transportation Research Board* 1068: 42–51.

Jovanis, P. P., and J. Delleur. 1983. "Exposure-based analysis of motor vehicle accidents." *Transportation Research Record: Journal of the Transportation Research Board* 919: 1–7.

Khattak, A. J., R. J. Schneider, and F. Targa. 2003. "Risk factors in large truck rollovers and injury severity: analysis of single-vehicle collisions." 82nd Annual Meeting of Transp. Res. Board Meeting, Washington, DC.

Khattak, A. J., M. D. Pawlovich, R. R. Souleyrette, and S. L. Hallmark. 2002. "Factors related to more severe older driver traffic crash injuries." *Journal of Transportation Engineering* 128(3): 243-249.

Kim, J.-Ki, G. F. Ulfarsson, V. N. Shankar, and F. L. Mannering. 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.

Khorashadi, A., D. N., V. N. Shankar, and F. L. Mannering. 2005. “Differences in rural and urban driver-injury severities in accidents involving large-trucks: an exploratory analysis.” *Accident Analysis and Prevention* 37(5): 910-921.

King, W. D., N., M. H., W. E. Hardwick, and P. A. Palmisano. 1994. “Urban/rural differences in child passenger deaths.” *Pediatrict Emergency Care* 10(1): 34–36.

Kockelman, K. M., and Y. -J. Kweon. 2002. “Driver injury severity: an application of ordered probit models.” *Accident Analysis and Prevention* 34(3): 313–321.

Koppelman, F. S., and C. G. Wilmot. 1982. “Transferability analysis of disaggregate choice model.” *Transportation Research Record: Journal of Transportation Research Board* 895: 18–24.

Krantz, P. 1979. “Differences Between Single- and Multiple-automobile Fatal Accidents.” Accident Analysis and Prevention 11(3): 225 – 236.

Kostyniuk, L. P., F. M. Streff, and J. Zakrajsek. 2002. “Identifying Unsafe Driver Actions that Lead to Fatal Car-Truck Crashes.” AAA Foundation for Traffic Safety, 1440 New York Avenue, N.W., Suite 201, Washington, DC 20005.

Lee, C., and M. Abdel-Aty. 2005. “Comprehensive analysis of vehicle–pedestrian crashes at intersections in Florida.” *Accident Analysis and Prevention* 37(5): 775–786.

Lee, J., and F. L. Mannering. 2002. “Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis.” *Accident Analysis and Prevention* 34(2): 149–161.

Lee, L.-Fei. 1992. “On efficiency of method of simulated moments and maximum simulated maximum estimation of discrete response models.” *Econometric Theory* 8(3): 518–552.

Lemp, J. D., K. M. Kockelman, and A. Unnikrishnan. 2011. “Analysis of large truck crash severity using heteroskedastic ordered probit models.” *Accident Analysis and Prevention* 43(1): 370-380.

Lenne´, M. G., T. J. Triggs, and J. R. Redman. 1997. “Time of day variation in driving performance.” *Accident Analysis and Preventio*n 29(4): 431-437.

Levine, N. 2006. Houston, Texas, Metropolitan Traffic Safety Planning Program, *Transportation Research Record: Journal of the Transportation Research Board* 1969: 92–100.

Lundy, R A. 1956. “Effects of traffic volumes and number of lanes on freeway accident rates.” *Transportation Research Record: Journal of the Transportation Research Board* 99:138-156.

Langlois, P., M. H. Smolensky, B. P. Hsi, and F. W. Weir, F. 1985. Temporal patterns of reported single-vehicle car and truck accidents in Texas, USA, during 1980–1983. *Chronobiology International: Journal of Biological and Medical Rhythm Research* 2(2): 131–146.

Limdep, 2007. NLOGIT Version 4.0. Econometric Software, Planeview, NY.

Loeb, P. D., and W. A. Clarke. 2007. "The determinants of truck accidents." *Transportation Research Part E: Logistics and Transportation Review* 43(4): 442–452.

Lord, D., S. P. Washington, and J. N. Ivan. 2005. "Poisson, Poisson-gamma and Zero Inflated regression models of motor vehicle crashes: balancing statistical fit and theory." *Accident Analysis and Prevention* 37(1): 35–46.

Lord, D., S. R. Geedipally, B. N. Persaud, S. P. Washington, I. V. Schalkwyk, J. N. Ivan, C. Lyon, and T. Jonsson. 2008. Methodology for estimating the safety performance of multilane rural highways. *In: NCHRP Web-Only Document 126. National Cooperation Highway Research Program*, Transportation Research Board of the National Academies, Washington, D.C.

Lund, U. J. 2005. "The effect of seating location on the injury of properly restrained children in child safety seats." *Accident Analysis and Prevention* 37(3), 435–439.

Maddala, G. S. 1983. Limited-dependent and Qualitative Variables in Econometrics, NY: Cambridge University Press.

Malyshkina, N. V., and F. L. Mannering, and A. P. Tarko. 2009. "Markov switching negative binomial models: an application to vehicle accident frequencies." *Accident Analysis and Prevention* 34(2): 149–161.

Malyshkina, N. V., and F. L. Mannering. 2009. "Markov switching multinomial logit model: An application to accident-injury severities." *Accident Analysis and Prevention* 41(4): 829–838.

Malyshkina, N. V., and F. L. Mannering. 2010. "Empirical assessment of the impact of highway design exceptions on the frequency and severity of vehicle accidents Analysis." *Accident Analysis and Prevention* 42(1): 131-139.

Massie, D. L., K. L. Campbell, and A. F Williams, 1995. "Traffic accident involvement rates by driver age and gender." *Accident Analysis and Prevention* 2(1): 73–87.

McDonald, J. F., and R. A Moffitt. 1980. "The uses of tobit analysis." The *Review and Economics Statistics* 62(2): 318–321.

McFadden, D. 1981. Econometric models of probabilistic choice. In: Manski, D., McFadden (Eds.), A Structural Analysis of Discrete Data with Econometric Applications. The MIT Press, Cambridge, MA..

McFadden, D., and P. A. Ruud. 1994. "Estimation by simulation." *The Review of Economics and Statistics* 76(4): 591-608.

McFadden, D., and K. Train. 2000. "Mixed MNL models for discrete response." *Journal of Applied Econometrics* 15(5): 447-470.

McKelvey, R. D., and W. Zavoina. 1975. "Statistical model for analysis of ordinal level dependent variables." *Journal of Mathematical Sociology* 4(1): 103–120.

Mensah, A., and E. Hauer. 1998. "Two problems of averaging arising in the estimation of the relationship between accidents and traffic flow." *Transportation Research Record: Journal of the Transportation Research Board* 635: 37–43.

Martin, J-L, 2002. "Relationship between crash rate and hourly traffic flow on Interurban motorways." *Accident Analysis and Prevention* 34(5): 619–629.

Miaou, S.P., P.S. Hu, T. Wright, A.,K. Rathi, S.C. Davis. 1992. Relationships Between Truck Accidents and Highway Geometric Design: A Poisson Regression Approach. Presented at 71st Annual Transportation Research Board Meeting, Washington DC.

Miller, T. R. 1993. "Costs and Functional Consequences of U.S. Roadway Crashes." *Accident Analysis and Prevention* 25(5): 593–607.

Miller, T. R., D. T. Levy, R. S. Spicer, and D. C. Lestina. 1998. "Allocating the cost of motor vehicle crashes between vehicle types." *Transportation Research Record: Journal of the Transportation Research Board* 1635: 81–97.

Milton, J. C., V. N. Shanker, and F. L. Mannering, 2008. "Highway accident severities and the mixed logit model: an exploratory empirical analysis." *Accident Analysis and Prevention* 40(1): 260-266.

Milton, J. C., and F. L. Mannering. 1998. "The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies." *Journal of Transportation Engineering* 25(4): 395–413.

Moeltner, K., and D. F. Layton. 2002. "A Censored Random Coefficients Model for Pooled Survey Data with Application to the Estimation of Power Outage Costs." *The Review of Economics and Statistics* 84(3): 552–561.

Moore, D. N., W. H. Schneider, P. T. Savolainen, and M. Farzaneh, 2011. "Mixed logit analysis of bicyclist injury severity resulting from motor vehicle crashes at intersection and non-intersection locations." *Accident Analysis and Prevention* 43(3): 621-630.

