Advanced Econometric Approaches to Modeling Driver Injury Severity

by

Shamsunnahar Yasmin

Civil Engineering and Applied Mechanics
McGill University
Montreal, Canada

March, 2015

Dissertation
submitted to McGill University in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

ABSTRACT

The objective of the dissertation is to develop advanced econometric frameworks to address methodological gaps in safety literature while employing these models developed to study important empirical issues. Crash severity analysis has evolved on examining the influence of several factors, comprising of driver characteristics, vehicle characteristics, roadway attributes, environmental factors and crash characteristics on traffic crash related severities. These associated risk factors are critical to assist decision makers, transportation officials, insurance companies, and vehicle manufacturers to make informed decisions to improve road safety, thereby providing empirical evidence regarding the critical factors would allow us to suggest remedial measures to reduce the negative consequences of crash outcomes. To that extent, the current dissertation contributes to the severity analysis with a specific focus on driver injury severity analysis.

Road safety researchers have employed several statistical formulations for analyzing the relationship between injury severity and crash related factors. However, there are still several methodological and empirical gaps in safety literature. The specific emphasis of the current dissertation is to contribute substantially towards methodological gaps in the state of the art for driver injury severity analysis along six directions: (1) appropriate model framework, (2) underreporting issue in severity analysis, (3) exogenous factor homogeneity assumption (4) multiple dependent variables in severity analysis, (5) continuum of fatal crashes and (6) data pooling from multiple data sources. In the dissertation, several econometric models are formulated, estimated and validated to address the aforementioned methodological issues through five different empirical studies.

The relevance of alternate discrete outcome frameworks for modeling driver injury severity is examined by empirically comparing several ordered (ordered logit, generalized ordered logit and mixed generalized ordered logit) and unordered outcome (multinomial logit, nested logit, ordered generalized extreme value logit and mixed multinomial logit) models. The research also explores the effect of potential underreporting on alternative frameworks by artificially creating an underreported data sample from the driver injury severity sample. The performance of the alternative frameworks are examined in the context of model estimation and validation (at the aggregate and disaggregate level) by using a host of comparison metrics. Further, the performance of the model frameworks in the presence of underreporting is explored – with and without corrections to the estimates. The results from these extensive analyses point towards the emergence of the mixed generalized ordered logit framework as a strong competitor to the mixed multinomial logit model in modeling driver injury severity.

In addressing the exogenous factor homogeneity assumption, a latent segmentation based generalized ordered logit model is formulated, estimated and validated in the context of driver injury severity. The proposed model probabilistically allocates drivers (involved in a crash) into different injury severity segments based on crash characteristics to recognize that the impacts of exogenous variables on driver injury severity level can vary across drivers based on both observed and unobserved crash characteristics. The results clearly highlight the need for segmentation based on crash characteristics. Overall, the comparison exercise supports the hypothesis that latent segmentation based generalized ordered logit model is a promising ordered outcome framework for accommodating population heterogeneity and for relaxing the fixed threshold assumption in examining driver injury severity.

An analysis is conducted to examine the hypothesis that collision type fundamentally alters the injury severity pattern under consideration. Towards this end, we propose a joint modeling

framework to study collision type and injury severity sustained as two dimensions of the severity process. We employ a copula based joint framework that ties the collision type (represented as a multinomial logit model) and injury severity (represented as an ordered logit model) through a closed form flexible dependency structure to study the injury severity process. The proposed approach also accommodates the potential heterogeneity (across drivers) in the dependency structure. Further, the study incorporates collision type as a vehicle-level, as opposed to a crash-level variable as hitherto assumed in earlier research, while also examining the impact of a comprehensive set of exogenous factors on driver injury severity. The findings of this study provide a more complete picture of injury severity profile associated with different collision type, thus target based countermeasures could be devised to address the entire profile of collision mechanism.

Another study contributes to continuing research on fatal crashes. Specifically, rather than homogenizing all fatal crashes as the same, our study analyzes the fatal injury from a new perspective by examining fatality as a continuous spectrum based on survival time ranging from dying within thirty days of crash to dying instantly (as reported in the Fatality Analysis Reporting System database). The fatality continuum is represented as a discrete ordered dependent variable and analyzed using the mixed generalized ordered logit model. We also propose to estimate a two equation model that comprises of a regression equation for emergency medical response time and mixed generalized ordered logit model for fatality continuum with residuals from the emergency medical response time model to correct for endogeneity bias on the effect of exogenous factors on the timeline of death. Such research attempts are useful in determining what factors affect the time between crash occurrence and time of death so that safety measures can be implemented to prolong survival. The model estimates are augmented by conducting elasticity analysis to highlight the important factors affecting time-to-death process.

Finally, a study focuses on developing a framework for pooling of data from Fatality Analysis Reporting System and Generalized Estimates System databases. The validation of the pooled sample against the original Generalized Estimates System sample (un-pooled sample) is carried out through two methods: univariate sample comparison and econometric model parameter estimate comparison. The validation exercise indicates that parameter estimates obtained using the pooled data model closely resemble the parameter estimates obtained using the un-pooled data. After we confirm that the differences in model estimates obtained using the pooled and un-pooled data are within an acceptable margin, we also simultaneously examine the whole spectrum of injury severity on an eleven point ordinal severity scale – no injury, minor injury, severe injury, incapacitating injury, and seven refined categories of fatalities ranging from fatality after 30days to instant death – using the pooled dataset. We also demonstrate how our approach can be employed to identify factors affecting potentially fatal crashes (non-instantaneous) and improving the chances of survival of motor vehicle occupants involved through the elasticity exercise.

The econometric models developed in the dissertation are estimated using police reported crash databases at the regional and the national level from different industrialized countries. Specifically, the dissertation research is undertaken employing General Estimates System and Fatality Analysis Reporting System of the United States and the Victoria crash database of Australia. In addition to making the aforementioned methodological contributions, the dissertation also makes a substantial empirical contribution to the existing safety literature. Specifically, several policy measures in terms of engineering, enforcement, education and emergency response strategies are identified to improve safety situation and to reduce road crash related fatalities.

RÉSUMÉ

Cette dissertation vise à développer des modèles économétriques avancés afin d'adresser des lacunes méthodologiques dans la littérature sur la sécurité routière, tout en employant ces modèles afin d'étudier des questions empiriques importantes. L'analyse de la gravité des collisions routières a évolué afin d'examiner l'influence de nombreux facteurs, comprenant les caractéristiques des chauffeurs, des véhicules, des chaussées, des facteurs environnementaux ainsi que des collisions routières elles-mêmes, sur les fatalités résultant de ces collisions. Ces facteurs de risque associés sont cruciaux afin d'informer les décideurs politiques, les responsables des transports, les compagnies d'assurance ainsi les fabricants de véhicules afin qu'ils puissent prendre des décisions éclairées pour améliorer la sécurité routière. Ainsi les résultats d'analyses empiriques concernant ces facteurs critiques nous permettraient-ils de suggérer des mesures de remédiation qui pourraient réduire les conséquences négatives des collisions routières. Dans ce contexte, cette dissertation contribue à la littérature sur l'analyse de la sévérité des collisions routières tout en portant une attention particulière à l'analyse de la sévérité des blessures des conducteurs.

Les chercheurs en sécurité routière ont employé de nombreuses formulations statistiques afin d'analyser les liens entre la gravité des blessures et les caractéristiques des collisions. Cependant, il existe encore de nombreuses lacunes méthodologiques et empiriques dans la littérature. L'objectif principal de cette dissertation est de contribuer à corriger ces lacunes méthodologiques dans les approches d'analyse de sécurité routière de pointe, en adressant six thèmes principaux : (1) cadre de modélisation approprié, (2) problèmes liés à la sous-déclaration dans l'analyse de sévérité, (3) hypothèse d'homogénéité des facteurs exogènes, (4) variables dépendantes multiples en analyse de sévérité, (5) continuum de collisions fatales et (6) regroupement de données de sources multiples. Dans cette dissertation, de nombreux modèles économétriques sont formulés, estimés et validés afin d'adresser les lacunes méthodologiques susmentionnées à travers cinq études empiriques distinctes.

La pertinence de divers cadres statistiques discrets dans la modélisation de la gravité des blessures des conducteurs est examinée à travers la comparaison de nombreux modèles ordonnés (logit ordonné, logit ordonné généralisé, logit ordonné généralisé mixte) et non-ordonnés (logit multinomial, logit imbriqué, logit ordonné généralisé à valeur extrême, logit multinomial mixte). Notre recherche examine également les effets de la sous-déclaration potentielle sur les divers cadres statistiques employés à travers la création d'un échantillon de données artificiel à partir de l'échantillon portant sur la gravité des blessures des conducteurs. Les performances de ces cadres divers sont examinées dans le contexte de l'estimation et de la validation des modèles (aux niveaux agrégés et désagrégés) en utilisant une large gamme d'outils de comparaison. De plus, la performance des cadres de modélisation en présence de sous-déclaration est investiguée – avec et sans correction des estimations. Les résultats de ces analyses étendues démontrent que le logit ordonné généralisé mixte pourrait rivaliser avec le logit multinomial mixte dans le cadre de la modélisation de la gravité des blessures des conducteurs.

Afin d'adresser l'hypothèse d'homogénéité des facteurs exogènes, un logit ordonné généralisé à base de segmentation latente est formulé, estimé, et validé dans le contexte de l'analyse de gravité des blessures des conducteurs. Le modèle proposé alloue de façon probabiliste les conducteurs (victimes d'une collision) dans divers segments de gravité des blessures basé sur les caractéristiques de la collision afin de reconnaître que l'impact des variables exogènes sur le niveau de gravité des blessures des conducteurs peut varier selon le conducteur en fonction de caractéristiques observées et non-observées des collisions. Les résultats démontrent clairement le

besoin de segmentation basée sur les caractéristiques des collisions. Globalement, l'exercice de comparaison soutien l'hypothèse que le logit ordonné généralisé basé sur la segmentation latente est un cadre ordonné prometteur afin d'accommoder l'hétérogénéité de la population, ainsi que pour assouplir l'hypothèse de seuils fixes lors de l'analyse de la gravité des blessures des conducteurs.

Une analyse est menée afin de déterminer si le type de collision a un impact fondamental sur les types de blessures encourues. Dans ce contexte, nous proposons un cadre statistique combiné afin d'examiner le type de collision et les blessures encourues comme deux dimensions de l'analyse de gravité. Nous employons un cadre combiné à base de copules qui relie le type de collision (représenté à travers un modèle logit multinomial) et la gravité des blessures (représenté à travers un modèle logit ordonné) à travers une structure dépendante flexible à forme fermée afin d'étudier le processus de gravité des blessures. L'approche proposée adresse également l'hétérogénéité potentielle (des conducteurs) dans la structure dépendante. En outre, l'analyse considère que les caractéristiques d'une collision dépendent du type de véhicules plutôt que du type de collision elle-même, comme il est généralement considéré dans la littérature sur le sujet; tout en examinant l'impact d'un ensemble complet de variables exogènes sur la gravité des blessures des conducteurs. Les conclusions de cette étude permettent une meilleure compréhension des profils de gravité des blessures associés aux divers types de collision ainsi que le développement de contre-mesures ciblées afin d'adresser toute la gamme des mécanismes de collisions.

Une autre étude contribue à la recherche sur les collisions fatales. Spécifiquement, plutôt que d'homogénéiser toutes les collisions fatales en une catégorie unique, notre analyse propose une nouvelle approche qui considère les collisions fatales comme un spectre continu basé sur le temps de survie, qui va de décès dans les trente jours suivant la collision à décès immédiat (tel que rapporté dans la base de données « Fatality Analysis Reporting System Database »). Le spectre de décès est représenté comme une variable dépendante ordonnée discrète et est analysé en utilisant un modèle logit ordonné généralisé. Nous proposons également d'estimer un modèle à deux équations qui comprend une équation de régression représentant le temps de réponse de l'équipe médicale et d'un logit ordonné généralisé mixte pour le spectre de décès, avec les résiduels du modèle pour le temps de réponse de l'équipe médicale qui corrigent pour les distorsions endogènes sur les effets des facteurs exogènes sur les délais de décès. Ces efforts de recherche sont utiles afin de déterminer les facteurs qui affectent le délai entre la collision et l'heure du décès, dans le but de concevoir des mesures de sécurité qui puissent prolonger la survie des victimes. Les résultats de l'analyse sont complémentés par une analyse des élasticités afin de souligner les facteurs importants qui affectent l'heure du décès.

Enfin, nous présentons une analyse portant sur le développement d'un cadre de mise en commun des données des bases de données « Fatality Analysis Reporting System » et « Generalized Estimates System ». La validation de l'échantillon regroupé comparé à l'échantillon original de « Generalized Estimates System » (échantillon non regroupé) est effectuée en utilisant deux méthodes : comparaison des échantillons univariés et comparaison des estimations de paramètres de modèles économétriques. L'exercice de validation indique que les estimations de paramètres obtenues en utilisant les données regroupées sont très similaires aux estimations obtenues en utilisant les données non regroupées. Une fois confirmé que les différences des estimations de paramètres obtenues à partir des données regroupées et non-regroupées sont dans des marges acceptables, nous examinons simultanément tout le spectre de gravité des blessures sur une échelle de gravité à onze niveaux – aucune blessure, blessure légère, blessure grave, blessure incapacitante, et sept niveaux de décès allant de décès dans les trente jours suivant l'accident à décès immédiat – en utilisant les données regroupées. Nous démontrons également comment cette approche peut être utilisée afin d'identifier les facteurs affectant les collisions potentiellement fatales (décès non-immédiat) et améliorer les chances de survie des passagers à travers l'exercice d'analyse des élasticités.

Les modèles économétriques développés dans cette dissertation sont estimés grâce aux données contenues dans les bases de données de police aux échelles régionales et nationales de divers pays développés. Plus précisément, cette recherche est basée sur les bases de données « General Estimates System » et « Fatality Analysis Reporting System » des États-Unis et la base de données « Victoria crash database » de l'Australie. En plus des contributions méthodologiques susmentionnées, cette dissertation constitue également une contribution empirique importante à la littérature de sécurité routière. Spécifiquement, de nombreuses mesures politiques liées à la conception, la mise en vigueur, l'éducation et les stratégies de réaction d'urgence sont identifiées afin d'améliorer la sécurité des passagers et de réduire les décès liés aux collisions routières.

ACKNOWLEDGEMENT

I would like to take this opportunity to acknowledge all the people who have supported, inspired and helped me throughout the journey of my doctoral education.

First, I would like to convey my deepest appreciation to the Faculty of Engineering of McGill University for awarding me with the McGill Engineering Doctoral Award (MEDA). I was honored to receive the award in the form of a prestigious named fellowship – the Emil Nenniger Memorial Fellowship along with the Graduate Excellence Fellowships (GEF). I would like to express my deepest gratitude and gratefulness to Mrs. F.S. Nenniger and Dr. Emil H. Nenniger for their generous donation for the award that I received.

I owe my deepest gratitude to my supervisor, Dr. Naveen Eluru, for his valuable advice and financial support. This dissertation would not have been possible without his encouragement, dedication and guidance. He is unprecedented in supervising his students. He has always amazed me with his passion, dedication and enthusiasm for research and I can undoubtedly say that he is the best supervisor that I could have ever worked with for my doctoral education. I consider myself fortunate and privileged for getting the opportunity to work under his supervision.

I am grateful to Dr. Marianne Hatzopoulou for being my co-supervisor. I am also thankful to Dr. Ahmed El-Geneidy, Dr. Luis F. Miranda-Moreno, Prof. Luc E. Chouinard, Dr. Tom Gleeson and Dr. Dimitrios G. Lignos for serving on my dissertation committee and for their valuable suggestions. I would like to take this opportunity to thank Professor Richard Tay for his invaluable advice and encouragement. He has been a great mentor throughout my graduate studies and has always guided me with my educational pursuits.