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

Muelleman, R.L, and K. Mueller. 1996. "Fatal motor vehicle crashes: variations of crash characteristics within rural regions of different population densities." *Journal of Trauma* 41(2): 315–320.

National Automotive Sampling System (NASS) 2005. General Estimates System (GES). Analytical User's Manual 1988-2005. http://www-nrd.nhtsa.dot.gov/Pubs/AUM05.pdf. (accessed on September 10, 2010).

National Highway Traffic Safety Administration, 2009. Traffic Safety Facts: Motorcycles, DOT HS 811 159. http://www-nrd.nhtsa.dot.gov/Pubs/811172.pdf. (accessed on May 14, 2012).

National Highway Traffic Safety Administration. 2006. Traffic Safety Facts 2005: Older Population (DOT HS 810 622). Washington, D.C.

National Highway Traffic Safety Administration. Traffic Safety Facts 2010 Data, NHTSA's National Center for Statistics and Analysis (DOT HS 811 628). Washington, D.C., 2012.

National Institute for Occupational Safety and Health (NIOSH), 2007. Truck driver occupational safety and health. Conference report and selective literature review, Centers for Disease Control and Prevention, Department of Health and Human Services, 2003.

NHTSA 2008. Traffic Safety Facts 2007: Large Trucks, DOT HS 810 805, http://www-nrd.nhtsa.dot.gov/Pubs/810989.PDF. (accessed on October 1, 2010).

NHTSA 2009. Traffic Safety Facts 2009: Large Trucks, DOT HS 811388, http://www-nrd.nhtsa.dot.gov/Pubs/811388.pdf. (accessed on April 26,2011).

NHTSA, 1996. Rural and urban crashes—a comparative analysis. DOT HS-042 422. Research Note (August). US Department of Transpiration, National Highway Traffic Safety Administration, National Center for Statistics and Analysis (NCSA), 400 Seventh St. SW,Washington, DC 20590.

NHTSA, 2001. Traffic Safety Facts 2001—Rural/urban Comparison. DOT HS-809 524. US Department of Transpiration,National Highway Traffic Safety Administration, National Center for Statistics and Analysis (NCSA),Washington, DC.

NHTSA, 2009, Traffic Safety Facts, Rural/Urban Comparsion, DOT HS 811 395, http://www-nrd.nhtsa.dot.gov/pubs/811395.pdf. (accessed on May 10, 2012).

NHTSA, 2010, Traffic Safety Facts. http://www-nrd.nhtsa.dot.gov/Pubs/811659.pdf. (accessed on May 25, 2011).

NHTSA, 2012. Traffic Safety Facts, Research Notes, DOT HS 811 552, http://www-nrd.nhtsa.dot.gov/Pubs/811552.pdf. (accessed on July 14, 2012).

NHTSA. 2009 Traffic Fatalities by State and Percent Change from 2008 – State: USA. http://www-fars.nhtsa.dot.gov/States/StatesCrashesAndAllVictims.aspx. (accessed Mar 21, 2012).

NHTSA. Traffic Safety Facts, 2008 Data: State Traffic Data. DOT HS 811 168. http://www-nrd.nhtsa.dot.gov/Pubs/811168.pdf. (accessed Mar 05, 2012).

O'Donnell, C. J., and D. H. Connor. 1996. "Predicting the severity using models of ordered multiple choices." *Accident Analysis and Prevention* 28(6): 739–753.

Ostrom, M., and A. Eriksson. 1993. "Single-vehicle crashes and alcohol: a retrospective study of passenger car fatalities in Northern Sweden." *Accident Analysis and Prevention* 25(2): 171–176.

Otmani, S., J. Rog´e, and A. Muzet. 2005. "Sleepiness in professional drivers: Effect of age and time of day." *Accident Analysis and Prevention* 37(5): 930–937.

Pai, C.-W., and W. Saleh. 2008. “Modeling motorcyclist injury severity by various crash types at T-junctions in the UK.” *Journal of Safety Science* 46(8): 1234–1247.

Park, B.-J, D. Lord, and J. D. Hart. 2010. “Bias properties of Bayesian statistics in finite mixture of negative binomial regression models in crash data analysis.” *Accident Analysis and Prevention* 42(2): 741–749.

Parody, T.E. 1977. “An analysis of disaggregate mode choice models in prediction.” *Transportation Research Record: Journal of the Transportation Research Board* 649: 23–31.

Patil, S, S. R. Geedipally, and D. Lord. 2012. “Analysis of crash severities using nested logit model—Accounting for the underreporting of crashes.” *Accident Analysis and Prevention* 45(1): 646– 653.

Persaud, B. N., and K. Mucsi. 1995. “Microscopic accident potential models for two-lane rural roads.” *Transportation Research Record: Journal of the Transportation Research Board* 1485: 134–139.

Poch, M., and F. L. Mannering. 1996. “Negative Binomial Analysis of Intersection Accident Frequencies.” *Journal of Transportation Engineering* 122(2): 105–113.

Qin, X., J. N. Ivan, and N. Ravishanker. 2004. “Selecting exposure measures in crash rate prediction for two-lane highway segments.” *Accident Analysis and Prevention*, 36(2): 183–191.

Qin, X., J. N. Ivan, N. Ravishanker, J. Liu, and D. Tepas. 2006. “Bayesian estimation of hourly exposure functions by crash type and time of day.” Accident *Analysis and Prevention* 38(6): 1071–1080.

Quddus, M. A., R. B. Noland, and H. C. Chin. 2002. “An analysis of motorcycle injury and vehicle damage using ordered probit models.” *Journal of Safety Research* 33(4): 445–462.

Rakauskas, M. E., N. J Ward, and S. G. Gerberich, 2009. “Identification of differences between rural and urban safety cultures.” *Accident Analysis and Prevention* 41(5): 931–937.

Revelt, D, and K. Train, 1999. Customer specific taste parameters and mixed logit. Working Paper, University of California, Berkeley, Department of Economics,.

Revelt, D., and K. Train. 1997. “Mixed logit with repeated choices: households’ choice of appliance efficiency level.” *The Review of Economics and Statistics* 80(4): 647-657.

Roncek, D. W. 1992. “Learning more from tobit coefficients: extending a comparative analysis of political protest.” *American Sociological Review* 57(4): 503–507.

Savolainen, P. T., and A. P. Tarko. 2005. “Safety Impacts at Intersections on Curved Segments.” *Transportation Research Record: Journal of the Transportation Research Board* 1908: 130–140.

Savolainen, P. T., F. L. Mannering, D. Lord, and M. A. Quddus. 2011. “The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives.” *Accident Analysis and Prevention* 43(5): 1666– 1676.

Savolainen, P. T., and F. L. Mannering. 2007. “Probabilistic models of motorcyclists’ injury severities in single- and multi-vehicle crashes.” *Accident Analysis and Prevention* 39(5): 955–963.

Shankar, V. N., and F.L. Mannering 1996. “An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity,” *Journal of Safety Research* 27(3) :183–194.

Shankar, V. N., F. L. Mannering, and W. Barfield. 1995. “Effect of roadway geometrics and environmental Factors on rural accident frequencies.” *Accident Analysis and Prevention* 27(3): 371–389.

Shankar, V. N, J. Milton, and F. L Mannering. 1997. “Modeling accident frequencies as zero-altered probability processes: an empirical inquiry.” *Accident Analysis and Prevention* 29(6): 829–837.

Shankar, V. N., A. B. Albin, J. Milton, and F. L. Mannering.1998. “Evaluating median crossover likelihoods with clustered accident counts: an empirical inquiry using the random effects negative binomial model.” *Transportation Research Record: Journal of the Transportation Research Board* 1635: 44–48.

Smith, K., and J. O’Day. 1982. “Simpson's paradox: An example using accident data from the state of Texas.” *Accident Analysis and Prevention* 14(2): 131-133.

Stamatiadis, N., and J. A. Deacon.1995. “Trends in Highway Safety: Effects of an aging population on accident propensity.” *Accident Analysis and Prevention* 27(4): 443–459.
Statistical Analysis System, SAS® 9.2 for Windows. 2011. Cary, NC: SAS Institute, Inc.

Stern, S. 1997. “Simulation-based estimation.” *Journal of Economic Literature* 35(4): 2006-2039.

Svenson, J.E., C. Spurlock, , and M. Nypaver, , 1996. “Factors associated with the higher traumatic death rate among rural children.” *Annual Emergency Medicine* 27(5): 625–632.

Texas Motor Vehicle Crash Statistics, 2011. Texas Motor Vehicle Traffic Crash Highlights. http://ftp.dot.state.tx.us/pub/txdot-info/trf/crash_statistics/2011/01_2011.pdf. (accessed on March 23, 2011).

Tobin, J. 1958. “Estimation of relationships for limited dependent variables.” *Econometrica* 26(1): 24–36.

Train, K. 1997. Mixed logit models for recreation demand. In: Kling, C., Herriges, J. (Eds.), Valuing the Environment Using Recreation Demand Models. Elgar Press, NY.

Train, K. 1999. Halton sequences for mixed logit, Working Paper, University of California Berkley, Department of Economics.

Train, K. 2009. Discrete choice methods with simulation. Cambridge University Press, New York, NY.

Train, K. 2003. Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge, UK..

U.S. Department of Transportation (USDOT). 2005. Report to Congress on the Large Truck Crash Causation Study. Federal Motor Carrier Safety Administration, U.S. Department of Transportation.

Ulfarsson, G. F., and F. L. Mannering. 2004. "Differences in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger car accidents." *Accident Analysis and Prevention* 36(2): 135–147.

United Nations, 2010. Resolution of the United Nations General Assembly, A/RES/64/255, 2010.http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/64/255. (accessed on March 23, 2011).

USDOT/BTS, 2008. 2007 Commodity Flow Survey, http://www.bts.gov/publications/ commodity flow survey/preliminary_tables_december 2008/index.html. (accessed on November 01, 2010).

Veall, M. R., and K. F. Zimmermann. 1996. "Pseudo-R2 Measures for some common limited dependent variable models." *Journal of Economic Surveys* 10(3): 241-259.

Walker, J., and M. E. Ben-Akiva, 2002. "Generalized random utility model." *Mathematical Social Sciences* 43(3): 303–343.

Wang, S.C. 1998. An Aging Population: Fragile, Handle with Care.. http://www-nrd.nhtsa.dot.gov/departments/nrd-50/ciren/um_fragile.html.Washington, D.C., National Highway Traffic Safety Administration. (accessed June 20, 2012).

Washington, S. P., M. G. Karlaftis, and F. L. Mannering. 2011. Statistical and Econometric Methods for Transportation Data Analysis, 2nd Ed. Chapman & Hall/CRC, Boca Raton, FL.

Weiss, S. J., R. Ellis, A. Ernst, R. F. Land, and A. Garza. 2001. "A comparison of rural and urban ambulance crashes." *The American Journal of Emergency Medicine* 19(1): 52–56.

Wylie M. A. T., and K. B. Richard. 1997. "Rural motor vehicle crashes: can we prevent them?" *Journal of Emergency Nursing* 23(6):542–544.

Xie, Y., K. Zhao, and N. Huynh. 2012. "Analysis of driver injury severity in rural single-vehicle crashes," *Accident Analysis and Prevention* 47: 36–44.

Xie, Y., Y. Zhang, and F. Liang 2009. "Crash Injury Severity Analysis Using Bayesian Ordered Probit Models," *Journal of Transportation Engineering* 135(1):18-25.

Yamamoto, T., and V. N. Shankar. 2004. "Bivariate ordered-response probit model of driver's and passenger's injury severities in collisions with fixed objects." *Accident Analysis and Prevention* 36(5): 869–876.

Yau, K. K. W., and H. P. Lo, and H. H. F. Sherrice. 2006. "Multiple-vehicle traffic accidents in Hong Kong." *Accident Analysis and Prevention* 38(6): 1157–1161.

Ye, F., and D. Lord. 2011. Comparing three commonly used crash severity models on sample size requirements: multinomial logit, ordered probit and mixed logit models, 90th Annual Meeting of the Transportation Research Board (Paper #11-1988), Washington D.C.

Ye, F., and D. Lord. 2011. "Investigating the effects of underreporting of crash data on three commonly used traffic crash severity models: multinomial logit, ordered probit and mixed logit models." 90th Annual Meeting of Transportation Research Board, Washington, DC.

Zajac, S. S., and J. N. Ivan. 2002. "Factors influencing injury severity of motor vehicle-crossing pedestrian crashes in rural Connecticut." *Accident Analysis and Prevention* 35(3): 369–379.

Zaloshnja, E., and T. R. Miller. 2004. "Costs of large truck-involved crashes in the united states." *Accident Analysis and Prevention* 36(5): 801-808.

Zaloshnja, E., and T. R. Miller. 2006. "Unit Costs of Medium and Heavy Truck Crashes." Final Report for Federal Motor Carrier Safety Administration and Federal Highway Administration, 400 Seventh Street, SW, Washington, DC.

Zhu, X., and S. Sivaramakrishnan. 2011. "A Comprehensive Analysis of Factors Influencing the Injury Severity of Large-Truck Crashes." *Accident Analysis and Prevention* 43(1): 49-57.

Zlatoper, T. J. 1989. "Models explaining motor vehicle death rates in the United States." *Accident Analysis and Prevention* 21(2):125–154.

Zwerling, C., C. Peek-Asa, P. S. Whitten, S-W Choi, N.L. Sprince, and M.P. Jones. 2005. "Fatal motor vehicle crashes in rural and urban areas: decomposing rates into contributing factors." *Journal of Injury Prevention* 11(1): 24-28.

Appendix – A

Chapter 10: An Application of A Random Parameters Ordered Probit model for Large-truck involved Crashes

10.1 Introduction

Trucking is a vital component of any prospering and growing economy. It is the backbone of many logistic and supply chain systems. Growing concerns related to large truck (Gross Vehicle Weight Rating (GVWR) greater than 10,000 pounds) crashes have increased in recent years due to the potential level of injury severity and economic impact. Recent statistical data indicate that large trucks have a higher rate of crash involvement than passenger vehicles in the United States (US) controlling for the number of registered vehicles and vehicle-miles traveled (VMT) (FHWA 2008; NHTSA 2008). Although large trucks accounted for four percent of registered vehicles and eight percent of VMT in 2008, eleven percent of motor vehicle deaths in 2008 were a result of large truck crashes (FHWA, 2008). Large trucks heavily impact the national economy through daily freight movements. However, large-truck involved crashes also impact the level of injury severity of collision partners and incur high societal cost associated with fatalities, injuries and property damages.

The cost associated with these large-truck involved crashes is of great concern and can be substantial. Based on 2005 dollars, the estimated cost of a police-reported crash involving a large truck, considering all truck-tractor with trailer configurations, averaged $91,112 (Zaloshnja and Miller, 2006). In addition, this study estimated the average cost per fatality, non-fatality, and property-damage-only as $3,604,518, $195,258, and $15,114, respectively. An earlier study by Zaloshnja and Miller (2004) estimated the cost associated with different configurations of large trucks. According to this study

(Zaloshnja and Miller, 2004), of all configurations of truck-tractor carrying different number of trailers and multiple combination trucks (i.e., large truck carrying two or three trailers) had the highest cost at $88,483 per crash based on 2000 dollars. The crash costs based on 2000 dollars per 1,000 truck miles were $157 for single unit trucks, $131 for single combination trucks, and $63 for multiple combinations (Zaloshnja and Miller 2004). These costs illustrate the potential monetary impacts large truck crashes have on society. Hence, any increase in the number or severity of these types of crashes is of great concern to organizations that operate, maintain, and construct the transportation system as well as to trucking companies.

With this in mind, this study aims to analyze the injury severity of large-truck involved crashes through an econometric modeling approach. We utilize the random parameters ordered probit model to shed light on the factors contributing to large truck crashes. To achieve this, the fusion of three datasets from the National Automotive Sampling System General Estimates System (NASS-GES) crash database was used. Through this fused dataset we hope to provide an improved understanding of the complex interactions between contributing factors influencing large truck crash results. The three fused datasets pertain to human, vehicle, and road-environment factors. To capture these complexities using the NASS-GES database, consideration of random parameters provides a mechanism to account for any unobserved heterogeneity. To the best of the authors' knowledge this study is one of the first attempts at modeling large truck injury severity focusing on the US interstate system by utilizing the NASS-GES dataset. Although the random parameters ordered probit model has been applied to large-truck crash severity from different modeling perspectives (i.e., these studies use limited datasets and don't consider exclusively the potential effect of unobserved factors on crash severity outcomes, see following section), this research extends the current literature and introduces additional significant variables related to human factors in regards to large-truck crashes.