Special thanks to Sabreena, both personally and professionally, for her help, advice, continuous support, for all the "desi foods" that she cooked and obviously for being my bonu and taking care of me. Thanks to Nilufar for visiting Montreal once in a while from Kelowna, even though for very short durations, and giving me a chance to re-live all the BUET memories through our chit-chats. Many thanks to all my "UNO buddies" – Adham, Timothy, Imtiaz, Rifat, Golnaz and Ahmad – for all the fun and crazy time that we spent together. Also thanks to Adham and Frederic for their help in French translation. I am thankful to Vincent, Frederic and Miguel for all the fun and serious talks. I will cherish their friendship for the rest of my life. Thanks to Juliette &

Chocolat for making the most delicious chocolate desserts which were the best treat for myself after every accomplishment.

Most importantly, my deepest gratitude goes to my mother, my father, my sisters and my uncle for their endless love, unconditional sacrifice, continuous encouragement and for providing me the best possible environment to grow up and proceed forward. Very special thanks goes to my grandmother – the "hero" of my life. Without her sacrifice and unconditional love, I would not be where I am today. Lastly, but by no means least, I would like to express my gratefulness to the Almighty, who blessed me with the courage to complete this dissertation successfully.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

AUTHOR CONTRIBUTIONS

The dissertation contains empirical studies from five full length journal articles. Amon these five articles, three have already been accepted and published in several journals and the other two articles are under review for publications in different peer-reviewed journals. Shamsunnahar Yasmin is the primary author for all of the manuscripts. She has contributed to the articles in terms of data preparation, model estimations and writing. The respective co-authors have contributed to the manuscripts by sharing their valuable insights and editing the manuscripts. The publication details of the manuscripts are presented in the following section.

Chapter 2 is based on the article: **S. Yasmin** and N. Eluru, Evaluating Alternate Discrete Outcome Frameworks for Modeling Crash Injury Severity. Accident Analysis and Prevention, Vol. 59, No. 1, 2013, pp. 506-521.

Chapter 3 is based on the article: **S. Yasmin**, N. Eluru, C. R. Bhat and R. Tay, A Latent Segmentation based Generalized Ordered Logit Model to Examine Factors Influencing Driver Injury Severity. Analytic Methods in Accident Research, Vol. 1, 2014, pp. 23-38.

Chapter 4 is based on the article: **S. Yasmin**, N. Eluru, C. A. R. Pinjari and R. Tay, Examining Driver Injury Severity in Two Vehicle Crashes - A Copula Based Approach. Accident Analysis and Prevention, Vol. 66, 2014, pp. 120-135.

Chapter 5 is based on the article: **S. Yasmin**, N. Eluru, C. and A. R. Pinjari, Analyzing the Continuum of Fatal Crashes: A Generalized Ordered Approach. The paper is under review in the journal of Analytic Methods in Accident Research.

Chapter 6 is based on the article: **S. Yasmin**, N. Eluru, C. and A. R. Pinjari, Pooling Data from Fatality Analysis Reporting System (FARS) and Generalized Estimates System (GES) to explore the Continuum of Injury Severity Spectrum. The paper is under review in the journal of Accident Analysis and Prevention.

CHAPTER 1 Introduction

1.1 Background

Road traffic crashes and their consequences such as injuries and fatalities are acknowledged to be a serious global health concern. These incidents result in physical and emotional trauma as well as huge financial losses for the individuals involved, their families and the society at large. Across the world, these crashes account for 18 deaths and 1,136 disability-adjusted life years (DALY) lost per 100,000 individuals annually (WHO, 2013a; WHO, 2013b). The death toll of road traffic crashes is expected to become the fifth (currently eighth) leading cause of death by the year 2030 (WHO, 2013b) if appropriate remedial measures are not implemented. Researchers and practitioners are constantly seeking remedial measures to reduce the burden of these unfortunate events. In fact, most developed countries, through coordinated multi-sectoral responses to road safety issues, have been able to achieve a reduction in the crash related fatalities. For example, between 1970 and 2011, the annual road fatality rate of the United States (US) declined from 25.7 deaths per 100,000 population to 10.4 deaths per 100,000 population, while at the same time the fatality rate declined from 30.4 to 5.6 fatalities per 100,000 population in Australia (IRTAD, 2013). In spite of these strides in improving road safety, traffic crashes still lead to substantial economic (approximately 2.3% and 2.6% of Gross Domestic Product of the US and Australia, respectively (IRTAD, 2013)) and emotional losses to the society.

Given the import of the consequences of motor vehicle crashes, the issue has received significant attention from researchers and practitioners. In particular, the focus is on identifying and gaining a comprehensive understanding of the factors that contribute to the negative consequence (property damage, injuries and fatalities) of crash. However, road traffic crashes occur due to the complex interactions among several factors, comprising of driver characteristics, vehicle characteristics, roadway design attributes, environmental factors and crash characteristics. To identify the factors and their influence on the severity of crashes, researchers have formulated and employed several modeling frameworks. The current chapter provides an overview of the methodological developments in examining the crash injury severity outcome. Finally, a description of the objectives of the dissertation and its organization are provided.

The remainder of this chapter is organized as follows. The role of severity analysis in road safety and the importance of driver injury severity are briefly discussed in Section 1.2 and 1.3, respectively. Section 1.4 provides a methodological overview of driver injury severity, while Section 1.5 presents and discusses the gaps in existing safety literature. In Section 1.6 we discuss objectives of the dissertation. Section 1.6 provides outlines of the rest of the dissertation.

1.2 The Role of Severity Analysis in Road Safety

The historical background of road safety has initiated with the motor age. In fact, concern with the road traffic crashes goes back to the nineteenth century, when for the first time a person was killed in a fatal car crash (Offaly History, 2007). With the increase in road transportation and motorization, the issue has received significant attention. In today's context, road traffic crashes are identified as a national health problem since these incidents affect the society as a whole both emotionally and economically (Subramanian, 2006; Blincoe et al., 2002). Given the societal impact of road crashes, safety research has mostly been developed based on national crash databases in order to monitor the consequences of road crashes and to develop effective countermeasures both at the local and the national level. These national crash databases are usually compiled from the police reported crash records. However, in police reported crash databases many property damage and minor injury crashes might go underreported since lower crash severity levels make reporting to authorities less likely (Savolainen and Mannering, 2007). Given the limitations of traditional police reported crash databases, a number of proactive (conflict technique, driving simulation) and reactive (crash reconstruction) approaches has been proposed and employed in safety research. While these approaches and the traditional crash analysis are complimentary to one another, the statistical analysis using police reported crash databases has been so far the most prevalent method in traffic safety research. Moreover, it is very challenging and expensive to compile collision data by using other resources. Hence, safety researchers continue to use police reported collision data for safety analysis.

The domain of literature in transportation safety using police reported crash database has evolved along two major streams: the first stream of research is focused on identifying attributes that result in traffic crashes and propose means to reduce the occurrence of traffic crashes (see Lord and Mannering, 2010 for a review of these studies); the second stream of work examines crash events and identifies factors that impact the crash outcome and suggests countermeasures to reduce crash related consequences (injuries and fatalities) (see Savolainen et al., 2011 for a review). The analysis of crash-frequency analysis is predominantly based on non-crash-specific attributes and is focused on identifying the effective countermeasure to improve the roadway design and operational attributes. While improving road infrastructure design to reduce traffic crash occurrence is essential, it is also important to provide solutions to reduce the consequences in the unfortunate event of a traffic crash. To that end, crash severity analysis has evolved on examining the influence of several factors, comprising of driver characteristics, vehicle characteristics, roadway attributes, environmental factors and crash characteristics on traffic crash related severities. These associated risk factors are critical to assist decision makers, transportation officials, insurance companies, and vehicle manufacturers to make informed decisions to improve road safety. In summary, the severity analysis studies contribute to road safety by identifying the various factors that affect the crash severity to assist policy makers in developing appropriate remedial measures.

1.3 Importance of Driver Injury Severity Analysis

In crash severity analysis, researchers have considered and investigated severity by considering outcomes at different levels of crash victims: severity level of each person involved in the crash, most severely injured person of the crash, most severely injured person of each vehicle, driver of the vehicle, pedestrian, bicyclist or user of motorized two/three wheeler vehicles. Among these road user groups, pedestrian, bicyclists and user of motorized two/three wheeler vehicles bear the major burden of road traffic crash related fatalities in low- and middle-income countries of the world (WHO, 2013b). On the other hand, in high-income countries, occupants of four wheeled motorized vehicle, specifically drivers, constitute the highest proportion of crash related fatalities. For example, drivers of passenger vehicles (sedans and light vehicles) represent approximately 50% and 47% fatalities among all road user groups in the US and Australia, respectively (FARS, 2010; Ministry of Infrastructure and Transport, 2010).These statistics clearly indicate that driver safety of passenger vehicles is of great concern for high income countries. Any effort to reduce the social burden of these crashes and enhance driver safety would necessitate the examination of factors that contribute significantly to crash likelihood and/or driver injury severity in the event of a crash.

1.4 Methodological Overview of Driver Injury Severity Studies

A number of research efforts have examined driver injury severity to gain a comprehensive understanding of the factors that affect injury severity. In our review of earlier research we focus on studies examining severity at a disaggregate accident or individual level models of driver injury severity. Within these approaches, we specifically focus on econometric models that employ a discrete variable representation for injury severity analysis. A summary of earlier research on driver injury severity analysis from the perspective of the various ordered and unordered outcome econometric models is provided in Table 1.1. The information presented in the table includes model structures employed for the analysis and identifies the variable categories considered in the analysis from the five broad categories of variables as presented in the table. The following observations may be made from the table. First, the most prevalent mechanisms to study driver injury severity are logistic regression[1] and ordered outcome models (twenty four out of thirty one). The number of studies employing unordered models has been steadily increasing in recent years. Second, the most prevalent unordered outcome structure considered is the multinomial logit model. Third, it is evident from the analysis that very few studies (except Abdel-Aty, 2003; Ye and Lord, 2011) have empirically examined the different frameworks for modeling injury severity[2]. Fourth, only few studies have addressed the effect of observed and unobserved heterogeneity in examining driver injury severity outcomes. Finally, the maturity of the transportation safety community in examining driver injury severity is highlighted by the fact that a majority of studies (seventeen out of thirty one) have considered exogenous variables from all broad categories of variables.

1.5 Gaps in Existing Safety Literature

The preceding section of the driver injury severity studies clearly indicates that literature in severity analysis is vast and growing. These studies offer many useful insights on what factors affect crash severity outcomes. However, there are still several methodological and empirical gaps

[1] To be sure, the logistic regression with two alternatives can be regarded as an ordered logit model with two alternatives.

[2] To be sure, Ye and Lord (2011) have compared the ordered probit, multinomial logit and mixed logit model in terms of underreported data. The authors conclude that all the three models considered in the study perform poorly in the presence of underreported data. The exact impact of underreporting on these model frameworks needs further investigation. The study employs data simulation; however, the models are estimated with just one parameter and for a particular aggregate sample share.

(as highlighted recently in the article by Mannering and Bhat, 2014), suggesting continual needs to develop advanced econometric frameworks to address these gaps in safety literature. In the subsequent discussion, various methodological gaps in the state of the art for driver injury severity analysis are presented and discussed along six directions: (1) appropriate model framework, (2) underreporting issue in severity analysis, (3) exogenous factor homogeneity assumption (4) multiple dependent variables in severity analysis, (5) continuum of fatal crashes and (6) data pooling from multiple data sources.

1.5.1 Appropriate Model Framework

The commonly available traffic crash databases compile injury severity data, primarily, as an ordinal discrete variable (for example: no injury, minor injury, major injury, and fatal injury). Naturally, many earlier studies examining the influence of exogenous factors employ ordered discrete outcome modeling approaches to evaluate their influence on crash severity (for example O'Donnell and Connor, 1996; Renski et al., 1999; Eluru et al., 2008). However, researchers have also employed unordered discrete outcome frameworks to study the influence of exogenous variables (for instance Shankar et al., 1996; Chang and Mannering, 1999; Khorashadi et al., 2005). The ordered outcome models represent the decision process under consideration using a single latent propensity. The outcome probabilities are determined by partitioning the uni-dimensional propensity into as many categories as the dependent variable alternatives through a set of thresholds. Unordered discrete outcome frameworks offer a potential alternative to the analysis of ordered discrete variables. These models are characterized, usually, by a latent variable per alternative and an associated decision rule. The unordered models, usually, allow for additional parameter specification because they are tied to alternatives as opposed to a single propensity in the ordered models. The applicability of the two frameworks for analyzing ordinal discrete variables has evoked considerable debate on using the appropriate model for analysis. Yet, there is little research on empirically examining the differences between the ordered and unordered frameworks, specifically in the context of crash injury severity.

1.5.2 Underreporting Issue in Severity Analysis

Another issue that has received little attention in road safety literature is the influence of underreporting associated with conventional crash databases on alternative model frameworks.

The estimation of injury severity models correspond to the assumption of random sampling of severities from a population, where the probability of occurring for each individual crash is equal (Savolainen et al., 2011). However, the unknown population shares of such outcome-based crash severity data make the estimation of parameters even more challenging. Moreover, most of the crash data are sampled from police reported crash database. Several previous studies (Elvik and Mysen, 1999; Yamamoto et al., 2008) have provided evidence of underreporting issues related to the police-reported crash database. In such cases, the application of traditional econometric frameworks may result in biased estimates (Yamamoto et al., 2008). In the presence of underreported data, the unordered outcome framework is considered to be more effective compared to the ordered response framework. In the case of an underreported decision variable, the traditional multinomial logit (MNL) model provides estimates that are unbiased i.e. the elasticity effects of the variables are not affected by the underreported data. This is often considered as a strong reason for promoting the use of unordered models over ordered models in modeling injury severity. It is important to recognize that the potential advantage applies only to MNL models under the condition that the dataset under examination satisfies the Independence of Irrelevant Alternatives (IIA) property (Ben-Akiva and Lerman, 1985). Hence, the nested logit and other advanced logit models that relax the IIA property are unlikely to yield unbiased estimates in the presence of underreporting. Moreover, the comparison of these two frameworks has mostly been undertaken in the context of traditional ordered models. The Generalized Ordered Logit (GOL) framework with its improved flexibility will provide the true benchmark for a fair comparison. Therefore, it is also essential to examine how alternative modeling frameworks are impacted by underreporting; thus allowing us to adopt frameworks that are least affected by underreporting.

1.5.3 <u>Exogenous Factor Homogeneity Assumption</u>

The widely employed discrete outcome formulations (ordered, generalized ordered, or unordered frameworks) typically restrict the impact of exogenous variables to be the same across the entire population of crashes (Eluru et al., 2012; Xie et al., 2012; Yasmin et al., 2014). But, the impact of control variables on crash injury severity might vary across individuals based on different crash attributes. Ignoring such heterogeneous impact of variable might result in incorrect coefficient estimates.

One approach to extend these formulations to allow heterogeneity effects (variations in the effects of variables across the population) is to specify random coefficients (rather than impose fixed coefficients) (for example, see Eluru and Bhat, 2007; Paleti et al., 2010; Srinivasan, 2002; Morgan and Mannering, 2011; Kim et al., 2013). But, while the mean of the random coefficients can be allowed to vary across drivers based on observed crash-specific variables, the random coefficients approach usually restricts the variance and the distributional form of a random coefficient to be the same across all drivers. Thus, in a crash context, the impact of a rear-end crash (relative to an angular crash) may lead to a certain distribution of injury risk propensity due to unobserved factors. This distribution may be tight for low speed crashes (that is, the injury risk may be negative in the mean and tightly distributed about this mean), but more variant for high speed crashes (that is, the injury risk may be quite volatile in high-speed situations, with rear-end collisions leading to high injury severity in some cases and low injury severity in some other cases). This is a case of the distribution on the rear-end crash variable being dependent on another variable (low speed or high speed crashes). Such possibilities cannot be easily accommodated in random coefficients models. Besides, an *a priori* distribution form has to be imposed on the random coefficients, and the normal distribution assumption is usually imposed even though there is no reason why other distribution forms may not be more appropriate.

A second approach to allow heterogeneity effects is to consider segmenting the population based on exogenous variables (such as collision type, initial impact point of collision, speed, and location of impact) and estimate separate models for each segment (see Aziz et al., 2013 for segmentation based on location; Islam and Mannering, 2006 for segmentation based on driver demographics). However, because there may be many variables to consider in the segmentation scheme, the number of segments (formed by the combination of the potential segmentation variables) can explode rapidly. This causes problems in estimation because of very small sample sizes in some of the segments, and thus analysts tend to fall back to segmenting along 2-3 variable utmost (see Bhat, 1997 for a good discussion of these issues). To address this limitation, more advanced approaches such as clustering techniques that allow to segment based on a multivariate set of factors have been employed (Mohamed et al., 2013; Depaire et al., 2008). However, the approach still requires allocating data records exclusively to a particular segment, and does not consider the possible effects of unobserved factors that may moderate the impact of observed exogenous variables.

A <u>third approach</u> to accommodate heterogeneity is to undertake an endogenous (or sometimes also referred to as a latent) segmentation approach (see Bhat, 1997). In this approach, the drivers involved in collisions are allocated probabilistically to different segments, and segment-specific injury severity models are estimated for each segment. At the same time, each segment is identified based on a multivariate set of exogenous variables. Such an endogenous segmentation scheme is appealing in many respects: (a) each segment is allowed to be identified with a multivariate set of exogenous variables, while also limiting the total number of segments to a number that is much lower than what would be implied by a full combinatorial scheme of the multivariate set of exogenous variables, (b) the probabilistic assignment of drivers to segments explicitly acknowledges the role played by unobserved factors in moderating the impact of observed exogenous variables, and (c) there is no need to specify a distributional assumption for the coefficients (Greene and Hensher, 2003).

This third approach may be viewed as a combination of the two earlier approaches, in that it considers a multivariate set of exogenous variables in the segmentation and also allows unobserved variable effects to moderate the impact of exogenous variables. In fact, the third approach is equivalent to specifying a (discrete) non-parametric distribution on the coefficients (rather than the continuous parametric distribution assumption of the first approach), while also allowing the non-parametric distribution shape to be a function of a multivariate set of exogenous variables. The approach has been employed recently in the safety literature (Eluru et al., 2012; Xie et al., 2012; Xiong and Mannering, 2013; Yasmin et al., 2014). But these studies have employed either traditional ordered (ordered logit/ordered probit) or traditional unordered (multinomial logit) outcome frameworks in examining injury severity levels within latent segmentation based approach. However, the traditional ordered response formulation imposes a restrictive monotonic assumption regarding the impact of exogenous variables on the injury severity levels and the unordered outcome model does not recognize the inherent ordering of the crash severity outcome. To recognize the ordinality of the injury severity levels, as well as to provide as much flexibility as the unordered response formulation, Eluru et al. (2008) proposed the generalized ordered outcome formulation that bridges the divide between the traditional ordered outcome and the traditional unordered outcome formulations. Thus, the severity analysis would benefit from employing GOL framework in the latent segmentation based approach[3].

1.5.4 Multiple Dependent Variables in Severity Analysis

The most commonly identified exogenous factor that significantly affects traffic crash injury severity outcome is the collision type variable. Most of the earlier studies consider the collision type as an explanatory variable in modeling injury severity (except Ye et al., 2008 and Rana et al., 2010). In this approach, the analyst imposes the assumption that the injury severity profile for vehicle occupants in all types of crashes is the same and any potential differences between different collision types can be accurately captured by employing the collision type variable as an explanatory variable. However, it is possible that various collision types might lead to distinct vehicle occupant injury severity profiles i.e., the overall manifestation of injury severity is different by collision type. For example, consider the impact of the gender variable in injury severity models. It is possible that males due to their higher physiological strength are more equipped to resist severe injuries in crashes. However, in a head-on crash due to the greater dissipation of kinetic energy, the physiological advantage might be inadequate. At the same time, the additional strength might be beneficial for male occupants to avoid severe injury in the event of other collision types such as side-swipe. This is an example of how a collision type variable moderates the impact of gender. It is plausible to visualize that collision type variables might similarly affect multiple exogenous variables – indicating that the injury severity profile itself is moderated by the collision type. Thus, estimating a single injury severity model, when such distinct profiles of injury severity exist, will result in incorrect and biased estimates. In fact, several studies have recognized this in safety literature and estimated injury severity focused on a specific type of collision - Head-on collision: Gårder, 2006; Conroy et al., 2008; Zuxuan et al., 2006; Zhang and Ivan, 2005; Rear-end collision: Khattak, 2001; Yan et al., 2005; Das and Abdel-Aty, 2011; Abdel-Aty and Abdelwahab, 2003; and Angular collision: Jin et al., 2010; Chipman, 2004. These studies provide evidence that collision type has a fundamentally distinct effect on injury severity sustained in the crash.

[3] The more recent work of Xiong and Mannering, (2013) proposes a latent segmentation model that further specifies unobserved heterogeneity in each segment-level injury severity model using a continuous multivariate normal distribution for the coefficients. This is tantamount to a discrete mixture-of-normals approach. Though, GOL does not account for unobserved heterogeneity in the segment level models in this paper, GOL can accommodate the more realistic case of injury reporting in more than two injury severity levels (the study by Xiong and Mannering, (2013) on the other hand, was a binary choice model of injury severity).

Given the possibility of distinct injury severity profiles – the estimation of separate injury severity models for various collision types seems the appropriate solution. At the same time, it is also important to investigate the factors that result in crashes of a particular collision type. This necessitates a model for collision type; an unordered outcome variable that can be studied using a multinomial logit model. Within this system, it is possible that the collision type and resulting injury severity are influenced by the same set of observed and unobserved factors. Accommodating for the impact of observed factors is relatively straightforward within the traditional discrete outcome models by estimating distinct outcome models for collision type (multinomial logit) and injury severity (ordered logit). The process of incorporating the impact of unobserved factors poses methodological challenges. Essentially, accommodating the impact of unobserved factors recognizes that the two dimensions of interest are realizations from the same joint distribution. Traditionally, in econometric literature, such joint processes are examined using simulation based approaches that stitch together the processes through common unobserved error terms (see Eluru and Bhat 2007; Abay et al., 2013 for examples in safety literature). In this direction, Ye et al. (2008) propose a simulation based simultaneous equation framework to study the collision type and injury severity dimensions. The framework employs maximum simulated likelihood approach and requires simulation in the order of the dimension of collision type variables. The process of applying simulation for such joint processes is likely to be error-prone in model estimation as well as inference – particularly the estimation of standard errors (see Bhat, 2011 for a discussion). At the same time, ignoring the presence of such potential jointness may lead to biased and inconsistent parameter estimates in modeling injury severity outcome (Chamberlain, 1980; Eluru and Bhat, 2007; Washington et al., 2003).

More recently, a closed form approach that obviates the need for simulation has been proposed in transportation literature for examining joint decision processes. The approach, referred to as Copula Approach, allows for flexible dependency structures across joint dimensions while retaining the closed form structure (see Bhat and Eluru, 2009). In fact, Rana et al., (2010) employed a copula based approach to consider the crash type and injury severity as a joint process with success. However, both of these studies (Ye et al., 2008, Rana et al., 2010) that jointly model the collision type and injury severity outcome describe the collision type as a crash level variable. But, depending on the position of driver and the initial point of impact, it is possible that the individual vehicle might have different effects in the manner of collision for the same type of collision (see

Khattak, 2001 for a discussion in the context of rear-end collision). Moreover, the earlier approach considers the dependency parameter in the copula model to be the same across the entire crash database. However, it is possible that several exogenous factors might actually affect the dependency profile. In other words, the correlation between collision type and injury severity might be stronger or weaker depending on the various attributes of the particular crash. Allowing for such flexibility in the dependency profile allows for more accurate model estimation.

1.5.5 <u>Continuum of Fatal Crashes</u>

In identifying the critical factors contributing to crash injury severity, safety researchers have focused on either examining fatal crashes (involving at least one fatally injured vehicle occupant) or traffic crashes that compile injury severity spectrum at an individual level (such as no injury, possible injury, non-incapacitating injury, incapacitating injury and fatality). In the US, the former category of studies predominantly use the Fatality Analysis Reporting System (FARS) database (see Evans and Frick, 1988; Preusser et al., 1998a; Zador et al., 2000; Gates et al., 2013) while the latter group of studies typically employ the General Estimates System (GES) database (see Kockelman and Kweon, 2002; Eluru and Bhat, 2007). FARS database compiles crashes if at least one person involved in the crash dies within thirty consecutive days from the time of crash. Further, FARS database reports the exact timeline of the fatal occurrence within thirty days from the time to crash.

A number of research efforts have examined the impact of exogenous characteristics associated with fatal crashes employing FARS. These studies offer many useful insights on what factors affect crash related fatality, particularly in the context of fatal vs. non-fatal injury categorization. However, there is one aspect of fatal crashes that has received scarce attention in the traditional safety analysis. These studies assume that all fatal crashes in the FARS dataset are similar. Keeping all else same, a fatal crash that results in an immediate fatality is clearly much more severe than another crash that leads to fatality after several days. In fact, there is evidence from epidemiological studies (Tohira et al., 2012; Sauaia et al., 1995) that the risk factors associated with early and late trauma deaths of crash victims are different from the risk factors associated with late trauma deaths. For instance, Tohira et al. (2012) reported that older drivers (aged 65 years or older) and/or crash victims with a depressed level of consciousness were at increased risk of late trauma death. Research attempts to discern such differences are useful in determining what factors affect the time between crash occurrence and time of death so that countermeasures can be implemented to improve safety situation and to reduce road crash related fatalities. Early EMS (Emergency Medical Service) response is also argued to potentially improve survival probability of motor vehicle crash victims (Clark and Cushing, 2002; Clark et al., 2013). In fact, Meng and Weng (2013) reported 4.08% decrease in the risk of death from one minute decrease in EMS response time, while Sánchez-Mangas et al. (2010) reported that a ten minutes EMS response time reduction could decrease the probability of death by one third. Given the import of this variable, it is also important to explore the effect of EMS response time in examining crash fatalities. Moreover, the first hour after crash occurrence – most popularly known as the "golden hour" (Cowley et al., 1973; Stewart, 1990) – is the most important phase in trauma care to ensure the best chance of a crash victim survival. Besides, identification of critical crash attributes that contribute to major trauma is crucial not only for preclinical trauma care but also for the optimal use of emergency medical service (EMS) resources (such as selecting appropriate patient transport method from the accident scene) (Weninger and Hertz, 2007; Meng and Weng, 2013). The detailed information available in FARS provides us a continuous timeline of the fatal occurrences from the time of crash to death. This allows for an analysis of the survival time of victims before their death. To be sure, earlier research efforts focused on examining the factors influencing the time period between road accident and death (Golias and Tzivelou, 1992; Marson and Thomson, 2001; Feero et al., 1995; Al-Ghamdi, 1999; Gonzalez et al., 2006; Gonzalez et al., 2009; Brown et al, 2000). These studies demonstrated that nature of injury, EMS response time and prehospital trauma care were the main factors affecting the time till death and concluded that timely EMS response with proper prehospital trauma care may improve the survival outcome. For analysis of the time to death data, these studies employed univariate statistical analysis (such as descriptive analysis or Fisher's exact test, Student t test). Most recently, Ju and Sohn (2014) analyzed the factors that are potentially associated with variation in the expected survival time by using Weibull regression approach and identified that survival probabilities and expected survival times are related to changes in delta V, alcohol involvement, and restraint systems. But, none of these studies investigate the timeline of death at the disaggregate level as a function of exogenous characteristics for a crash victim.

1.5.6 Data Pooling from Multiple Data Sources

The FARS database is a census (not a sample) of all fatal crashes in the US; i.e., crashes that lead to at least one fatality within thirty consecutive days from the time of crash, as discussed in the preceding section. The GES database, on the other hand, comprises a sample of road crashes across the US involving at least one motor vehicle travelling on a roadway and resulting in property damage, injury or death to the road users. The two datasets employed in the safety literature have their own advantages and limitations. The FARS focuses exclusively on fatal crashes. Therefore, one cannot reliably use this data to analyze the factors that increase or decrease the probability of fatality (because the data does not include crashes that do not lead to fatalities). The GES fills this gap by compiling data on a sample of roadway crashes involving all possible severity consequences (no injury, possible injury, non-incapacitating injury, incapacitating injury and fatality) providing a more representative sample of traffic crashes in the US. One of the advantages of FARS, however, is that the collected information includes the date and time of occurrence of the fatalities resulting within a 30-day time period from the crash. This detailed information provides us a continuous timeline of the fatal occurrences from the time to crash (instead of considering all fatalities to be the same). This allows for an analysis of the survival time of victims before their death. The GES, on the other hand, does not offer such detailed information except identifying who died in the crash. Examining the impact of various exogenous factors on all levels of injury severity as well as on the survival time of fatalities can potentially play a critical role in field triage - screening process to determine the more severe cases. While using the FARS data is very helpful for understanding the differences across different fatal crashes, it inherently excludes crashes with other possible, non-fatal injury severity outcomes. This makes it difficult to generalize the findings to the overall crash population. Besides, while analyzing the survival time of only fatal crash victims (using FARS data) helps in deriving the influence of various exogenous factors on survival time conditional upon the occurrence of a fatality, it doesn't allow the analyst to derive the influence of those factors in increasing the chances of survival. This is because the FARS data doesn't provide a representative sample of non-fatal crashes. One way to address this issue is combining information from both the FARS and GES datasets into a single, disaggregate crash-level database. This will bring together the strengths of both datasets – the representativeness of crashes with all injury severity outcomes from the GES data and the detailed information on fatal crashes from the FARS data. However, none of the earlier studies pooled FARS and GES datasets into a single, disaggregate crash level database that combines information from both the datasets. A pooled dataset would allow us to examine the whole spectrum of injury severity ranging from no injury to fatality, along with differentiating fatal crashes based on survival time. Moreover, the simultaneous interpretation of information would allow researchers to provide recommendations using a single modeling framework, rather than making inferences from the results of separate econometric models from different datasets[4].

1.6 Objectives

Literature in severity analysis is vast and growing. However, there are still several methodological and empirical gaps as has been discussed in the preceding section. The objective of the current dissertation is to develop advanced econometric frameworks to address these gaps in safety literature while employing these models developed to study important empirical issues. The specific objectives of the current dissertation are fivefold as discussed below:

The first objective is to evaluate the performance of alternate outcome frameworks for modeling driver injury severity. Specifically, the study provides a comprehensive comparison of ordered and unordered outcome models for examining the impact of exogenous factors on driver injury severity by using an observed and an underreported data samples. Further, we also evaluate the performance of model frameworks in the presence of underreporting – with and without corrections to the estimates.

The second objective is to formulate, estimate and validate econometric models accounting for systematic heterogeneity in the context of driver injury severity. Specifically, the study formulates and estimates latent segmentation based generalized ordered logit model. Moreover, it also compares the performance of the formulated model with its traditional counterparts to demonstrate the advantages of accommodating the effect of both observed and unobserved heterogeneity in examining driver injury severity.

[4] To be sure, the reader would note that there have been compilation of GES and FARS datasets to obtain the Annual Traffic Safety Facts (see NHTSA, 2012). However, in these efforts, there is no attempt to pool data from the two sources. The report provides trends separately for FARS and GES datasets.

The third objective is to apply and validate a multiple dependent variable model for accommodating endogeneity in the context of driver injury severity. Specifically, we develop a copula based joint modeling framework to study collision type and injury severity sustained as two dimensions of the severity process. Moreover, we enhance the copula based methodology by incorporating parameterization of dependency profile in an unordered and ordered joint structure.

The fourth objective is to examine fatality as a continuum rather than a single discrete state. Specifically, rather than homogenizing all fatal crashes as the same, we analyze the fatal injury from a new perspective and examine fatality as a continuous spectrum based on survival time ranging from a death occurring within thirty days of the crash up to instantaneous death. The disaggregate level models are estimated for the discrete representation of the continuous fatality timeline, while also accounting for endogeneity bias of EMS arrival time using ordered outcome modeling framework with endogeneity treatment.

The fifth objective is to develop a framework for pooling data from two crash datasets. Specifically, we propose and test the efficacy of a simple yet statistically valid approach to fuse two different datasets into a single, disaggregate crash level database that combines information from both the datasets. We also simultaneously examine the whole spectrum of injury severity by considering a very refined categorization of fatal crashes along with other non-fatal crashes.