10.2 Literature review

In this section we present a synthesis of previous research with special attention given to the methodological approaches that establish any links between crash characteristics and injury severity with human, vehicle, and road environment factors.

A recent study by Lemp et al. (2011) introduced a Heteroskedastic Ordered Probit (HOP) model to analyze the injury severity in crashes involving at least one large truck. The analysis of large-trucks in their study was limited to long-combination trucks (LVCs) of two or more trailers with a GVWR of 36,287 kg (80,000 lb). They utilized a HOP model specification for greater flexibility in parameter estimation over the standard Ordered Probit (OP) model to address the issue of heteroskedasticity (i.e., non-constant variance) across the observations. Their model results indicated that given an injurious large truck crash, the likelihood of an injury of lesser severity was greatly increased when a crash occurred at curved sections of a roadway, a crash occurred on a roadway sag, if the trucks were overweight, if a driver was under the influence of illegal drugs, and/or if a driver exhibited aggressive driving behavior other than speeding. Their study was limited to data provided from the Large Truck Crash Causation Study (LTCCS) collected between 2001 and 2003. Moreover, their modeling approach did not explicitly account for any unobserved factors (i.e., factors not captured in the dataset, but that may be contributing to the injury severity outcomes).

Zhu and Srinivasan (2011) applied an ordered probit model using the data from large-truck involved crashes. The aim of the study was to address the need for research regarding large-truck safety in addition to contributions to transportation policy, improvement of motor carrier operations, and incident-cost reductions. The study found that head on collisions and collisions at intersections were the most serious, whereas crashes on multi-lane highways were less severe. Their estimation results indicated that particular attention must be paid to truck corridors near major tourist spots because of network (route) unfamiliarity issues of passenger car drivers. That is, distracted driving, driving under

the influence or emotional distraction increased the likelihood of severe crashes. Other factors such as truck-driver fatigue, aggression, and seat belt usage turned out to be statistically insignificant. This study was based on a small dataset extracted from the LTCCS study.

As with our work, Chistoforou et al. (2010) applied a random parameter ordered probit framework to large truck crashes, but it was applied to road users to address the challenge of external cost estimation and roadway safety. The data they utilized for their analysis period is not continuous (2000-2002 and then 2006 were considered). This discontinuity in the analysis period could lead to estimation errors. That is, some observed and unobserved factors may vary from year-to-year. For example, weather may vary from year-to-year, geometry (e.g., widening shoulder or median or installing roadside barrier), or policy related factors (e.g., change in posted ‘Speed Limit’). These variations could lead to variations in the injury severities that are experienced from year-to-year and could lead to erroneous estimates (Tarko et al., 2011).

Duncan et al. (1998) applied an ordered probit model for rear-end collision between truck and passenger cars was estimated to better understand the complex interaction of vehicles crash dynamics where each observation of the developed model refers to the vehicular occupants. The authors found when a rear end crash outcome occurs between a passenger car and truck, factors related to darkness, high speed differentials, high speed limits, grades (especially when wet), being in a car struck to the rear, driving while drunk, and being female highly increase the likelihood of injury severity for passenger vehicle occupants. The study also found car rollovers and a situation where a car is rear-ended by a truck at a high speed differential and to be significant. Driving on snowy or icy roads, congested roads, the use of a child restraint in passenger vehicles, and being in a rear-ended station wagon instead of a sedan are likely to reduce the severity outcome of the crashes. This study was limited to the inadequate traffic data on certain segments of highway at the time of crash. That is, unavailable road inventory data, such as the number of lanes and traffic flow data, not present in the crash dataset could

lead to unobserved factors influencing injury severity outcomes, thus leading to biased estimates of the parameters. For instance, number of directional lanes indirectly related to the ease or tendency of lane changing behavior of the passenger and heavy vehicle drivers. Similarly, the traffic flow is related to time of day, types of crashes such as single- and multi-vehicle (Qin et al., 2006) as well as speed and injury severities of the involved vehicles. To limit biased estimates, a comprehensive database should be considered. In this study we consider the GES system because it is a nationally representative sampled crash database.

From a logit based approach, Chang and Mannering (1999) applied a nested logit structure to model injury severity based on vehicle occupancy in terms of exposure effects to address the issue of severe injury caused by large trucks (i.e., nesting structure was based on vehicle occupancy). Their results indicate that the effects of trucks on crash injury severity are greater for multi-occupant vehicles over single-occupant vehicles, even though the multi-occupant vehicle crash dataset was limited in the number of observations. The authors considered non-incapacitating injury, incapacitating injury, and fatal into one single group which could be separated and addressed in different modeling approaches such as ordered probit models where all five categories – fatal, incapacitating, non-incapacitating, possible injury and property-damage-only (no-injury) could be addressed. However, considering disaggregate injury outcomes (i.e., considering non-incapacitating injury, incapacitating injury, and fatal separately rather than combined) could be challenging in the model estimation which depends on data regarding those injury outcomes in the sample.

Khorashadi et al. (2005) investigated injury severity through the use of a multinomial logit model to address large-truck crashes and fatalities, and the limitations of other studies regarding crash frequencies. Although not incorporated in their study, they identify the urgency of additional research regarding human factors such as perceptual, cognitive and response demands of drivers to explain the

complex interaction of factors in crashes. Their study was focused on the exclusive use of crash data from California.

A mixed logit framework was implemented by Milton et al. (2008) to capture injury severities of crashes involving all vehicles, which was not explicitly focusing as large-truck crashes, on roadway segments to address the challenges of methodological approaches related to count models. Their model results indicate that the Average Daily Traffic (ADT) per lane (in thousands of vehicles) for slightly less than half of the roadway segments in the sample result in a decrease in PDO crashes which implies an increase in more severe injury outcomes. They found that increasing pavement friction decreases the likelihood of possible injury and increases the likelihood of PDO and injury; whereas a greater snowfall increases the likelihood of PDO and consequently decreases the likelihood of more severe injury outcomes. The number of horizontal curves and the number of grade breaks, both used as a fixed parameter, and the number of interchange per mile in the roadway segment reduces the likelihood of injury crashes. Their work was limited to an aggregated dataset to avoid missing event specific information.

The current literature related to the modeling of large truck crash severity is very wide and extensive in nature. However, various aspects of the problem have not been addressed with particular emphasis, for instance, on human factors focusing on pre-crash driving behaviors over US interstate system. It is evident through the presented works that data sample age and size is an issue. Our research differentiates itself from the presented works through the use of a comprehensive nationally representative sampled crash database which incorporates human related factors, vehicle factors, and road environment variables. We consider additional human related factors which are not only limited to the driving behavior in the pre-crash phase, but also include drivers' demographics into the modeling framework. Furthermore, we utilize a large database to minimize the effects of unaccounted for factors

dealing with human behavior. Also, we define large-trucks to be greater than 4,536 kg (10,000 lb) as defined by the Insurance Institute for Highway Safety (IIHS).

10.3 Empirical settings

The data for large truck crashes was collected from the National Automotive Sampling System General Estimated System (NASS-GES) crash database maintained by National Highway Traffic Safety Administration (NHTSA). A large truck, as defined by the IIHS and for this study, can be classified as a tractor-trailer, single-unit truck, or cargo van having Gross Vehicle Weight Rating (GVWR) greater than 4,536 kg (10,000 lb). According to Analytical Users' Manual (NASS-GES, 2008) the GES database is based on a nationally representative probability sample selected from an estimated 5.8 million police-reported crashes which occur annually resulting in a fatality or injury and those involving major property damage.

According to a technical report by NHTSA (2009), there are 25% of minor injury crashes and half no injury crashes being unreported (Savolainen et al., 2011). In this study, we considered the GES sample data over a period of four years from 2005 to 2008 for large-truck involved crashes. Despite the issues of under reporting for minor and no personal injury along with the multi-stage sampling scheme in GES database, the GES focuses on crashes of greatest concern to the highway safety community and the general public (NASS-GES, 2008).