1.7 Outline of the Dissertation

The remainder of the dissertation is divided into five chapters structured as follows:

Chapter two contributes to objective one and focuses on the relevance of alternate discrete outcome frameworks for modeling driver injury severity. The ordered outcome and unordered outcome models are empirically compared in the context of driver injury severity in traffic crashes. The alternative modeling approaches considered for the comparison exercise include: for the ordered outcome framework- ordered logit (OL), generalized ordered logit (GOL), mixed generalized ordered logit (MGOL) and for the unordered outcome framework - multinomial logit (MNL), nested logit (NL), ordered generalized extreme value logit (OGEV) and mixed multinomial logit (MMNL) model. A host of comparison metrics are computed to evaluate the performance of these alternative models. A comprehensive comparison exercise of the performance of ordered and unordered outcome models are provided for examining the impact of exogenous factors on driver injury severity. The effect of potential underreporting on alternative frameworks is also explored in this chapter by artificially creating an underreported data sample from the driver injury severity sample. The empirical analysis is based on the 2010 General Estimates System (GES) database of the US. The performance of the alternative frameworks are examined in the context of model estimation and validation (at the aggregate and disaggregate level). Further, the performance of the model frameworks in the presence of underreporting is explored – with and without corrections to the estimates. The empirical examination of alternative approaches in the context of injury severity analysis would allow us to determine the preferred model frameworks.

Chapter three contributes to objective two by formulating, estimating and validating an econometric model, referred to as the latent segmentation based generalized ordered logit (LSGOL) model, for examining driver injury severity. The proposed model probabilistically allocates drivers (involved in a crash) into different injury severity segments based on crash characteristics to recognize that the impacts of exogenous variables on driver injury severity level can vary across drivers based on both observed and unobserved crash characteristics. The model is estimated using Victorian Crash Database from Australia for the years 2006 through 2010.

Chapter four contributes to objective three by examining the hypothesis that collision type fundamentally alters the injury severity pattern under consideration. Towards this end, a copula based joint framework is employed that ties the collision type (represented as a multinomial logit model) and injury severity (represented as an ordered logit model) through a closed form flexible dependency structure to study the injury severity process. The proposed approach also accommodates the potential heterogeneity (across drivers) in the dependency structure. Further, the chapter incorporates collision type as a vehicle-level, as opposed to a crash-level variable as hitherto assumed in earlier research, while also examining the impact of a comprehensive set of exogenous factors on driver injury severity. The proposed modeling system is estimated using collision data from the province of Victoria, Australia for the year 2006 through 2010.

Chapter five focuses on objective four and examines fatality as a continuous spectrum based on survival time ranging from dying within thirty days of crash to dying instantly by using FARS database for the year 2010. The fatality continuum is represented as a discrete ordered dependent variable and analyzed using the mixed generalized ordered logit (MGOL) model. By doing so, we expect to provide a more accurate estimation of critical crash attributes that contribute to death. We also propose to estimate a two equation model that comprises of a regression equation for EMS response time and MGOL for fatality continuum with residuals from the EMS model to correct for endogeneity bias on the effect of exogenous factors on the timeline of death. Such research attempts are useful in determining what factors affect the time between crash occurrence and time of death so that safety measures can be implemented to prolong survival. The model estimates are augmented by conducting elasticity analysis to highlight the important factors affecting time-to-death process.

Chapter six contributes to objective five. This chapter focuses on developing a framework for pooling of data from FARS and GES data. The validation of the pooled sample against the original GES sample (unpooled sample) is carried out through two methods: (1) univariate sample comparison and (2) econometric model parameter estimate comparison. Generalized Ordered Logit (GOL) model (also referred to as Partial Proportional Odds model) is employed on the pooled dataset to analyze the influence of a variety of exogenous factors on traffic crash injury severity, while considering a very refined characterization of fatal crashes along with other, non-fatal injury severity outcomes. Finally, elasticity measures are also computed to identify important factors affecting vehicle occupant injury severity outcomes.

Chapter seven concludes the dissertation by summarizing the findings, and identifies directions for future research.

Table 1.1: Summary of Existing Driver Injury Severity Studies

Paper	Methodological Approach	Driver injury Severity Representation	Crash Attributes Considered: Driver Characteristics	Vehicle Characteristics	Roadway Design & Operational Attributes	Environmental Factors	Crash Characteristics
Shibata and Fukuda (1994)	Logistic Regression	Fatal; Non-fatal	Yes	-	-	-	Yes
Krull et al. (2000)	Logistic Regression	Fatal/Incapacitating Injury; Non-incapacitating/ Possible/ No injury	Yes	Yes	Yes	Yes	Yes
Toy and Hammitt (2003)	Logistic Regression	Serious injury/Death; Non-fatal	Yes	Yes	-	-	Yes
Conroy et al. (2008)	Logistic Regression	Severe injury	Yes	Yes	-	-	Yes
Fredette et al. (2008)	Logistic regression	Fatality, Major injury (hospitalized)	Yes	Yes	Yes	-	Yes
Bédard et al. (2002)	Multivariate Logistic Regression	Fatal; Non-fatal	Yes	Yes	-	-	Yes
Dissanayake and Lu (2002)	Sequential Binary Logistic Regression	No injury; Possible injury; Non-incapacitating injury; Incapacitating Injury; Fatality	Yes	-	Yes	Yes	Yes

Huang et al. (2008)	Bayesian Hierarchical Binomial Logistic Regression	Fatal/Severe injury; Slight/No injury	Yes	Yes	Yes	Yes	Yes
Khattak et al. (2002)	Ordered Probit	Fatality; Incapacitating injury; Evident injury; Possible injury	Yes	Yes	Yes	Yes	Yes
Kockelman and Kweon (2002)	Ordered Probit	No injury; Minor injury; Severe injury; Fatal injury	Yes	Yes	-	Yes	Yes
Abdel-Aty (2003)	Ordered Probit, Ordered Logit, Multinomial Logit, Nested Logit	Property damage only, Possible injuries, Evident injuries, Severe/fatal injuries	Yes	Yes	Yes	Yes	Yes
Khattak and Rocha (2003)	Ordered Logit	No injury; Minor injury; Moderate injury; Serious injury; Severe injury; Critical injury; Max injury	Yes	Yes	Yes	-	Yes
Kweon and Kockelman (2003)	Ordered Probit & Poisson Model	No injury; Not severe injury; Severe injury; Fatal injury	Yes	Yes	-	-	-
Khattak et al. (1998)	Binary Probit & Ordered Probit	Fatal; Severe injury; Moderate Injury; Minor injury	Yes	Yes	Yes	Yes	-
Yamamoto and Shankar (2004)	Bivariate ordered-response probit	Property damage only, Possible injury, Evident injury, Disabling injury, Fatality	Yes	Yes	Yes	Yes	Yes

Yamamoto et al. (2008)	Sequential Binary Probit Model; Ordered- Probit Model	Property damage only; Possible injury; Evident injury; Disabling injury; Fatality	Yes	Yes	Yes	Yes	Yes
Xie et al. (2009)	Bayesian Ordered Probit	No injury, Possible injury, Non-incapacitated injury, Capacitated injury, and Fatal injury	Yes	Yes	Yes	Yes	Yes
Eluru and Bhat (2007)	Mixed Joint Binary Logit-Ordered Logit	No injury; Possible injury; Non-incapacitating injury; Incapacitating injury; Fatal injury	Yes	Yes	Yes	Yes	Yes
Paleti et al. (2010)	Random Coefficients Heteroscedastic Ordered-Logit	No injury; Possible injury; Non-incapacitating injury; Incapacitating/Fatal injury	Yes	Yes	Yes	Yes	-
de Lapparent (2008)	Bivariate Ordered Probit	No injury; Light injury; Severe injury; Fatal injury	Yes	-	Yes	Yes	Yes
Srinivasan (2002)	Ordered Logit; Ordered Mixed Logit	No Injury/ Property Damage; Moderate injury; Severe injury; Fatal injury	Yes	Yes	-	Yes	Yes
Ulfarsson and Mannering (2004)	Multinomial Logit	No injury; Possible injury; Evident injury; Fatal/Disabling injury	Yes	Yes	Yes	Yes	Yes
Rana et al. (2010)	Copula-based Joint Ordered Logit–Ordered Logit; Copula-Based Joint Multinomial	No injury; Possible injury; Non-incapacitating injury; Incapacitating injury; Fatal injury	Yes	Yes	Yes	Yes	Yes

	Logit–Ordered Logit						
Eluru et al. (2012)	Latent Segmentation Based Ordered Logit	No injury; Injury; Fatal injury	Yes	Yes	Yes	Yes	-
Eluru et al. (2010)	Copula Based Approach	No injury; Possible injury; Non-incapacitating injury; Incapacitating/ Fatal injury	Yes	Yes	Yes	Yes	Yes
Khorashadi et al. (2005)	Multinomial Logit	No injury; Complaint of pain; Visible injury; Severe/Fatal injury	Yes	Yes	Yes	Yes	Yes
Islam and Mannering (2006)	Multinomial Logit	No injury; Injury; Fatality	Yes	Yes	Yes	Yes	Yes
Awadzi et al. (2008)	Multinomial Logit	No injury; Injury; Fatality	Yes	Yes	Yes	Yes	Yes
Schneider et al. (2009)	Multinomial Logit	Property damage only; Possible injury; Non-incapacitating injury; Incapacitating injury; Fatal	Yes	Yes	Yes	Yes	Yes
Morgan and Mannering (2011)	Mixed Multinomial Logit	Severe injury, Minor injury, No injury	Yes	Yes	Yes	Yes	Yes
Kim et al. (2013)	Mixed Multinomial Logit	Fatal injury, Severe injury, Visible injury, Complaint of pain/no injury	Yes	Yes	-	Yes	Yes

Xie et al. (2012)	Latent Class Logit	No injury; Possible injury; Non-incapacitating injury; Incapacitating injury; Fatal injury	Yes	Yes	Yes	Yes	Yes

CHAPTER 2 Evaluating Alternate Discrete Outcome Frameworks for Modeling Crash Injury Severity

2.1 Introduction

The applicability of ordered and unordered frameworks for analyzing ordinal discrete variables has evoked considerable debate on using the appropriate model for analysis. Yet, there is little research on empirically examining the differences between the ordered and unordered frameworks, specifically in the context of crash injury severity. There are many strengths and weaknesses for the ordered framework vis-à-vis the unordered framework (Eluru, 2013). The ordered outcome models explicitly recognize the inherent ordering within the decision variable whereas the unordered outcome models neglect the ordering or require artificial constructs to consider the ordering (for example the ordered generalized extreme value logit model). On the other hand, the traditional ordered outcome models restrict the impact of exogenous variables on the outcome process to be same across all alternatives while the unordered outcome models allow the model parameters to vary across alternatives (see Eluru et al., 2008 for a discussion). The restricted number of parameters ensures that ordered outcome models have a parsimonious specification. The unordered outcome models might not be as parsimonious but offer greater explanatory power because of the additional exogenous effects that can be explored. In fact, several studies highlight the advantages of multinomial logit model over the ordered outcome models (see for example Bhat and Pulugurta, 1998). Hence, an empirical examination of alternative approaches in the context of injury severity analysis will allow us to determine the preferred model. Yet, there is little research on empirically examining the differences between the ordered and unordered frameworks (except Abdel-Aty, 2003; Ye and Lord, 2011). Further, the recent revival of generalized ordered logit (GOL) model (proposed by Terza, 1985) offers an ordered framework that allows the analyst to estimate the same number of parameters as the multinomial logit for an ordinal discrete variable. Hence, an exercise comparing the alternative frameworks is incomplete without considering GOL. The GOL framework with its improved flexibility will provide the true benchmark for a fair comparison.

An accurate estimation of the associated risk factors is critical to assist decision makers, transportation officials, insurance companies, and vehicle manufacturers to make informed decisions to improve road safety. Yet, there is little research on empirically examining the differences between the ordered and unordered frameworks. Further, the influence of underreporting on alternative model frameworks has also received little attention. Given the significance of examining the influence of exogenous variables on injury severity it is important that we undertake a comparison based on the performance of alternative frameworks. Towards that end, the current chapter proposes a framework to compare and contrast the alternative frameworks available for modeling driver injury severity. However, the conventional police/hospital reported crash databases may not include precious behavioural, physiological and psychological characteristics of individual involved in collisions. Due to the presence of such unobserved information, the effect of exogenous variables might not be the same across individuals in the event of a crash (see for example Srinivasan, 2002; Eluru et al., 2008; Morgan and Mannering, 2011; Kim et al., 2013). For example, careful driving on behalf of a safe driver might moderate the severity outcome of a crash during night-time and while less cautious driving of an aggressive driver might exacerbate the crash severity in the same situation. In non-linear models, neglecting the effect of such unobserved heterogeneity can result in inconsistent estimates (Chamberlain, 1980; Bhat, 2001). Our study also incorporates the influence of unobserved heterogeneity in both the ordered and unordered outcome frameworks. Further, the chapter also incorporates the underreporting issue associated with traditional crash databases. Specifically, the current research examines the performance of alternative modeling frameworks in the context of estimation from an observed sample and also in the context of an artificially created underreported data sample. In doing so, the study generates elasticity measures for the true and underreported samples to illustrate the influence of underreporting. The parameters from these model estimations are also used on a validation hold-out sample to evaluate model predictions (in the true as well as underreported case). The alternative modeling approaches considered for the exercise include: for the <u>ordered outcome framework</u> - ordered logit (OL), generalized ordered logit (GOL), mixed generalized ordered logit (MGOL) and for the <u>unordered outcome framework</u> - multinomial logit (MNL), nested logit (NL), ordered generalized extreme value logit (OGEV) and mixed multinomial logit (MMNL) model. We generate a series of measures to evaluate model performance in estimation and prediction thus allowing us to draw conclusions on model applicability for injury severity analysis.

In summary, the current chapter contributes to literature on driver injury severity in multiple ways. <u>First</u>, it provides a comparison exercise of the performance of ordered and unordered outcome models by employing a host of comparison metrics for examining the impact of exogenous factors on driver injury. Second, we compare the performance of the various models in the presence of underreporting. Finally, we undertake the examination of driver injury severity using a comprehensive set of exogenous variables.

The rest of the chapter is organized as follows. Section 2.2 provides details of the various econometric model frameworks used in the analysis. In Section 2.3, the data source and sample formation procedures are described. The model comparison results, elasticity effects and validation measures are presented in Sections 2.4 and 2.5. Section 2.6 concludes the chapter.

2.2 Econometric Framework

In this section, we provide a brief description of the methodology of all the models considered for examining driver injury severity in this research.

2.2.1 Standard Ordered Logit Model

In the traditional ordered models, the discrete injury severity levels (y_i) are assumed to be associated with an underlying continuous latent variable (y_i^*). This latent variable is typically specified as the following linear function:

$$y_i^* = \boldsymbol{X}_i\boldsymbol{\beta} + \varepsilon_i, for\ i = 1,2, \dots \dots \dots, N \qquad (2.1)$$

where,

i ($i = 1,2, \dots \dots \dots, N$) represents the drivers

$\boldsymbol{X}_i$ is a vector of exogenous variables (excluding a constant)

$\boldsymbol{\beta}$ is a vector of unknown parameters to be estimated

ε is the random disturbance term assumed to be standard logistic

Let j ($j = 1,2, \dots \dots \dots, J$) denotes the injury severity levels and τ_j represents the thresholds associated with these severity levels. These unknown τ_js are assumed to partition the propensity into $J - 1$ intervals. The unobservable latent variable y_i^* is related to the observable ordinal variable y_i by the τ_j with a response mechanism of the following form:

$$y_i = j,\ if\ \tau_{j-1} < y_i^* < \tau_j, for\ j = 1,2, \dots \dots \dots, J \qquad (2.2)$$

In order to ensure the well-defined intervals and natural ordering of observed severity, the thresholds are assumed to be ascending in order, such that $\tau_0 < \tau_1 < \ldots\ldots\ldots < \tau_J$ where $\tau_0 = -\infty$ and $\tau_J = +\infty$. Given these relationships across the different parameters, the resulting probability expressions for individual i and alternative j for the OL take the following form:

$$\pi_{ij} = Pr(y_i = j|X_i) = \Lambda\left(\tau_j - \boldsymbol{X_i\beta}\right) - \Lambda\left(\tau_{j-1} - \boldsymbol{X_i\beta}\right) \tag{2.3}$$

where $\Lambda(.)$ represents the standard logistic cumulative distribution function.