To investigate human, vehicle, and road environment factors, a sample of 8,291 data observations is used in where each observation is a crash representing the most severely injured occupants (i.e., the worst injury level) involving at least a large truck on the interstate system from 2005 to 2008. The non-truck vehicles are broadly classified as passenger cars and their derivatives and light truck (having GVWR less than 4,536 kg (10,000 lb)) comprises about 71% and 21% of the vehicle involved with large trucks in this sample, respectively. The statistics clearly indicate that quite a large

number of passenger vehicles are involved in large truck crashes. This truck involved data sample (i.e., 8,291 observations) was extracted from the GES crash dataset with an average of 56,970 crashes (i.e., truck and non-truck involved crashes) reported each year over time period from 2005 to 2008. The crash dataset was fused to vehicle and person dataset through the appropriate linking variable and crash number, while vehicle and person datasets were linked through vehicle and crash number using the Statistical Analysis System (SAS) (SAS, 2011). The random parameter ordered probit modeling framework was modeled in Limdep (NLOGIT 4.0) (Greene, 2007).

Descriptive statistics of key variables used in the model (i.e., all independent variables) are presented in Table 10.1. The dependent variable has five levels of injury categories – fatality (K), incapacitating (A), non-incapacitating (B), possible (C) injuries, and non-injury (O) or property-damage-only (PDO) which represent 56 (0.06%), 258 (3.1%), 527 (6.3%), 593 (7.1%), and 6857 (82.7%) of the sample size considered in this study, respectively.

Human factors cover occupant's demographics, driving behavior, restraint usage, and driving or living area vicinity. Turning to Table 10.1 demographics, "males" make up about 93.8% of the sample and age between 55 and 65 accounts for about 12.7% of the sample. Speed as a contributing factor in the crash identified through the investigation process accounts for 14.4% of the sample. Restraint usage (i.e., lap and shoulder belt) accounts for around 82.2% of the sample. Drivers residing in the state of Texas account for about 10.2% of the sample. Turning to vehicular characteristics, trucks carrying single trailing unit account for about 75.2% of sample. On average, there are two vehicles involved in the crashes in this study. Vehicular role being passive in crashes (i.e., being struck by another vehicle in the crash) accounts for around 37.1% of the sample. Then, road and environmental characteristics such as curved sections of the highways account for about 13.5% of the sample. Lighting condition at the time of crash such as dark condition represents approximately 13.4% of the sample. Temporal characteristics such as weekend (i.e., Saturdays and Sundays) and summer months (i.e., June to August) account for

roughly 14.5%, and 23.9% of the sample, respectively. Finally, crash mechanism such as rollover, departing the roadway, sideswipe in the same direction, lane changing, and going straight in the lane accounts for about 80.2%, 13.5%, 26.3%, 9.8%, and 64.9% of the sample, respectively.

The correlation matrix for the ordered probit model was performed and indicated none of variables of interest have a correlation value of more than ∓0.50. The correlation matrix shows the maximum correlation between departing roadway and number of vehicles involved is -0.477. Similarly, the crash mechanism such as lane changing and going straight shows a correlation of -0.450. However, these two situations show logical relationship in terms of their signs (i.e., they are negatively correlated). This is possibly due to a single vehicle running off the road rather than multiple vehicles. Similarly, lane changing and going straight (i.e., lane keeping) deal with dissimilar characteristics with respect to driving maneuvers and in reality may not indicate any degree of multi-collinearity.

Table 10.1: Descriptive Statistics of Key Variables in the Model

Variable	Meaning of Variables in the Model	Mean	Std. Dev.
CURVE	Alignment of highway section (1 for curved section, 0 otherwise)	0.1353	0.3421
WEEKEND	Day of the week (1 if weekend, 0 otherwise)	0.1455	0.3526
SUMMER	Months of the year (1 if summer months (June - August), 0 otherwise)	0.2387	0.4263
DARK	Light condition of street (1 if dark, 0 otherwise)	0.1341	0.3408
VEH_INVL	Number of vehicles involved in the crash	2.0526	0.8034
TRAIL1	Trailing unit when (1 if one trailer, 0 otherwise)	0.7523	0.4317
PASSIVE	Vehicle role (1 if struck by other vehicle, 0 otherwise)	0.3711	0.4831
RLOVER	The most harmful event (1 if rollover, 0 otherwise)	0.8019	0.3985
LRRDDEP	Vehicle maneuver during pre-crash situation (1 if left or right side departure, 0 otherwise)	0.1354	0.3422
SSWIPESD	Orientation of vehicle at the time of crash (1 if sideswipe in the same direction, 0 otherwise)	0.2632	0.4404
LANECHNG	Vehicle maneuver just prior to impending crash (1 if changing lane, 0 otherwise)	0.0988	0.2984
GOSTRGHT	Vehicle maneuver just prior to impending crash (1 if going straight, 0 otherwise)	0.6490	0.4773
SPEEDFAC	Factor of crash identified in the investigation (1 if speed, 0 otherwise)	0.1438	0.351

Variable	Meaning of Variables in the Model	Mean	Std. Dev.
TEXAS1	Drivers' working/residing place according to license record (1 if Texas, 0 otherwise)	0.1019	0.3026
LAPSHLD	Occupants' use of available vehicle restraints (1 if lap and shoulder belt used, 0 otherwise)	0.8229	0.3817
MALE	Gender of the occupants (1 if male, 0 otherwise)	0.9388	0.2396
AGE5565	Age of occupants (1 if age group is 55 to 65, 0 otherwise)	0.1271	0.3331

10.4 Methodology

To obtain a better understanding of the factors associated with large-truck involved crashes, we propose a random parameters ordered probit modeling approach to capture the injury severity experienced to accommodate unaccounted for heterogeneity (Zhu and Srinivasan 2011; Chistoforou et al. 2010; McKelvey and Zavoina 1975). Since the level of injury is ordinal in nature, the KABCO scale ('K' for Fatal, 'A' for Incapacitating injury, 'B' for Non-incapacitating Injury, 'C' for Possible Injury and 'O' for Property-Damage-Only) is followed. This is done to account for any under-reporting tendency in the crash data reporting system, hence descending order of injury severity level 0 for K (Fatal), 1 for A (Incapacitating Injury), 2 for B (Non-incapacitating injury), 3 for C (Possible injury) and 4 for O (Property-Damage-Only) is considered in the random parameters ordered probit methodology presented in this study. Ye and Lord (2011) show that formulating in descending order of injury severity of (KABCO) rather than ascending order (OCBAK) reduces the bias and variability in the estimation of the parameters for ordered probit model which has been tested under different simulation scenarios.

With this in mind, the ordered probit models has been widely applied to model the marginal probability effects of several contributory factors on injury severity by considering '0' for no injury/PDO, '1' for possible injuries, '2' for non-incapacitating injury, '3' for incapacitating injury, and '4' for fatality (Chistoforou et al. 2010; Abdel-Aty 2003; Gray et al. 2008; Kockelman and Kweon 2002; Lee and Abdel-Aty 2005; O'Donnell and Connor 1996; Pai and Saleh; Quddus et al. 2002; Xie et al.

2009; Zajac and Ivan 2002). However, in this study we model the level of injuries as five levels of ordinal categories of the dependent variables and is as follows: '0' for Fatal, '1' for Incapacitating Injury, '2' for Non-incapacitating Injury, '3' for Possible Injury and '4' for Property-Damage-Only .

We start formulating the model by defining an unobserved variable y^* as a modeling basis of ordinal ranking of the data, with y^* specified as a latent and continuous measure of injury severity of each observation (Washington et al., 2011):

$$y^* = \boldsymbol{\beta X} + \varepsilon \tag{10.1}$$

where:

y^* : is the dependent variable (specified as a latent and continuous measure of injury severity of each observation n),

$\boldsymbol{\beta}$: is a vector of estimable parameters,

$\boldsymbol{X}$: is a vector of explanatory variables (e.g., human, roadway segment, vehicle, and crash mechanism characteristics),

ε : is a random error term (assumed to be normally distributed with 0 mean and a variance of 1).