2.2.2 Generalized Ordered Logit Model

The GOL model relaxes the constant threshold across population restriction to provide a flexible form of the traditional OL model. The basic idea of the GOL is to represent the threshold parameters as a linear function of exogenous variables (Maddala, 1983; Terza, 1985; Srinivasan, 2002; Eluru et al., 2008). Thus the thresholds are expressed as:

$$\tau_j = fn(Z_{ij}) \tag{2.4}$$

where, Z_{ij} is a set of exogenous variable (including a constant) associated with j th threshold. Further, to ensure the accepted ordering of observed discrete severity $\left(-\infty < \tau_1 < \tau_2 < \ldots\ldots\ldots < \tau_{J-1} < +\infty\right)$, we employ the following parametric form as employed by Eluru et al. (2008):

$$\tau_j = \tau_{j-1} + exp(\boldsymbol{\delta_j Z_{ij}}) \tag{2.5}$$

where, $\boldsymbol{\delta}_j$ is a vector of parameters to be estimated. The remaining structure and probability expressions are similar to the OL model. For identification reasons, we need to restrict one of the $\boldsymbol{\delta_j}$ vectors to zero.

2.2.3 Mixed Generalized Ordered Logit Model

The MGOL accommodates unobserved heterogeneity in the effect of exogenous variable on injury severity levels in both the latent injury risk propensity function and the threshold functions (Srinivasan, 2002; Eluru et al., 2008). Let us assume that $\boldsymbol{\alpha}_i$ and $\boldsymbol{\gamma}_{ij}$ are two column vectors representing the unobserved factors specific to driver i and his/her trip environments in equation 2.1 and 2.5, respectively. Thus the equation system for MGOL model can be expressed as:

$$y_i^* = (\boldsymbol{\beta} + \boldsymbol{\alpha}_i)\boldsymbol{X}_i + \varepsilon_i, for\ i = 1,2, \dots \dots \dots, N \tag{2.6}$$

and

$$\tau_{i,j} = \tau_{i,j-1} + exp[(\boldsymbol{\delta_j} + \boldsymbol{\gamma}_{i,j})\ \boldsymbol{Z}_{i,j}] \tag{2.7}$$

In equations 2.6 and 2.7, we assume that $\boldsymbol{\alpha}_i$ and $\boldsymbol{\gamma}_{ij}$ are independent realizations from normal distribution for this study. Thus, conditional on $\boldsymbol{\alpha}_i$ and $\boldsymbol{\gamma}_{ij}$, the probability expressions for individual i and alternative j in MGOL model take the following form:

$$\begin{aligned} \pi_{ij} &= Pr\left(y_i = j \middle| \boldsymbol{\alpha}_i, \boldsymbol{\gamma}_{ij}\right) \\ &= \Lambda[\tau_{i,j-1} + exp((\boldsymbol{\delta}_j + \boldsymbol{\gamma}_{i,j})\ \boldsymbol{Z}_{i,j}) - (\boldsymbol{\beta} + \boldsymbol{\alpha}_i)\boldsymbol{X}_i] - \Lambda[\tau_{i,j-2} \\ &\qquad + exp((\boldsymbol{\delta}_{j-1} + \boldsymbol{\gamma}_{i,j-1})\ \boldsymbol{Z}_{i,j}) - (\boldsymbol{\beta} + \boldsymbol{\alpha}_i)\boldsymbol{X}_i] \end{aligned} \tag{2.8}$$

The unconditional probability can subsequently be obtained as:

$$P_{ij} = \int_{\boldsymbol{\alpha}_i, \boldsymbol{\gamma}_{ij}} [Pr\left(y_i = j \middle| \boldsymbol{\alpha}_i, \boldsymbol{\gamma}_{ij}\right)] * \boldsymbol{dF}\left(\boldsymbol{\alpha}_i, \boldsymbol{\gamma}_{ij}\right)\boldsymbol{d}(\boldsymbol{\alpha}_i, \boldsymbol{\gamma}_{ij}) \tag{2.9}$$

In this study, we use a quasi-Monte Carlo (QMC) method proposed by Bhat (2001) for discrete outcome model to draw realization from its population multivariate distribution. Within the broad framework of QMC sequences, we specifically use the Halton sequence (200 Halton draws) in the current analysis (see Eluru et al., 2008 for a similar estimation process).

2.2.4 Multinomial Logit Model

Let us consider the probability of a driver i ending in a specific injury-severity level j. The alternative specific latent variables for MNL take the form of:

$$U_{ij} = \boldsymbol{\beta_j} \boldsymbol{X_{ij}} + \varepsilon_{ij} \tag{2.10}$$

where

$\boldsymbol{\beta_j}$ is a vector of coefficients to be estimated for outcome j

$\boldsymbol{X_{ij}}$ is a vector of exogenous variables

U_{ij} is a function of covariates determining the severity

ε_{ij} is the random component assumed to follow a gumbel type 1 distribution.

Thus, the MNL probability expression is as follows:

$$P_i(j) = \frac{exp[\boldsymbol{\beta_j X_{ij}}]}{\sum_{j=1}^{J} exp[\boldsymbol{\beta_j X_{ij}}]} \tag{2.11}$$

2.2.5 Nested Logit Model

The NL model allows the incorporation of correlation across alternatives and results in two kinds of alternatives: those that are part of a nest (*i.e.* alternatives that are correlated) and alternatives that are not part of nest. The crash severity probabilities for the nested alternatives in the NL are composed of the nest probability as well as the alternative probability (same structure as the MNL applies).

In the first step, the probability of choosing the nest is determined followed by the probability of choosing alternative within the nest

$$P_i(j) = \frac{exp[\boldsymbol{\beta_j X_{ij}} + \boldsymbol{\theta_j L_{ij}}]}{\sum_{j \in J} exp[\boldsymbol{\beta_j X_{ij}} + \boldsymbol{\theta_j L_{ij}}]}$$

$$P_i(k|j) = \frac{exp[\boldsymbol{\beta_{k|j} X_{ij}}]}{\sum_{k \in K} exp[\boldsymbol{\beta_{k|j} X_{ij}}]} \tag{2.12}$$

where,

$P_i(j)$ is the unconditional probability of ith crash falling in nest j

$P_i(k|j)$ is the conditional probability of ith crash having severity outcome k (lower level) conditioned on the nest j (higher level)

J is the actual severity and K is the alternative represented by the nest

$\boldsymbol{L_{ij}}$ is the inclusive value (log sum) representing the expected value of the attributes from the nest j

$\boldsymbol{\theta_j}$ is the nesting coefficient

The alternative probabilities for non-nested alternatives take a form similar to the MNL probabilities while considering the utility of the nested alternatives as a composite alternative. To be consistent with the NL derivation, the value of the $\boldsymbol{\theta_j}$ should be greater than 0 and less than 1 (McFadden, 1981). If the estimated value of $\boldsymbol{\theta_j}$ is not significantly different from 1, then the NL model collapses to a simple MNL model.

2.2.6 Ordered Generalized Extreme Value Model

Injury levels of a crash are typically progressive (ranging from non-injury to fatal). MNL and NL models do not account for any inherent ordering in the outcomes. Small (1987) proposed the OGEV model for such ordered discrete outcomes. The OGEV model allows for the correlations between the error terms of outcomes which are close to each other in the ordered scale.

We employ the structure proposed in Wen and Koppelman (2001) for the OGEV model with j alternatives as follows:

$$P_i(j) = \sum_{m=i}^{i+L} P_{i|m}.P_m$$

$$= \sum_{m=i}^{i+L}\left[\frac{(w_{mi}.e^{U_{ij}})^{1/\mu_m}}{\sum_{k\epsilon N_m}(w_{mj}.e^{U_{ik}})^{1/\mu_m}} * \frac{\left\{\sum_{k\epsilon N_m}(w_{mj}.e^{U_{ik}})^{1/\mu_m}\right\}^{\mu_m}}{\sum_{s=1}^{J+L}\left\{\sum_{k\epsilon N_s}(w_{sj}.e^{U_{ik}})^{1/\mu_s}\right\}^{\mu_s}}\right] \tag{2.13}$$

The probability of alternative j in a crash for driver i is computed as the sum of probability computed from all nests to which i belongs. In the above notation, L is the number of contiguous alternatives considered in a nest, w_{mi} represents the allocation weight for each alternative i to nest m, The total number of nests is given as a combination ${}^{J}C_L$. The allocation parameter satisfies the property $\sum_i w_{mi}$=1. μ_m represents the log-sum parameter for nest m. N_m represents the set of alternatives in nest m. In our analysis we set L = 1 *i.e.* we consider the following nests 1, 1 2, 2 3, 3 4, and 4 (where 1= No Injury, 2= Possible Injury, 3= Non-incapacitating Injury and 4= Incapacitating/Fatal Injury).

2.2.7 Mixed Multinomial Logit Model

The MMNL is a generalized version of traditional MNL model. It allows the parameters for exogenous variables to vary across individual involved in the collision by accommodating unobserved heterogeneity on the utility functions for different injury severity levels. Let us assume that $\boldsymbol{\omega}_{ij}$ is a column vectors representing the unobserved factors specific to driver i and his/her trip environments in equation 2.10. Thus the equation system for MMNL model can be expressed as:

$$U_{ij} = (\boldsymbol{\beta_j}+\boldsymbol{\omega}_{ij})\boldsymbol{X_{ij}} + \varepsilon_{ij} \tag{2.14}$$

In equation 2.14, we assume that $\boldsymbol{\omega}_{ij}$ is an independent realization from normal distribution for this study. Thus, conditional on $\boldsymbol{\omega}_{ij}$, the probability expression for individual i and alternative j in MMNL model take the following form:

$$P_{ij}|\boldsymbol{\omega}_{ij} = \frac{exp[(\boldsymbol{\beta_j}+\boldsymbol{\omega}_{ij})\boldsymbol{X_{ij}}]}{\sum_{j=1}^{J} exp[(\boldsymbol{\beta_j}+\boldsymbol{\omega}_{ij})\boldsymbol{X_{ij}}]} \tag{2.15}$$

The unconditional probability can subsequently be obtained as:

$$P_{ij} = \int_{\boldsymbol{\omega}_{ij}} (P_{ij}|\boldsymbol{\omega}_{ij}) * dF(\boldsymbol{\omega}_{ij})d\boldsymbol{\omega}_{ij} \tag{2.16}$$

To estimate the MMNL model, we apply the QMC simulation techniques in a similar fashion as described in MGOL model section.

2.3 Data

2.3.1 Data Source

The data for the current chapter is sourced from the "General Estimates System (GES)" database for the year 2010. The GES database is a nationally representative sample of road crashes collected and compiled from about 60 jurisdictions across the US. The data is obtained from the US Department of Transportation, National Highway Traffic Safety Administration's National Center for Statistics and Analysis (ftp://ftp.nhtsa.dot.gov/GES/GES10/). The data includes information of reports compiled by police officers for crashes involving at least one motor vehicle travelling on a roadway and resulting in property damage, injury or death to the road users. The GES crash database has a record of 46,391 crashes involving 81,406 motor vehicles and 116,020 individuals for the year of 2010. A five point ordinal scale is used in the database to represent the injury severity of individuals involved in these crashes: 1) No injury; 2) Possible injury; 3) Non-incapacitating injury; 4) Incapacitating injury and 5) Fatal injury. Further, the dataset compiles information on a multitude of factors (driver characteristics, vehicle characteristics, roadway design and operational attributes, environmental factors and crash characteristics) representing the crash situations and events. Accordingly, a number of crash-related factors are extracted from this database in order to explore the variables that might influence the driver injury severity.

2.3.2 Sample Formation and Description

The main focus of this study is injury severity of drivers of passenger vehicles (passenger car, sport utility vehicle, pickup or van). Thus, the following criteria were employed for sample formation:

- The crashes that involve only non-commercial (private) passenger vehicle drivers are selected (to avoid the potential systematic differences between commercial and non-commercial driver groups).
- The passenger vehicle crashes that involve another passenger vehicle or a fixed object are examined.
- The crashes that involve more than two vehicles are excluded from the analysis.

The final dataset of non-commercial driver of passenger vehicles, after removing records with missing information for essential attributes consisted of about 30,371 records. In this final sample of crashes the percentage of fatal crashes sustained by drivers is extremely small (0.7%). Therefore, both the fatal and incapacitating injury categories are merged together to ensure a representative share for each alternative crash level. From this dataset, a sample of 12,170 records is sampled out for the purpose of model estimation and 18,201 records are set aside for validation. In the final estimation sample, the distributions of driver injury severities are: no injury 65.9%, possible injury 15.1%, non-incapacitating injury 12.1 % and incapacitating/fatal injury 6.9%.

2.4 Empirical Analysis

2.4.1 Variables Considered

In our analysis of this chapter, we selected a host of variables from five broad categories: Driver characteristics (including driver gender, driver age, restraint system use, alcohol consumption and drug use), Vehicle characteristics (including vehicle type and vehicle age), Roadway design and operational attributes (including roadway class, speed limit, types of intersection and traffic control device), Environmental factors (including time of day and road surface condition) and Crash characteristics (including driver ejection, vehicle rolled over, air bag deployment, manners of collision and collision location). It should be noted here that several variables such as presence of shoulder, shoulder width, point of impact, number of lanes, lighting condition could not be considered in our analysis because either the information was entirely unavailable or there was a large fraction of missing data for these attributes in the dataset. To be sure, we employ the manner of collision and time of day variables to act as surrogates for point of impact and lighting condition, respectively. In the final specification of the model, statistically insignificant variables were removed (95% confidence level). Further, in cases where the variable effects were not significantly different, the coefficients were restricted to be the same.

2.4.2 Overall Measures of Fit

In the research effort of this chapter, we estimated seven different models: 1) OL, 2) GOL, 3) MGOL, 4) MNL, 5) OGEV, 6) NL and 7) MMNL model. After extensively testing for different nesting structures for NL and parametric assumptions for OGEV models we found that these models collapsed to the MNL model. Hence, the entire comparison exercise is focussed on five models: OL, GOL, MGOL, MNL and MMNL. Prior to discussing the estimation results, we compare the performance of these models in this section.

The log-likelihood values at convergence for the various frameworks are as follows: (1) OL (with 29 parameters) is -10617.51; (2) GOL (with 50 parameters) is -10517.83, (3) MGOL (with 55 parameters) is -10506.97, (4) MNL (with 57 parameters) is -10517.59 and (5) MMNL (with 61 parameters) is -10508.76. The corresponding value for the "constant only" model is -12164.58. We can compare the ordered models (OL, GOL and MGOL) among those by using likelihood ratio (LR) test for selecting the preferred model. Similarly, the MNL and MMNL models can be compared using LR test. However, to compare the ordered approaches with the unordered approach, the LR test is not appropriate because these structures are not nested within one another. Hence, to undertake the comparison we employ a two-step process. In the first step, we use the LR test to determine the superior model within each framework. Subsequently, we compare the best model from each framework using the non-nested measures applicable for such comparison.

2.4.2.1 Comparison within Ordered and Unordered Frameworks

The LR test statistic is computed as:

$$LR = 2[LL_U - LL_R] \quad (2.17)$$

where LL_U and LL_R are the log-likelihood of the unrestricted and the restricted models, respectively. The computed value of the LR test is compared with the $\aleph^2$ value for the corresponding degrees of freedom (*dof*). The resulting LR test values for the comparison of OL/GOL, OL/MGOL and GOL/MGOL models are 199.36 (21 *dof*), 221.08 (26 *dof*) and 21.72 (5 *dof*), respectively. The LR test values indicate that MGOL outperforms the OL model at any level of statistical significance. The MGOL outperforms the GOL model at the 0.001 significance level indicating that MGOL offers superior fit compared to both OL and GOL models. In the unordered context, the LR test value (17.66, 4 *dof*) for the comparison of MNL/MMNL indicates that MMNL offers superior fit over MNL model at the 0.001 significance level.

2.4.2.2 Comparison between Ordered and Unordered Frameworks - Non-nested Test

To evaluate the performance of the ordered and unordered models, we employ different measures that are routinely applied in comparing econometric models including: 1) Bayesian Information Criterion (BIC), 2) Akaike Information Criterion corrected (AICc)[5] and 3) Ben-Akiva and Lerman's adjusted likelihood ratio (BL) test. The BIC for a given empirical model is equal to:

$$BIC = -2LL + K\, ln(Q) \quad (2.18)$$

Also, the AICc values are computed for each of the four models as:

$$AIC_c = 2K - 2ln(LL) + \frac{2\,K(K+1)}{(Q - K - 1)} \quad (2.19)$$

where LL is the log-likelihood value at convergence, K is the number of parameters, and Q is the number of observations. The model with the *lower* BIC and AICc values is the preferred model. The BIC (AICc) values for the final specifications of the MGOL and MMNL models are 21531.31 (21124.45) and 21591.33 (21140.14), respectively.

The BL test statistic (Ben-Akiva and Lerman, 1985) is computed as:

$$\lambda = \Phi\left\{-\left[\sqrt{-2(\bar{\rho}_2^2 - \bar{\rho}_1^2)L(C) + (M_2 - M_1)}\right]\right\} \quad (2.20)$$

[5] AICc is a more stringent version of the AIC [AIC = 2K− 2ln(L)] in penalizing for additional parameters

where $\overline{\rho}^2$ represents the McFadden's adjusted rho-square value for the model. It is defined as $\bar{\rho}_2^2 = 1 - \frac{L_i(\beta)-M_i}{L(C)}$, where $L_i(\beta)$ represents log-likelihood at convergence for the i^{th} model, $L(C)$ represents log-likelihood at sample shares and M_i is the number of parameters in the model (Windmeijer, 1995). The $\Phi(.)$ represents the cumulative standard normal distribution function. The resulting λ value for the comparison of MGOL and MMNL is 0, clearly indicating that MGOL offers superior fit compared to MMNL model. The comparison exercise clearly highlights the superiority of the MGOL in terms of data fit compared to MMNL model. In the subsequent section, we discuss the results from MGOL and MMNL frameworks.

2.4.3 Estimation Results

Table 2.1 presents the results of the MGOL and MMNL models. The reader would note that the interpretation of the MGOL is slightly different from the MMNL model. In MGOL, when the threshold parameter is positive (negative) the result implies that the threshold is bound to increase (decrease); the actual effect on the probability is quite non-linear and can only be judged in conjunction with the influence of the variable on propensity and other thresholds. MMNL represents the effect of exogenous variables on each injury category relative to the base category. In the following sections, the estimation results are discussed by variable groups.

2.4.3.1 Driver Characteristics

In safety research, driver demographics, particularly driver's age and gender have always been considered to have a significant influence on injury severity. In the current research, the effects of these variables are found to be significant. In particular, MGOL estimates indicate that compared to the female drivers, the latent injury propensity is lower for male drivers, while the negative sign of threshold demarcating the possible and non-incapacitating injury indicates a higher likelihood of non-incapacitating and incapacitating/fatal injuries for the male drivers. It is important to note that the variable impacts in propensity and thresholds are counteracting one another and the exact impact realized is specific to every individual. Corresponding results from MMNL indicate that male drivers are more likely to evade injury relative to their counterparts. The estimates associated with driver age, from both the MGOL and MMNL, suggest a reduction in the likelihood of severe injuries for the young drivers (age<25) compared to middle-aged drivers (age 25 to 64). However, the parameter characterizing the effect of older age (age≥65) on driver injury severity is found significant in the MMNL model only. The result suggests that the odds of suffering an incapacitating/fatal injury are significantly higher for the older drivers compared to the middle-aged drivers.

Seat belt use is found to have a significant influence on driver injury severity. Consistent with several previous studies (Preusser et al., 1991; Janssen, 1994; Eluru and Bhat, 2007), our analysis showed an unequivocal benefit for employing seat belts. MGOL model estimates for the driver not wearing safety belts results in a parameter that is normally distributed with a mean 1.528 and standard deviation 0.844, which indicates that almost 96% of the drivers involved in the collision cannot evade injury if they do not wear seat belts at the time of crash. MMNL model estimates indicate that the likelihood of suffering from possible, non-capacitating and incapacitating/fatal injuries is higher for the unrestrained driver and these effects are fixed.

As expected, drivers under the influence of alcohol are likely to have a higher injury risk propensity compared to the sober drivers. Positive sign of the latent propensity of MGOL model estimate indicates that the latent injury risk propensity is higher for drivers who are impaired by alcohol, while the negative sign of threshold demarcating the non-incapacitating and incapacitating/fatal injury indicates a higher likelihood of incapacitating/fatal injury for this group of drivers. MMNL model estimates also reveal that the odds of suffering incapacitating/fatal injury are higher for non-sober drivers. The effect of impairment by drugs is found significant in MMNL model only and the result shows that the drivers are more likely to suffer an incapacitating/fatal injury when they are impaired by drugs. The MGOL model is unable to pick such an effect of drugs involvement on driver injury severity and the reason might be attributed to a small share (0.9%) of drivers under the influence of drug in the dataset.

2.4.3.2 Vehicle Characteristics

With respect to driver's vehicle type, the MGOL model results indicate that the latent injury propensity is higher for the driver of a passenger car compared to the driver of other passenger vehicles (sports utility vehicle (SUV), pickup and vans). This is expected because in collisions with other vehicles or fixed objects, the drivers in passenger cars are usually the most likely to be severely injured (Mayrose and Jehle, 2002; O'Neill and Kyrychenko, 2004; Fredette et al., 2008). The corresponding results from MMNL suggest that the likelihood of sustaining possible, non-capacitating and incapacitating/fatal injuries is higher for the drivers in a passenger car relative to drivers in other passenger vehicles.

The vehicle age result of MGOL model demonstrates that the drivers in older vehicles (6-10 years and above 10 years) have a higher injury risk propensity compared to the drivers in newer vehicles (vehicle age<6 years). The MMNL model estimates indicate that the drivers in older vehicles (6-10 years old and above 10 years old) have a higher likelihood of suffering from possible, non-capacitating and incapacitating/fatal injuries relative to the drivers in newer vehicles. The higher injury risk of older vehicle's driver might be attributed to the mechanical defect, lack of safety equipment, exposure of younger driver to these vehicles or the involvement of suspended and unlicensed drivers of these vehicles (Lécuyer and Chouinard, 2006). The lower injury risk for the driver of new vehicles may reflect the advancement in the vehicle-based safety equipment (such as airbag, antilock braking system, center high-mounted stoplight, crash cage, shatter resistant windshield).

2.4.3.3 Roadway Design and Operational Attributes

With respect to the roadway functional class, the MGOL model estimates show that the injury risk propensity of drivers is higher when the crash occurs on an interstate highway. Again, the effect of "interstate highway" variable on the threshold demarcating non-incapacitating and incapacitating/fatal injuries shows a higher likelihood of incapacitating/fatal injuries of the drivers during crashes on an interstate highway. The MMNL model estimates show that the likelihood of both possible and incapacitating/fatal injury increases when crash occur on interstate highway. The MGOL results for speed limit indicate that latent injury propensities are higher for crashes occurring on roads with medium (26 to 50 mph) and higher (above 50 mph) speed limits relative to crashes on lower speed limit (less than 26 mph). The effect of speed limit variables on the threshold indicates increased likelihood of non-incapacitating and incapacitating/fatal injuries at higher speed limits. The corresponding results from MMNL suggest that the likelihood of sustaining possible, non-incapacitating and incapacitating/fatal injuries is higher for crashes on both the medium and higher speed limit roads compared to the crashes on lower speed limit roads. As is expected, within the two speed categories considered the higher speed category has a larger impact relative to the medium speed limit category.

With respect to the types of intersection, only four way intersections are found to have significant influence on driver injury severity. The MGOL model estimates reflect the higher injury risk propensity to drivers on a four-way intersection. The MMNL results also indicate very similar impact of four-way intersection on injury severity. The four way intersection reduces the likelihood of no injury crashes and in turn increases the likelihood of a driver sustaining severe injury. The presence of traffic control device is also found to have significant effect on the severity of crashes. MGOL estimates reveal that the presence of a traffic signal/stop/yield sign reduces the likelihood of injury risk propensity of the drivers relative to the absence of a control measure. The MMNL estimates show that the likelihood of non-incapacitating injury reduces with the presence of a traffic signal/stop/yield sign. However, MGOL estimates also indicate that the injury risk propensity increases when there are other traffic control system or a warning sign present on the roadway. The corresponding result of MMNL specify that the odds of suffering an incapacitating/fatal injury increase significantly with the presence of these control measures relative to uncontrolled measure.

2.4.3.4 Environmental Factors

Time-of-day and surface condition are two of the environmental factors that are found to significantly influence driver injury severity. Compared to the evening peak, the likelihood of injury risk propensities are found to be higher for both the morning peak and off-peak periods in the MGOL estimates. At the same time, the effect of night-time variable on the threshold demarcating possible and non-incapacitating injuries shows a higher likelihood of non-incapacitating and incapacitating/fatal injuries. The MMNL estimates reveal that the drivers are less likely to evade no injury during morning peak and off-peak period. However, the effect of night-time variable results in an estimate that is normally distributed with 0.032 and standard deviation 0.772. But, the mean coefficient for night-time is not significantly different from zero, while the standard deviation is highly significant. This result indicates that driver injury severity outcome varies widely during night-time crash and the exact nature of injury severity is determined by the unobserved factors specific to the crash.

The findings of MGOL estimates indicate that if collisions occur on a snowy road surface, the consequence is likely to be less injurious as compared to the accident on dry road surface. The MMNL results also indicate very similar impacts of snowy road surface on driver injury severity.

On a snowy road the drivers are more likely to evade serious injury relative to crashes on a dry surface. The effect of wet road surface condition is found significant only in the MMNL model estimates and the result indicates a lower likelihood of non-incapacitating injury on wet roads. The reduced risk of injury on snowy/wet road can be attributed to more careful driving and reduced speeding possibility (Edwards, 1998; Mao et al., 1997; Eluru and Bhat, 2007).

2.4.3.5 Crash Characteristics

Several crash characteristics considered are found to be significant determinants of driver injury severity. Among those, the injury risk propensities are observed to be higher in MGOL estimates when a driver is ejected out from his/her vehicle or when the vehicle rolled over. At the same time, the positive values of the first thresholds of driver ejection reflect an increase in possible injury probability. But, the first threshold of vehicle rolled over is found to be random with a statistically insignificant mean and a highly significant standard deviation. The result indicates that while injury risk propensity is likely to increase the impact on crash severity, the threshold is determined by unobserved factors specific to the crash.

The likelihood of injury risk propensity for the deployment of air bag is also found to be significant and normally distributed in the MGOL model estimate. The result implies that air bag deployment increases the probability of injury in almost 97% cases. At the same time, the positive values of the first thresholds of air bag deployment reflect an increase in possible injury probability. The corresponding results from the MMNL model estimates indicate that the drivers are less likely to avoid serious injury when the vehicle rolled over or an air bag deployed during a crash. However, none of the aforesaid two variable estimates are found to be random, while the effect of driver ejection is found to be insignificant both as fixed and random parameter in MMNL.

With respect to the collision object, MGOL and MMNL model estimates indicate very similar effects indicating that the odds of suffering serious injury is higher when a vehicle strikes a stationary object (such as: pole, guard rail, tree and post) compared to the crashes with a moving vehicle. However, the threshold demarcating non-incapacitating injury to incapacitating/fatal injury of MGOL is distributed normally. With the estimated parameter, 39.36% of the distribution is greater than zero and 60.64% of the distribution is less than zero. At the same time, MMNL model also results in a random parameter for incapacitating/fatal injury category, which indicates that 82.12% of the distribution is above zero and only 17.88% is less than zero. The parameters characterizing the effects of manner of collision in Table 2.1, for both MMNL and MGOL models, suggest that the drivers are less likely to evade serious injury in the event of head-on or angular collision relative to the rear-end collision. Side-swipe collisions with vehicles travelling in the same direction and rear to sideswipe collisions are less severe than rear-end collision.

Finally, both the MGOL and MMNL model estimates indicate that collision location has a significant influence on injury severity profile. Specifically, collisions at an intersection or entry/exit ramp or driveway access or intersection related collisions are less likely to result in injuries to the drivers in the event of a crash relative to non-intersection location. At the same time, the latent propensity of MGOL and the possible/non-incapacitating injury coefficient of MMNL for intersection related collision indicate the presence of significant unobserved heterogeneity in those estimates. The driveway access related variable also results in a random parameter for incapacitating/fatal injury category in only MGOL model. Further, the MGOL estimates show that collision on driveway access or entrance/exit ramp has a reduced likelihood of severe injury, while railway grade crossing has a positive impact on possible injury outcome. In the MMNL model, the variable representing through roadway results in a higher likelihood of possible and non-incapacitating injuries, while the variable representing other location reduces the likelihood of possible and non-incapacitating injuries.

The broad characterization of exogenous variable effects across the MGOL and MMNL model systems is similar with some differences. These differences can be attributed to the different model structures and different outcome mechanism. The reader would note that in both systems, the impact of exogenous variables was moderated by unobserved effects resulting in statistically significant standard deviation parameters.

2.5 Model Comparison

In the preceding section, we have presented a discussion of model results for the MGOL and the MMNL model. To investigate the comparison further, we examine the model performance under two contexts: (1) presence of underreporting and (2) validation on a hold-out sample.

2.5.1 Underreporting

In police reported crash database, many property damage and minor injury crashes might go underreported since lower crash severity levels make reporting to authorities less likely

(Savolainen and Mannering, 2007). Researchers have argued that underreporting of data will have minimal impact on the model estimation result of standard MNL model (Kim et al., 2007; Shankar and Mannering, 1996; Savolainen and Mannering, 2007; Islam and Mannering, 2006). On the other hand, ordered response models are particularly susceptible to underreporting issue (Savolainen and Mannering, 2007; Ye and Lord, 2011) and can result in biased or inconsistent parameter estimates. However, recent evidence on examining underreporting suggests that none of the models (including unordered response systems) are immune to the underreporting issue (Ye and Lord, 2011). This is expected because the presence of underreporting would not affect the unordered systems only when the dataset under consideration satisfies the independence of irrelevant alternatives (IIA) property. Hence, even the MNL model will yield biased estimates if the IIA property does not hold for the dataset. To reinforce this, we undertake a comparison in the context of underreported data. For this purpose, we generate an underreported data set by randomly removing 50% of no injury crash records from the estimation sample. This reduced dataset is used to re-estimate MGOL and MMNL models. To compare the differences between the estimates from "true" and "underreported" dataset we compute elasticity effects for a selected set of independent variables - Male, Age less than 25, Passenger car, High speed limit, Snowy road surface and Head-on collision (*see* Eluru and Bhat, 2007 for a discussion on computing elasticities). The elasticity estimates are presented in Table 2.2. For the ease of presentation, we focus on the elasticity effects for the two severe injury categories. The results from the "true" sample and underreported sample indicate that the underreported sample consistently obtains the wrong elasticities, as expected. The percentage error in computing elasticity for the selected variables for the two injury severity categories has an average of (33.69, 19.11) and (31.81, 25.96) while the range of the errors is (2.97, 75.99) and (5.85, 57.83) for MGOL and MMNL models, respectively. From the estimated measures we can argue that neither of the models results in unbiased estimates in the underreporting context.

In addition to direct comparison in the context of underreporting, we also undertake a comparison of the elasticity effects with corrections to the MMNL and MGOL models. The correction exercise for altering constants estimated from an underreported sample is relatively straight forward. Specifically, all parameter estimates are kept the same and the constants are altered to match the population shares in the "true" sample. A trial and error approach to alter the constants is employed to generate "corrected" constants for the MMNL model. Further, we employ a similar approach to correct the threshold parameters for the MGOL model. In the MGOL model the population share can be influenced by altering the threshold constants thus achieving the same correction process as the MMNL model. In both correction exercises, adequate care is taken to ensure that the population shares match with the "true" shares after the parameters are corrected. Subsequent to the constant and threshold corrections, the elasticity values are recomputed for the updated estimates. The results are presented in the last block of rows in Table 2.2.