Using Equation 1, and under the order probit framework the observed ordinal data y (e.g., injury severity) for each observation can be represented as (Washington et al., 2011):

$$\begin{aligned}
&y = 0 \quad if -\infty \leq y^* \leq \mu_0 \\
&y = 1 \quad if\ \mu_0 \leq y^* < \mu_1 \\
&y = 2 \quad if\ \mu_1 \leq y^* < \mu_2 \\
&y = \cdots \\
&y = I - 1 \quad if\, \mu_{I-2} \leq\ y^* <\ \mu_{I-1} \\
&y = I \quad if\, \mu_{I-1} \leq\ y^* <\ \infty
\end{aligned} \tag{10.2}$$

where:

μ : are estimable parameters or thresholds between two adjacent injury categories that define y and are estimated jointly with the model parameters $\boldsymbol{\beta}$, which corresponds to integer ordering, and I is the highest integer ordered response (e.g., PDO this is 4).

To estimate the probabilities of I specific ordered response for each observation n, ε is assumed to be normally distributed with 0 mean and variance 1. The ordered probit model with ordered selection probabilities is defined as follows:

$$
\begin{aligned}
&P_n(y=0) = \Phi(-\boldsymbol{\beta X}) \\
&P_n(y=1) = \Phi(\mu_1 - \boldsymbol{\beta X}) - \Phi(-\boldsymbol{\beta X}) \\
&P_n(y=2) = \Phi(\mu_2 - \boldsymbol{\beta X}) - \Phi(\mu_1 - \boldsymbol{\beta X}) \\
&\ldots \\
&P_n(y=I) = 1 - \Phi(\mu_{I-1} - \boldsymbol{\beta X})
\end{aligned}
\tag{10.3}
$$

where:

$P_n(y=I)$: is the probability that observation n has I as the highest ordered response index (for instance, injury outcomes, in our case PDO), given a crash occurred

$\Phi(\cdot)$: is the standard normal cumulative distribution function

Marginal effects are computed at the sample mean for each category (Washington et al. 2011, Greene 1997):

$$\frac{P_n(y=I)}{\partial \boldsymbol{X}} = [\phi(\mu_{I-2} - \boldsymbol{\beta X}) - \phi(\mu_{I-1} - \boldsymbol{\beta X})]\boldsymbol{\beta} \tag{10.4}$$

where:

$\phi(\cdot)$: is the probability mass function of the standard normal distribution

To accommodate any unaccounted for factors that may vary across observations, this study extends the standard ordered probit model to account for random parameters (Chistoforou et al. 2010; Train 1997; Revelt and Train 1997; Brownstone and Train 1999; McFadden and Train 2000; Bhat 2001;

Eluru et al. 2007; Anastasopoulos and Mannering 2009; Anastasopoulos et al. 2009a; Anastasopoulos et al. 2009b). The inclusion of random parameters provides a mechanism to preclude inconsistent, inefficient and biased parameter estimates (Washington et al., 2011). The fixed parameters ordered probit model would lead to biases in the results while neglecting the unobserved heterogeneity as shown in (Chistoforou et al., 2010). This would be the case since: a) some assumptions regarding the standard probit model limits its applicability, b) the marginal probability effects change their sign while moving from the smallest to the largest outcome, and c) possible unaccounted for factors among observations are not properly addressed.

Greene (2007) developed an estimation procedure that utilizes simulated maximum likelihood estimation to incorporate random parameters in the ordered probit modeling scheme (Greene, 2007). The random parameter ordered probit model is formulated by taking into account of an error term being correlated with the unobserved factors in ε_i (as shown in Equation 4.1) which translates the individual heterogeneity into parameter heterogeneity as follows (Greene, 1997):

$$\beta_{in} = \beta + \gamma_{in} \tag{10.5}$$

where:

γ_{in} : is randomly distributed term (for example a normally distributed term with mean 0 and variance σ^2).

Estimation of the random parameters model is done using a Halton sequence approach (Milton et al. 2008; Anastasopoulos and Mannering 2009; Halton 1960; Train 1997; Bhat 2003). 200 Halton draws are used to estimate the parameters that maximize the simulated log-likelihood function and normal, triangular and uniform distributions are considered for the functional form of the parameter density function (Chistoforou et al. 2010).

10.5 Empirical results and discussions

Fixed- and random-parameters ordered probit models are estimated using maximum likelihood and simulation-based maximum likelihood methods for parameter vector $\boldsymbol{\beta}$, respectively. With regard to the distribution of the random parameters in our analysis, consideration was given to the normal, lognormal, triangular, and uniform distributions. Only the normal distribution was found to be statistically significant. Two hundred Halton draws were used for the simulation-based maximum likelihood estimate. This number of draws has been empirically shown to produce accurate parameter estimates (Milton et al. 2008; Bhat 2003).

The estimated variables in both models were found to be statistically significant within a 95% confidence level. A likelihood ratio test comparing the fixed and random parameters ordered probit models was performed to test the null hypothesis that the fixed parameter model is statistically equivalent to the random parameters model and the procedure is as follows (Washington et al. 2011):

$$\chi^2 = -2[LL_{FIX}(\beta^{FIX}) - LL_{RAN}(\beta^{RAN})] \tag{10.6}$$

where:

$LL_{FIX}(\beta^{FIX})$:is the log-likelihood at convergence of the fixed parameters model (-4933.841)

$LL_{RAN}(\beta^{RAN})$:is the log-likelihood at convergence of the random parameters model (-4908.552)

The Chi-square statistic for the likelihood ratio test with six degrees of freedom gave a value greater than the 99.99% ($\chi^2 = 55.578$) confidence interval based on two-tailed p-value, indicating that the random parameter model is statistically superior to the corresponding fixed parameter model. This means that the null hypothesis, that the random parameters estimated model is no better than the fixed model comparison model, is rejected. Table 10.2 and 10.3 present the details of the fixed- and random-parameters models and marginal effects of the random parameters model, respectively.

The marginal effects illustrated in Table 10.3 provide additional information regarding what occurs with interior injury severity category, their corresponding probabilities as well as the magnitude of change across these categories. A negative value represents a decreased impact on injury severity probabilities. For example, the variable indicating alignment of highway section (1 for curved section, 0 otherwise) for PDO (Y=4) the negative sign (-0.023) indicates that on average the probability of severe injuries is higher given crashes occurring on curved sections. In contrast, the other categories are positive and on average their probabilities are lower.

Table 10.2: Large-truck involved Injury Severity Model Results

	Fixed Parameters Model		Random Parameters Model	
Variable	**Coeff.**	***t*-stat**	**Coeff.**	***t*-stat**
Constant (*Std. Dev*)	1.512	13.653	1.425 (0.207)	10.126 (11.036)
Alignment of highway section (1 for curved section, 0 otherwise)	-0.149	-2.795	-0.167	-2.761
Day of the week (1 if weekend, 0 otherwise)	0.102	2.107	0.150	2.763
Months of the year (1 if summer months (June - August), 0 otherwise)	-0.171	-4.543	-0.201	-4.706
Light condition of street (1 if dark, 0 otherwise) (*Std. Dev*)	-0.137	-2.857	0.227 (1.084)	3.433 (17.974)
Number of vehicles involved in the crash (*Std. Dev*)	0.050	2.175	0.341 (0.333)	12.068 (30.146)
Trailing unit when the crash occurred(1 if one trailer, 0 otherwise) (*Std. Dev*)	0.336	9.220	0.552 (0.501)	13.183 (20.529)
Vehicle role (1 if struck by other vehicle, 0 otherwise)	0.469	10.712	0.621	11.912
The most harmful event (1 if rollover, 0 otherwise)	0.412	10.345	0.578	13.017
Vehicle maneuver during pre-crash situation (1 if left or right side departure, 0 otherwise) (*Std. Dev*)	-0.481	-9.626	-0.441 (0.473)	-8.227 (12.317)
Orientation of vehicle at the time of crash (1 if sideswipe in the same direction, 0 otherwise)	0.357	7.110	0.480	8.158
Vehicle maneuver just prior to impending crash (1 if changing lane, 0 otherwise) (*Std. Dev*)	0.232	2.942	0.658 (0.816)	6.114 (10.138)
Vehicle maneuver just prior to impending crash (1 if going straight, 0 otherwise)	0.116	2.600	0.168	3.230
Factor of crash identified in the investigation (1 if speed, 0 otherwise)	-0.098	-2.179	-0.094	-1.905
Occupants' use of available vehicle restraints (1 if lap and shoulder restraint used, 0 otherwise)	0.305	7.531	0.401	8.809
Age of the occupants (1 if for age group of 55-65 years, 0 otherwise)	-0.114	-2.299	-0.161	-2.836
Gender of the occupants (1 if male, 0 otherwise)	0.182	2.790	0.254	3.455
Drivers' working/residing place according to license record (1 if Texas, 0 otherwise)	-0.140	-2.698	-0.202	-3.296