The elasticity errors reduce substantially for both MGOL and MMNL models as a result of the parameter corrections. The average percentage errors in computing elasticity for the selected variables ranges are (15.73, 12.12) and (18.80, 11.27) for MGOL and MMNL models with a range of (0.74, 38.41) and (1.2, 35.89), respectively. We can argue that both the unordered and ordered frameworks perform almost equivalently with underreported dataset and the performance for both of these structures can be improved with the correction measure if the true population share is available to the analyst.

2.5.2 Validation Analysis

A validation experiment is also carried out in order to ensure that the statistical results obtained above are not a manifestation of over fitting to data. For testing the predictive performance of the models, 100 data samples, of about 4000 records each, are randomly generated from the hold out validation sample consisting of 18,201 records. We evaluate both the aggregate and disaggregate measure of predicted fit by using these 100 different validation samples. For these samples, we present the average measure from the comparison, and also the confidence interval (C.I.), of the fit measures at 95% level.

At the disaggregate level we computed predictive log-likelihood (computed by calculating the log-likelihood for the predicted probabilities of the sample), AICc, BIC, predictive adjusted likelihood ratio index, probability of correct prediction, and probability of correct prediction >0.7. The results are presented in Table 2.3. In terms of disaggregate validation measures, the MMNL model consistently outperforms the MGOL model (except for probability of correct prediction >0.7). At the aggregate level, root mean square error (RMSE) and mean absolute percentage error (MAPE) are computed by comparing the predicted and actual (observed) shares of injuries in each injury severity level. We compute these measures for each set of full validation sample and specific sub-samples within that validation population - Driver age less than 25, Air bag deployed, Off-peak hour crash, Snowy surface and Passenger car. The results for aggregate measure computation are presented in Table 2.4.

The comparison of MGOL and MMNL model at the aggregate level is far from conclusive. However, it is clear that MGOL and MMNL models perform very well at the aggregate level. For the full sample, both the MAPE and RMSE values are very close for both models. The RMSE and MAPE values show that the predicted performance for the MGOL model is superior to that of the MMNL model for sub-samples air bag deployed and off-peak hour crash while the MMNL model is superior to that of the MGOL model for driver age less than 25, snowy surface and for passenger car. Thus, we can argue that the differences in the validation measures at aggregate level are not as conclusive as the measures at disaggregate level. Further, the differences in the aggregate level characteristics between the models are very small.

We extend the validation exercise to examine the performance of underreported sample estimates (uncorrected and corrected) as well on the 100 randomly selected validation samples. We compute these measures only for each of the full validation samples (results are presented in Table 2.5). Clearly, based on the underreported sample estimates, the overall errors at disaggregate and aggregate levels are much larger than previously for both systems. In the uncorrected system, MGOL has lower AICc and BIC values, but MMNL has lower RMSE and MAPE values. But in the corrected system, MGOL consistently outperforms the MMNL model (except for RMSE) and the aggregate predicted shares from MGOL model is closer to the actual shares for three out of four injury categories compared to those from MMNL model.

In summary, from the host of validation statistics we can argue that neither the ordered nor the unordered frameworks exclusively outperforms each other both at the aggregate and the disaggregate levels. The relatively close performance of the two model systems is further illustrated through the computation of the validation measures for various sub-samples of the population. Overall, the results indicate that MGOL and MMNL offer very similar prediction for the various sub-samples at the aggregate and disaggregate level. The results reinforce that MGOL model performs very close to the MMNL model in examining driver injury severity.

2.6 Summary

This chapter focuses on the relevance of alternate discrete outcome frameworks for modeling driver injury severity. The most prevalent framework employed to model injury severity is the ordered response mechanism. However, unordered response models were also employed in the past to model crash injury severity. The applicability of the two frameworks for analyzing ordinal discrete variables has evoked considerable debate on using the appropriate framework for analysis. An empirical examination of alternative approaches to modeling injury severity would enable us to determine the preferred model framework between the two modeling approaches.

Further, the two frameworks are also influenced by the underreporting issue associated with crash data sample. Most of the crash data are sampled from police reported crash database, where many property damage and minor injury crashes might go underreported. In the case of an underreported decision variable, the application of traditional econometric frameworks may result in biased estimates. Unfortunately, the unknown population shares of such outcome-based crash severity data make the estimation of parameters even more challenging. In this context, it is essential to examine how alternative modeling frameworks are impacted by underreporting; thus allowing us to adopt frameworks that are least affected by underreporting.

The current chapter addressed the aforementioned issues of identifying the more relevant framework to model crash injury severity by empirically comparing the ordered and unordered outcome models. The performances of these models were also tested in the presence of underreported crash data by creating an artificial reduced dataset. Elasticity measures were generated for the "true" dataset and the artificial underreported dataset to compare the predicted elasticities for the different models. Thus, the current research contributes to the safety analysis literature from both the methodological and empirical standpoint.

The alternative modeling approaches considered for the exercise include: for the ordered outcome framework - ordered logit, generalized ordered logit, mixed generalized ordered logit and for the unordered outcome framework - multinomial logit, nested logit, ordered generalized extreme value logit and mixed multinomial logit model. The empirical analysis was based on the 2010 General Estimates System (GES) database. The focus in the analysis was exclusively on non-commercial passenger vehicle driver crash-related injury severity. Several types of variables were considered in the empirical analysis, including driver characteristics, vehicle characteristics, roadway design and operational attributes, environmental factors and crash characteristics. The empirical results indicated the important effects of all of the above types of variables on injury severity. The model comparison for the estimation sample clearly indicated that the MGOL model outperforms the MMNL model.

To investigate the comparison further, we studied the model performance under two contexts: (1) presence of underreporting and (2) validation on a hold-out sample. We generated a series of measures to evaluate model performance in estimation and prediction thus allowing us to draw conclusions on preferred model frameworks for injury severity analysis. In the context of underreporting, the comparison between the elasticity estimates from "true" and "underreported" sample indicated that the underreported sample consistently obtained the wrong elasticities for both MGOL and MMNL models. The most striking finding was the fact that the MMNL model did not perform any better in the underreporting context than MGOL. Moreover, the correction measures for the thresholds/constants based on the true aggregate shares reduced the elasticity errors substantially for both MGOL and MMNL models. In the context of validation analysis at the aggregate and disaggregate level, we argued that neither the ordered nor the unordered frameworks exclusively outperforms each other. The relatively close performance of the two model systems was further illustrated through the computation of the validation measures for various sub-samples of the population and in the presence of underreporting. Overall, the results of the empirical comparison provided credence to the belief that an ordered system that allows for exogenous variable effects to vary across alternatives and accommodates unobserved heterogeneity offer almost equivalent results to that of the corresponding unordered systems in the context of driver injury severity.

Table 2.1: MGOL and MMNL Estimates

Variables	MGOL			MMNL			
	Latent Propensity	Threshold between Possible and Non-incapacitating Injury	Threshold between Non-incapacitating and Incapacitating/Fatal Injury	No Injury	Possible Injury	Non-incapacitating Injury	Incapacitating/ Fatal Injury
Constant	-1.819	0.208	0.624	–	-2.239	-2.989	-5.215
Driver Characteristics							
Driver gender (Base: Female)							
Male	-0.565 (0.046)	-0.258 (0.046)	–	–	-0.656 (0.057)	-0.540 (0.064)	-0.500 (0.090)
Driver age (Base: Age 25 to 64)							
Age less than 25	-0.441 (0.050)	–	–	0.411 (0.051)	–	–	–
Age above 65+	–	–	–	–	–	–	0.403 (0.137)
Restraint system use (Base: Restrained)							
Unrestrained	1.528 (0.142)	–	–	–	1.303 (0.065)	1.695 (0.073)	2.127 (0.101)
SD Unrestrained	0.844 (0.223)	–	–	–	–	–	–
Under the influence of alcohol	0.489 (0.130)	–	-0.353 (0.122)	–	–	–	0.887 (0.166)
Under the influence of drug	–	–	–	–	–	–	0.776 (0.293)
Vehicle Characteristics							
Vehicle Type (Base: SUV, pickup and vans)							
Passenger car	0.269 (0.046)	–	–	-0.262 (0.047)	–	–	–
Vehicle age (Base: Vehicle age less than 6)							

Vehicle Age 6 to 10	0.144 (0.052)	–	–	–	0.122 (0.057)	0.122 (0.057)	0.308 (0.111)
Vehicle age above 10	0.405 (0.055)	–	–	–	0.312 (0.067)	0.444 (0.073)	0.684 (0.111)
Roadway Design and Operational Attributes							
Interstate Highways	0.303 (0.088)	–	-0.246 (0.090)	–	0.224 (0.092)	0.224 (0.092)	0.672 (0.163)
Speed limit (Base: Speed limit less than 26 mph)							
Speed limit 26 to 50 mph	0.462 (0.072)	-0.127 (0.046)	–	–	0.268 (0.088)	0.541 (0.105)	0.985 (0.172)
Speed limit above 50mph	0.715 (0.089)	–	–	–	0.616 (0.107)	0.767 (0.123)	1.122 (0.196)
Types of Intersection							
Four way intersection	0.177 (0.062)	–	–	-0.172 (0.060)	–	–	–
Traffic Control Device (Base: Non traffic control device)							
Traffic signal/Stop/Yield sign	-0.119 (0.059)	–	–	–	–	-0.252 (0.073)	–
Other traffic control device	0.376 (0.142)	–	–	–	–	–	0.567 (0.239)
Environmental Factor							
Time (Base: 3 pm to 6 pm)							
6 pm to 6 am	–	-0.141 (0.048)	–	–	–	0.032 (0.091)	0.032 (0.091)
SD 6 pm to 6 am	–	–	–	–	–	0.772 (0.211)	0.772 (0.211)
6 am to 9 am	0.173 (0.069)	–	–	-0.214 (0.073)	–	–	–
9 am to 3 pm	0.195 (0.048)	–	–	-0.244 (0.052)	–	–	–
Surface condition (Base: Dry)							
Wet	–	–	–	–	–	-0.179 (0.087)	–
Snowy	-0.648 (0.120)	–	–	–	-0.592 (0.123)	-0.592 (0.123)	-1.041 (0.263)
Crash Characteristics							

Driver ejected out of the vehicle	6.040 (2.655)	1.583 (0.751)	–	–	–	–	–
Vehicle rolled over	2.111 (0.209)	0.177 (0.220)	–	–	1.923 (0.224)	1.923 (0.224)	2.877 (0.286)
SD Vehicle rolled over	–	0.989 (0.343)	–	–	–	–	–
Air bag deployment	1.595 (0.066)	0.270 (0.073)	–	–	1.303 (0.065)	1.695 (0.073)	2.127 (0.101)
SD Air bag deployment	0.844 (0.223)	–	–	–	–	–	–
Collision object (Base: Another moving vehicle)							
Collision with stationary object	0.774 (0.081)	-0.283 (0.074)	-0.226 (0.087)	–	0.416 (0.097)	0.936 (0.098)	1.203 (0.257)
SD Collision with stationary object	–	–	0.847 (0.233)	–	–	–	1.310 (0.379)
Collision with other object	-1.174 (0.189)	-1.162 (0.313)	–	–	-1.774 (0.329)	-0.647 (0.233)	–
Manner of collision							
Head on	0.966 (0.100)	–	-0.393 (0.100)	–	0.805 (0.109)	0.805 (0.109)	1.974 (0.175)
Angular	0.382 (0.063)	-0.150 (0.061)	-0.244 (0.067)	–	0.317 (0.068)	0.317 (0.068)	1.153 (0.155)
Side swipe-same direction	-0.534 (0.097)	–	0.316 (0.151)	–	-0.334 (0.122)	-0.512 (0.150)	-1.206 (0.330)
Rear to side collision	–	-3.683 (0.717)	2.309 (0.182)	–	–	–	–
Other manners of collision	-1.258 (0.627)	–	1.651 (0.178)	–	–	–	–
Collision location (Base: Non-intersection)							
Intersection	–	0.227 (0.061)	–	–	–	–	-0.369 (0.141)
Intersection related	-0.255 (0.071)	–	–	–	-0.430 (0.155)	-0.430 (0.155)	-0.530 (0.170)
SD Intersection related	0.007 (0.002)	–	–	–	0.915 (0.323)	0.915 (0.323)	–
Driveway access	-0.477 (0.243)	–	–	–	–	–	–

Entrance and exit ramp	-0.323 (0.150)	–	–	–	–	–	–
Railway grade crossing	–	1.181 (0.421)	-3.981 (0.987)	–	–	–	–
Driveway access related	-0.427 (0.087)	–	–	–	-0.335 (0.090)	-0.335 (0.090)	-2.649 (1.210)
SD Driveway access related	–	–	–	–	–	–	2.332 (0.896)
Through roadway	–	–	–	–	0.913 (0.428)	0.913 (0.428)	–
Other location	–	–	–	–	-0.768 (0.375)	-0.768 (0.375)	–

Table 2.2: Elasticity Effects

Variables	MGOL				MMNL			
	Non-incapacitating injury	Incapacitating /Fatal injury	% of error in Non-incapacitating injury	% of error in incapacitating /Fatal injury	Non-incapacitating injury	Incapacitating /Fatal injury	% of error in Non-incapacitating injury	% of error in Incapacitating/ Fatal injury
Estimation sample								
Male	-17.28	-20.35	-	-	-25.26	-14.51	-	-
Age less than 25	-24.07	-29.69	-	-	-19.97	-14.72	-	-
Passenger car	15.23	18.76	-	-	13.02	9.50	-	-
High speed limit	43.77	57.44	-	-	38.41	63.82	-	-
Snowy surface	-32.69	-38.40	-	-	-24.20	-44.32	-	-
Head-on collision	27.54	153.04	-	-	20.27	173.52	-	-
Underreported sample without corrections								
Male	-11.33	-12.06	34.44	40.74	-16.42	-7.07	35.00	51.25
Age less than 25	-18.47	-25.14	23.26	15.31	-18.80	-13.62	5.85	7.49
Passenger car	12.25	16.76	19.60	10.65	11.08	6.03	14.89	36.47
High speed limit	36.23	55.73	17.23	2.97	28.37	47.52	26.15	25.54
Snowy surface	-22.36	-28.92	31.61	24.70	-11.83	-34.33	51.13	22.53

Head-on collision	6.61	121.98	75.99	20.30	8.55	151.89	57.83	12.46
Average Error	-	-	33.69	19.11	-	-	31.81	25.96
Underreported sample with corrections								
Male	-15.57	-17.32	9.88	14.87	-23.19	-12.78	8.17	11.90
Age less than 25	-20.96	-26.24	12.93	11.62	-23.14	-17.43	15.88	18.39
Passenger car	13.95	17.49	8.43	6.78	17.69	10.42	35.89	9.70
High speed limit	43.44	58.88	0.74	2.51	38.87	57.10	1.20	10.53
Snowy surface	-24.85	-29.97	23.99	21.96	-16.83	-37.70	30.46	14.94
Head-on collision	16.96	130.13	38.41	14.96	24.56	177.21	21.19	2.13
Average Error	-	-	15.73	12.12	-	-	18.80	11.27

Table 2.3: Disaggregate Measures of Fit in Validation Sample

DISAGGREGATE MEASURE OF FIT IN VALIDATION SAMPLE		
Summary statistic	**MGOL predictions**	**MMNL predictions**
Number of observations	3993.9900	3993.9900
Number of parameters	55	61
Log-likelihood at zero	-5536.8458	-5536.8458
Log-likelihood at sample shares	-3962.5600	-3962.5600
Predictive Log-likelihood	-3671.0702	-3643.0636
C.I.	-3685.6638/-3656.4766	-3657.3289/-3628.7984
AICc	7453.7050	7410.0514
C.I.	7424.5207/7482.8892	7381.5246/7438.5782
BIC	7798.2252	7791.9668
C.I.	7768.9357/7827.5147	7763.3179/7820.6156
Predictive adjusted likelihood ratio index	0.0597	0.0652
C.I.	0.0578/0.0615	0.0638/0.0667
Average probability of correct prediction	0.6649	0.6663
C.I.	0.6636/0.6662	0.6650/0.6677
Average probability for chosen probability>0.70	0.4787	0.4620
C.I.	0.4774/0.4799	0.4609/0.4632