Threshold 1, μ1	0.772	27.782	1.1062	15.992
Threshold 2, μ2	1.350	59.431	1.906	26.086
Threshold 3, μ3	1.738	73.697	2.439	32.790
Log-likelihood at zero, LL(0)	-5493.706		-5493.706	
Log-likelihood at convergence, LL(β)	-4933.841		-4908.552	
Chi-square	1119.728		1170.308	
Number of observations, N	8,291		8,291	

Table 10.3: Marginal Effects Associated to the Random Parameters Model

Variable	Marginal Effects (Random Parameters Model)				
	Y = 0	Y = 1	Y = 2	Y = 3	Y = 4
Alignment of highway section (1 for curved section, 0 otherwise)	0.000	0.001	0.008	0.013	-0.023
Day of the week (1 if weekend, 0 otherwise)	0.000	-0.000	-0.005	-0.011	0.017
Months of the year (1 if summer months (June - August), 0 otherwise)	0.000	0.002	0.009	0.016	-0.027
Light condition of street (1 if dark, 0 otherwise)	0.000	-0.001	-0.008	-0.016	0.025
Number of vehicles involved in the crash	-0.000	-0.002	-0.014	-0.026	0.042
Trailing unit when the crash occurred(1 if one trailer, 0 otherwise)	-0.000	-0.005	-0.029	-0.049	0.084
Vehicle role (1 if struck by other vehicle, 0 otherwise)	-0.000	-0.003	-0.023	-0.043	0.069
The most harmful event (1 if rollover, 0 otherwise)	-0.000	-0.006	-0.033	-0.052	0.093
Vehicle maneuver during pre-crash situation (1 if left or right side departure, 0 otherwise)	0.000	0.005	0.024	0.039	-0.069
Orientation of vehicle at the time of crash (1 if sideswipe in the same direction, 0 otherwise)	-0.000	-0.002	-0.016	-0.032	0.050
Vehicle maneuver just prior to impending crash (1 if changing lane, 0 otherwise)	0.000	-0.002	-0.016	-0.035	0.054
Vehicle maneuver just prior to impending crash (1 if going straight, 0 otherwise)	0.000	-0.001	-0.007	-0.013	0.022
Factor of crash identified in the investigation (1 if speed, 0 otherwise)	0.000	0.001	0.004	0.007	-0.012
Occupants' use of available vehicle restraints (1 if lap and shoulder restraint used, 0 otherwise)	-0.000	-0.004	-0.021	-0.035	0.060
Age of the occupants (1 if for age group of 55-65 years, 0 otherwise)	0.000	0.001	0.007	0.013	-0.022
Gender of the occupants (1 if male, 0 otherwise)	-0.000	-0.002	-0.013	-0.022	0.037
Drivers' working/residing place according to license record (1 if Texas, 0 otherwise)	0.000	0.002	0.009	0.017	-0.028

Six parameters were found to be random with statistically significant standard deviations for their assumed distributions; the constant term, dark conditions, lane changing, one trailer trucks, left or right side departure and number of vehicles involved in the crashes. For the parameters whose standard

deviations were not statistically different from zero, the parameters were fixed to be constant across the observations in the model.

Turning to the results found on Table 10.2, roadway departure (left or right) was found to be significant with a random parameter that is normally distributed, with a mean of -0.441 and a standard deviation of 0.473. Given this estimates, roadway departure indicates 82.44% (less than zero) more severity for large-truck crash observations in the sample, whereas, 17.56% of the observations indicates less severity. The variability in the variable "roadway departure" is likely capturing the unaccounted heterogeneity in large-truck involved crashes—such as the under-reporting of injury outcomes or pre-crash evasive maneuvers as recorded by the investigating police officers at the crash scene.

In addition to the constant term, other explanatory variables were found to be significant. These variables are related to the fused datasets from GES and pertain to human, vehicle, and road-environmental factors.

10.5.1 Human related factors

Previous work related to large truck injury severity lacked variables related to human factors. As seen from Table 10.2, males are more likely to experience less severe injuries. A possible explanation is the greater physiological strength and injury sustaining capability of males over that of females (O'Donnell and Connor 1996). Also, Abdel-Aty (2003) finds that female drivers are more likely to be involved in more severe crashes. Another demographic character related to the occupants (i.e., both drivers and passengers) is the age-group from 55 to 65 is more likely to be injured severely. This might also be the same reasons of physical strength related to injury tolerance due to age effect.

The indicator variable for speed is found to be significant. High speed crashes involving large trucks may lead to greater injury severity levels due to larger kinetic forces especially when objects of

substantial mass are involved. Khattak et al. (2003) also find speeding to be a significant factor that impacts the level of injury severity experienced by the vehicle's occupants.

The proper use of in-vehicle restraints is proven to save lives. As seen in Table 10.2, vehicle occupants that restrained by lap/shoulder belts are less prone to be severely injured. One possible explanation is that the components inside the passenger compartment such as the dashboard, wheel, steering column, console, head-rest, A-pillar and windshield may inflict greater damage to unrestrained individuals as they hurl towards these objects inside the vehicles during the crash phase as opposed to properly restrained individuals. In addition, occupants may be ejected from the vehicles leading to increased levels of injury severity. Both, Abdel-Aty (2003) and Boufous et al. (2008) identified through their studies that not wearing a seat belt is a leading risk factor and may lead to higher injury severities suffered by vehicle occupants. The findings of this variable could be also substantiated on a study by Gkritza and Mannering (2008) on the presence of passenger on the seat-belt usage in single- and multi-occupant driving scenarios.

With regards to the indicator variable representing the drivers' working or residing in the State of Texas according to license record, is found to be significant in the model. Licensed and registered drivers in the State of Texas are more likely to be involved in more severe injury crashes. This variable may be capturing the driving complexities related to the diverse geographical nature and higher speed limit posted in the State of Texas. That is, the speed limit in Texas is higher and varies based on rural highways (i.e., 75, 80 and 85 mph) and urban highways (i.e., 75 mph) compared to other states (GHSA, 2012).

10.5.2 Vehicle related factors

As with the influential aspects of human related factors on large-truck involved crashes, factors associated with the vehicle are presented here. The indicator variable for a trailing unit (i.e., a truck is

hauling a trailer) is found to be statistically significant with a random parameter that is normally distributed, with mean 0.553 and standard deviation of 0.510. This indicates that for 82.0% (less than zero) of the crash occurrences where a truck with a trailing unit will experience a lower level of injury severity while 18.0% experience more severe injury. This variable may be capturing the varying degree of driving experience and training with hauling a trailing unit. The randomness in the coefficient may be accounting for driver experience levels with trucks with trailing units.

With respect to vehicle role, when large trucks are struck by other vehicles, the likelihood of the injury outcome is less severe. A possible explanation is that large trucks, when struck by other vehicles (such as passenger cars), may sustain less damage due to their structural design integrity, and size difference. Rather, the effect of passenger vehicles being struck by high velocity trucks at the rear-end results in higher injury outcomes (Duncan et al., 1998). Additionally, this variable may be capturing other vehicular dynamics specific to large trucks, for example damage sustainability potential.

As the number of vehicles involved in a crash increases, the level of injury severity decreases. This continuous variable is found to be statistically significant with a random parameter that is normally distributed, with mean 0.341 and standard deviation of 0.333. This indicates that for 97.6% (less than zero) of the crash occurrences, as the number of vehicles increase in the crash that resultant outcome is lower level of injury severity while 2.4% experience more severe injury. An explanation for this finding is that crashes with many cars, such as pile ups, lessen injury severity due to some unforeseen dynamics and preventive technologies present in vehicles (Chakravarthy et al. 2009).

10.5.3 Road and environmental related factors

In this section we discuss road and environment factors that are found to be significant. As seen from Table 10.2, the indicator variable representing curved road sections leads to more severe injury categories. Consequently, the drivers of large trucks often deal with a higher level of difficulty in

negotiating curved sections especially when considering weight and size of the vehicle. In addition, this variable may be capturing the influence of locational factors related to curve sections as well as skill level of drivers.

With regards to street lighting conditions, dark conditions (i.e., no lighting) was found to be significant with a random parameter that is normally distributed, with mean 0.227 and standard deviation of 1.084. This implies that for roughly 76.2% (less than zero) of the crash occurrences where street lighting conditions is classified as dark, a lower level of injury severity is a possible outcome, while for about 23.8% the opposite. Several studies have found that dark or limited lighting conditions could increase injury severity outcomes (Xie et al. 2009; Chimba and Sando 2009; Helai et al. 2008). Also, this variable may be capturing varying nighttime driving behavior in addition to visibility and sight distance related factors not reported. Truck drivers are usually more cautious in dark conditions on highways than passenger vehicle drivers.

Although not explicitly related to road and environmental factors, the indicator variable for weekend driving was found to be significant. If crashes occur on the weekend, the injury severity sustained by the occupants of large trucks is less severe. This variable might be capturing the effect of weekend driving patterns, density, and frequency of truck trips. Moreover, the indicator variable representing the summer months was also found to be significant and increased the possibility of injury severity. This may be the case since in the summer months there is higher number of vehicles on the road which increases the exposure of passenger vehicles on the highways. This exposure clearly indicates greater interaction of passenger vehicles with large trucks resulting in increasing likelihood of severe crashes. A similar result was found by Malyshkina and Mannering (2009).

10.5.4 Crash mechanism related factors

Lastly, variables related to crash mechanism are presented here. As illustrated in Table 10.2, the indicator variable representing truck rollovers was found to be significant and is more likely to lead to crashes that are less severe. The significance of this variable may stem from vehicular compartment rigidness and properly working restraint systems. The variable may also be capturing driver skill level as well as rollover locational factors. A study by Khattak et al. (2003) and Cate and Richards (2001) found that rollovers cause increased injury severities of occupants in large trucks. In addition, their study associated greater injury severity with curves of five degrees or more. The randomness of the variable coefficient may be accounting for these types of crashes.

Sideswipe in the same direction was found to be significant and is likely to lead to less severe large-truck crashes. A possible explanation may be that truck drivers' skill level and training in regards to steering control in the same direction is possibly minimizing the injury severity outcomes under sideswipe scenarios.

When large trucks depart from the traveled roadway either to left or to right side of the travel direction was found to be significant with a random parameter that is normally distributed, with mean -0.441 and standard deviation of 0.473. This indicates that for 99.9% (less than zero) of the crash occurrences where large-truck departing from the roadway will experience a higher level of injury severity while 0.1% will sustain less severe injuries. The increased level of severity for veering off the road may stem from the possibility of collisions with stationary objects or other dangers present on the traveled roadway. The wide range of variability may be due to the factors related to driver skill level and training. A study performed by Yamamoto and Shankar (2004) also found that running off the roadway may lead to increased injury severities sustained by vehicle occupants.

Lane changing as an evasive maneuver prior to an impending crash was found to be significant with a random parameter that is normally distributed, with mean 0.659 and standard deviation of 0.816.

This implies that for roughly 66.2% (less than zero) of the crash occurrences where lane changing was an evasive maneuver observed a lower level of injury severity, while for about 33.8% a higher level of injury severity may result. A possible explanation for the variability of this estimate may stem from unforeseen factors related to oncoming traffic, median types, or secondary crashes influencing the reported severity. Lane changing as a crash risk factor was also found to be significant by Gray et al. (2008) and Khattak et al. (2003). Similarly, driving straight as an evasive maneuver to avoid a crash was found to be fixed and resulted in less severe injury possibilities. This could be due to driver alertness and crash anticipation because drivers can more successfully brace themselves for impact if they are holding a straight trajectory.

10.6 Summary and future work

In this paper, we analyze large-truck injury severity through a random parameters ordered probit modeling framework. The random parameters ordered probit is an important approach because it allows us to account and correct for unobserved heterogeneity that can arise from factors such as human (i.e., drivers and passengers), vehicle, road-environment, weather, variations in police reporting, temporal and other unobserved factors not captured. The data used in this study was from NASS-GES database for the years of 2005 to 2008, and to the best of our knowledge is one of first studies to explicitly use this database for the modeling of large truck injury outcomes.

The results of the analyses provide some interesting findings. First, human related factors from the fused GES dataset were found to be significant in the model. Of these, the estimate for the factor of crash identified in the investigation “speed” indicator variable was found to be fixed. However, lane changing behavior and departing the roadway during pre-crash stages were found to be random and vary across observations. In contrast to variables related to gender, age of occupants, in-vehicle restraint usage, and Texas truck drivers’ parameter estimates were found to be fixed. Second, in terms of vehicle

related variables from the fused dataset the estimates for a trailing unit, number of vehicles involved were found to be random. While variables related to vehicle role in a collision, orientation of vehicle at the time of crash, vehicle maneuvering during pre-crash were also found to be fixed. Third, the dark indicator variables for road and environment related factors were found to be random. In addition, indicator variables related to highway alignment, day of the week, and summer months (serves as a proxy for traffic conditions) were also found to be fixed across large-truck crash occurrences.

A key finding is the change of signs from the “dark condition” observed between the fixed and random parameters models. Under the fixed model, this variable would increase the likelihood of severe injuries. In contrast, the random parameters model identified the variable coefficient to be random accounting for unobserved factors which leads to cases of severe injuries (i.e., above zero) or less severe cases (i.e., below zero).

Although this study is exploratory in nature, the modeling approach presented in this paper offers a methodology to analyze large-truck injury severity and at same time account for unobserved factors. Applying this approach to state specific datasets with available AADT data, car-following dynamics, and human response and for more years could potentially provide additional information with regards to large-truck crashes. In addition, datasets with driver skill and other cognitive processing information can greatly improve parameter estimates as well as help in the development and improvement of truck driver training.

In future work, we are currently working on a random parameters (mixed) logit modeling framework to see if we can obtain better coefficient estimates in addition to finding more statistically significant variables. Furthermore, we are working with local trucking companies on truck driver training to identify variables of interest for simulation-based safety analyses.

Vita

Mouyid Islam received his Bachelor of Science in Civil Engineering in 2003 from Bangladesh University of Engineering and Technology (BUET). He then received his Master of Engineering in Transportation Engineering from the Asian Institute of Technology (Thailand) in 2005 and a Master of Science in Civil Engineering from Purdue University in 2010 with a research concentration in transportation engineering. Mouyid Islam has been a graduate research and teaching assistant since he started his PhD studies in the Civil Engineering department at the University of Texas at El Paso (UTEP) in fall of 2010. Before pursuing his master degree in the US, he also worked as on-site crash investigator and reconstructionist for simulation-based in-depth studies at the Thailand Accident Research Center and research fellow at Accident Research Center at BUET.

Mouyid has presented his research in the Transportation Research Board (TRB) Annual meetings, Transportation Research Forum conferences, TexITE meetings, and graduate school presentations at UTEP during the period of 2010 to 2012. He has published his research in conference proceedings and several publications are under review in the transportation journals. Mouyid received the Dwight Eisenhower Transportation Fellowship as well as International Road Federation (IRF) Fellowship in 2012. In addition, Mouyid was awarded the “Outstanding Student” by the Texas Institute of Transportation Engineers (TexITE) (District 9) for his leadership and service for the ITE Student Chapter at UTEP. His research paper entitled “Identifying the Contributing Factors of Injury Severity for Large-truck involved Crashes on US Highways,” was selected for a student paper award in TexITE meeting. Mouyid was ITE UTEP chapter president, secretary, and currently ITE student member; member of Chi-Epsilon, member of American Society of Civil Engineers. Mouyid is also affiliated with Truck and Bus Safety subcommittee (ANB70) with TRB.

Permanent address: Mamtaz Mahal 404, 63/B North Dhanmondi, Kalabagan, Dhaka, Bangladesh

This thesis/dissertation was typed by Mouyid Islam.