Table 2.4: Aggregate Measures of Fit in Validation Sample

AGGREGATE MEASURE OF FIT IN VALIDATION SAMPLE				
	Injury categories/Measures of fit	Actual shares	MGOL predictions	MMNL predictions
	No injury	66.4311	65.8805	65.9509
	C.I. *	-	65.8118/65.9492	65.8842/66.0174
	Possible injury	15.0667	15.1281	15.0362
	C.I.	-	15.1034/15.1528	15.0139/15.0583
	Non-incapacitating injury	11.3647	12.0757	12.0754
	C.I.	-	12.0449/12.1064	12.0476/12.1032
	Incapacitating/Fatal injury	7.1375	6.9157	6.9376
	C.I.	-	6.8823/6.9492	6.9029/6.9722
	RMSE	-	0.6319	0.6105
	C.I.	-	0.5883/0.6756	0.5667/0.6544
	MAPE	-	3.7679	3.6586
	C.I.	-	3.7651/3.7706	3.6558/3.6613
Driver age less than 25	No injury	69.1630	67.9434	67.8363
	C.I.	-	67.8059/68.0809	67.71094/67.9617
	Possible injury	12.8669	14.1267	13.3549
	C.I.	-	14.0783/14.1751	13.3131/13.3967
	Non-incapacitating injury	11.2528	11.3173	11.7434
	C.I.	-	11.2599/11.3747	11.6869/11.7999
	Incapacitating/Fatal injury	6.7173	6.6126	7.0653
	C.I.	-	6.5453/6.6799	6.9988/7.1319
	RMSE	-	1.1199	1.0354
	C.I.	-	1.0377/1.2023	0.9641/1.1067
	MAPE	-	6.6456	6.1554
	C.I.	-	6.6408/6.6505	6.1509/6.1600
	Predictive Log-likelihood	-	-1028.3794	-1015.5878
	C.I.	-	-1036.0795/-1020.6794	-1023.1219/-1008.0537
Air bag deployed	No injury	34.6793	34.8052	34.4638
	C.I.	-	34.6797/34.9307	34.3658/34.5619
	Possible injury	23.7389	23.7988	23.4669
	C.I.	-	23.7434/23.8541	23.4176/23.5162

	Non-incapacitating injury	23.1525	23.0632	24.1901
	C.I.	-	22.9902/23.1361	24.1354/24.2449
	Incapacitating/Fatal injury	18.4293	18.3329	17.8792
	C.I.	-	18.2296/18.4362	17.7821/17.9762
	RMSE	-	1.2129	1.2902
	C.I.	-	1.1276/1.2984	1.1869/1.3934
	MAPE	-	4.2884	4.6403
	C.I.	-	4.2852/4.2915	4.6364/4.6441
	Predictive Log-likelihood	-	-1385.1886	-1318.6118
	C.I.	-	-1394.7695/-1375.6077	-1327.2064/-1310.0172
Off-peak period	No injury	66.9671	65.8187	65.6960
	C.I.	-	65.7138/65.9236	65.5950/65.7969
	Possible injury	15.8240	16.1584	16.4015
	C.I.	-	16.1176/16.1993	16.3655/16.4375
	Non-incapacitating injury	10.9846	11.9150	11.8761
	C.I.	-	11.8676/11.9624	11.8398/11.9123
	Incapacitating/Fatal injury	6.2242	6.1078	6.0265
	C.I.	-	6.0606/6.1550	5.9774/6.0755
	RMSE	-	0.9911	1.0427
	C.I.	-	0.9119/1.0703	0.9637/1.1218
	MAPE	-	5.7662	6.0102
	C.I.	-	5.7612/5.7711	6.0054/6.0150
	Predictive Log-likelihood	-	-1226.6454	-1207.2053
	C.I.	-	-1234.7771/-1218.5138	-1215.5970/-1198.8135
Snowy surface	No injury	73.0563	71.9579	71.7287
	C.I.	-	71.6597/72.2560	71.4244/72.0330
	Possible injury	10.7654	12.2862	11.4324
	C.I.	-	12.1692/12.4032	11.3389/11.5259
	Non-incapacitating injury	11.6573	9.9632	11.6253
	C.I.	-	9.8255/10.1009	11.4894/11.7612
	Incapacitating/Fatal injury	4.5210	5.7927	5.2135
	C.I.	-	5.6498/5.9356	5.0748/5.3523
	RMSE	-	2.1626	1.8423
	C.I.	-	1.9874/2.3379	1.6628/2.0217
	MAPE	-	20.8766	16.7887

	C.I.	-	20.8500/20.9033	16.7651/16.8122
	Predictive Log-likelihood	-	-150.5851	-149.2434
	C.I.	-	-153.9116/-147.2586	-152.4695/-146.0173
Passenger car	No injury	63.3983	62.5658	62.6231
	C.I.	-	62.4731/62.6584	62.5320/62.7141
	Possible injury	16.4833	16.3340	16.5008
	C.I.	-	16.3018/16.3661	16.4707/16.5309
	Non-incapacitating injury	12.3735	13.2977	13.3121
	C.I.	-	13.2583/13.3371	13.2753/13.3489
	Incapacitating/Fatal injury	7.7449	7.8026	7.5640
	C.I.	-	7.7552/7.8499	7.5178/7.6102
	RMSE	-	0.8573	0.8286
	C.I.	-	0.7917/0.9229	0.7598/0.8974
	MAPE	-	4.6446	4.5066
	C.I.	-	4.6412/4.6479	4.5029/4.5102
	Predictive Log-likelihood	-	-2311.8055	-2301.2185
	C.I.	-	-2322.8512/-2300.7599	-2313.0902/-2289.3468

**C.I.* =*Confidence Interval*

Table 2.5: Measures of Fit in Validation for Underreported sample

MEASURE OF FIT IN UDERREPORTED SAMPLE			
Injury categories/Measures of fit	**Actual shares**	**MGOL predictions**	**MMNL predictions**
No injury	66.4311	52.4731	52.6582
C.I.[*]	-	52.4051/52.5411	52.5779/52.7386
Possible injury	15.0667	21.6642	21.5562
C.I.	-	21.6359/21.6925	21.5045/21.6079
Non-incapacitating injury	11.3647	17.0554	16.9202
C.I.	-	17.0207/17.0901	16.8876/16.9528
Incapacitating/Fatal injury	7.1375	8.8073	8.8653
C.I.	-	8.7683/8.8463	8.8277/8.9029
RMSE	-	8.2760	8.1565
C.I.	-	8.2049/8.3470	8.0806/8.2324
MAPE	-	34.7376	34.3961
C.I.	-	34.7334/34.7418	34.3918/34.4005
Predictive Log-likelihood	-	-4080.7320	-4089.1194
C.I.	-	-4096.0726/-4065.3915	-4104.0381/-4074.2008
AICc	-	8264.8098	8293.9191
C.I.	-	8234.1313/8295.4884	8264.0853/8323.7529
BIC	-	8584.3790	8650.9086
C.I.	-	8553.6005/8615.1576	8620.9523/8680.8649
MEASURE OF FIT IN UDERREPORTED SAMPLE WITH CORRECTION			
Injury categories/Measures of fit	**Actual shares**	**MIXGOL predictions**	**MIXMNL predictions**
No injury	66.4311	69.4232	69.4094
C.I.	-	69.3574/69.4889	69.3349/69.4839
Possible injury	15.0667	13.7549	13.8957
C.I.	-	13.7262/13.7835	13.8526/13.9389
Non-incapacitating injury	11.3647	10.9293	10.8844
C.I.	-	10.8999/10.9586	10.8553/10.9135
Incapacitating/Fatal injury	7.1375	5.8926	5.8105
C.I.	-	5.8599/5.9253	5.7786/5.8423
RMSE	-	1.7944	1.7827
C.I.	-	1.7256/1.8633	1.7119/1.8536
MAPE	-	8.6295	8.7599

C.I.	-	8.6266/8.6325	8.7569/8.7629
Predictive Log-likelihood	-	-3853.4807	-3881.9877
C.I.	-	-3869.9209/-3837.0405	-3898.5934/-3865.3820
AICc	-	7810.3072	7879.6556
C.I.	-	7777.4290/7843.1853	7846.4471/7912.8641
BIC	-	8129.8764	8236.6451
C.I.	-	8096.9087/8162.8441	8203.3327/8269.9575

**C.I. =Confidence Interval*

CHAPTER 3 A Latent Segmentation based Generalized Ordered Logit Model – Heterogeneity in Driver Injury Severity Modeling

3.1 Introduction

Road safety researchers have employed several statistical formulations for analyzing the relationship between injury severity and crash related factors. But, as indicated earlier, the most prevalent formulation to study injury severity is the ordered outcome formulation. The traditional ordered outcome formulation imposes a restrictive monotonic assumption regarding the impact of exogenous variables on the injury severity levels (Eluru et al., 2008). To address this limitation, researchers have employed the unordered outcome formulation that allows the impact of exogenous variables to vary across injury severity levels. The most common model used under the unordered outcome formulation is the multinomial logit model (Khorashadi et al., 2005; Islam and Mannering, 2006; Awadzi et al., 2008; Schneider et al., 2009; Ulfarsson and Mannering, 2004). However, the unordered model does not recognize the inherent ordering of the crash severity outcome and, therefore, neglects vital information present in the data. To recognize the ordinality of the injury severity levels, as well as provide as much flexibility as the unordered model formulation, Eluru et al., (2008) proposed the GOL formulation that bridges the divide between the traditional ordered-response and the traditional unordered-response formulations (Eluru, 2013).

The current chapter contributes to the safety literature methodologically and empirically by building on the GOL formulation. In terms of methodology, we formulate and estimate a latent segmentation based generalized ordered logit (LSGOL) model. The LSGOL model relaxes the traditional GOL formulation assumption that the effects of exogenous variables on the injury risk propensity, and on the thresholds that map the risk propensity to injury severity outcomes, are fixed across all drivers involved in collisions. Empirically, the LSGOL model is estimated using driver injury severity data from the state of Victoria, Australia, employing a comprehensive set of exogenous variables.

The rest of the chapter is organized as follows. Section 3.2 provides details of the econometric model framework used in the analysis. In Section 3.3, the data source and sample formation procedures are described. The model comparison results, elasticity effects and validation measures are presented in Section 3.4, 3.5 and 3.6, respectively. Section 3.7 concludes the chapter by summarizing the major findings of the study.

3.2 Model Framework

The analysis in this chapter is undertaken at the level of drivers involved in a crash. That is, we focus on driver-level injury severity in a crash. Thus, in the case of a crash involving a single vehicle with an object, there is one driver record with the corresponding injury severity level sustained by the driver. In the case of a crash involving multiple drivers, each driver contributes a record, along with the injury severity level sustained by the driver.

The framework used for modeling driver-level injury severity assumes that drivers can be implicitly sorted into S relatively homogenous (but latent to the analyst) segments based on characteristics of the crash. Within each segment, the effects of exogenous variables are fixed across drivers in the segment. Let s be the index for segments ($s = 1,2, \ldots S)$), i be the index for drivers ($i = 1,2, \ldots, N$), and j be the index for driver injury severity levels ($j = 1,2, \ldots \ldots \ldots, J$). The crash outcomes are analyzed using a GOL model within each segment. Across segments, the parameters of the GOL model vary. In the GOL model, conditional on driver i belonging to segment s, the discrete injury severity levels (y_i) are assumed to be a mapping (or partitioning) of an underlying continuous latent variable (y_i^*) as follows:

$$y_i^* | (i \in s) = \boldsymbol{X}_i \boldsymbol{\beta}_s + \varepsilon_{is}\,, \;\; y_{is} = j, if\ \tau_{i,j-1,s} < \; y_i^* < \tau_{i,j,s} \tag{3.1}$$

where,

$\boldsymbol{X}_i$ is a row vector of exogenous variables

$\boldsymbol{\beta}_s$ is a corresponding column vector of unknown parameters specific to segment s

ε_{is} is a segment-specific idiosyncratic random disturbance term assumed to be identically and independently standard logistic

$\tau_{i,j,s}$ ($\tau_{i,0,s} = -\infty\,, \tau_{i,J,s} = +\infty$) represents the segment-specific upper threshold associated with driver i and severity level j, with the following ordering conditions: $\left(-\infty < \tau_{i1,s} < \tau_{i2,s} < \; \ldots \ldots \ldots < \tau_{iJ-1,s} < +\infty\right) \forall \;\; s = 1,2, \ldots S$.

To maintain the ordering conditions and allow the thresholds to vary across drivers within each segment, Eluru et al. (2008) propose the following non-linear parameterization of the thresholds as a function of exogenous variables:

$$\tau_{ij,s} = \tau_{ij-1,s} + exp(\boldsymbol{\delta_{js}} \boldsymbol{Z_{is}}) \tag{3.2}$$

where $\boldsymbol{\delta}_{js}$ is a segment-specific and injury level-specific row vector of parameters to be estimated and $\boldsymbol{Z_{is}}$ is a corresponding column vector of segment-specific exogenous variables ($\boldsymbol{Z_{is}}$ includes a constant as its first element, with the corresponding coefficient being $\boldsymbol{\delta}_{js,1}$; for identification, we need $\boldsymbol{\delta}_{1s,-1}$ to be a row vector of zero values, where $\boldsymbol{\delta}_{1s,-1}$ is a sub-vector of the vector $\boldsymbol{\delta}_{1s}$ minus the first element). The traditional ordered logit (OL) model assumes that the thresholds $\tau_{i,j,s}$ remain fixed across drivers ($\tau_{ij,s} = \tau_{j,s} \;\; \forall \;\; i$) for each segment; that is, it assumes that $\boldsymbol{\delta}_{js,-1}$ has all zero elements for all j values and all s values.

Given the above set-up, the probability that driver i suffers an injury severity outcome j, conditional on driver i belonging to segment s, may be written as:

$$P_i(j)|s = \Lambda\big(\tau_{ij-1,s} + exp(\boldsymbol{\delta_{js}} \boldsymbol{Z_{is}}) - \boldsymbol{X}_i \boldsymbol{\beta}_s\big) - \Lambda\big(\tau_{ij-2,s} + exp(\boldsymbol{\delta_{j-1,s}} \boldsymbol{Z_{is}}) - \boldsymbol{X}_i \boldsymbol{\beta}_s\big) \tag{3.3}$$

where $\Lambda(.)$ represents the standard logistic cumulative distribution function.

Of course, the analyst does not observe the segment to which driver i belongs. So, the analyst specifies this segment assignment to be a function of a column vector of observed crash factors $\boldsymbol{\eta_i}$. To also acknowledge the presence of unobserved factors that may influence this assignment, the analyst develops an expression for the probability of driver i belonging to segment s. While many parametric expressions may be used for this probability expression (the only requirement is that the probabilities sum to one across the segments for each driver i), the most commonly used form corresponds to the multinomial logit structure (see Bhat, 1997; Greene and Hensher, 2003; Eluru et al., 2012):

$$P_{is} = \frac{\exp[\boldsymbol{\alpha_s} \, \boldsymbol{\eta_i}]}{\sum_s \exp[\boldsymbol{\alpha_s} \, \boldsymbol{\eta_i}]} \tag{3.4}$$

where $\boldsymbol{\alpha_s}$ is a row vector of parameters to be estimated. Then, the unconditional probability of driver i leading up to injury severity level j can be written as:

$$P_i(j) = \sum_{s=1}^{S} (P_i(j)|s) \times (P_{is}) \tag{3.5}$$

The log-likelihood function for the entire dataset can be written as:

$$L = \sum_{i=1}^{N} log \left[\sum_{s=1}^{S} (P_i(j)|s) \times (P_{is}) \right] \tag{3.6}$$

The parameters to be estimated in the LSGOL model are the segment parameters ($\boldsymbol{\beta}_s$ & $\boldsymbol{\delta}_{js}$), the class probability parameters ($\boldsymbol{\eta_i}$) for each s, and the appropriate number of segments S. For identification reasons, we need to restrict one of the $\boldsymbol{\delta_{js}}$ vectors to zero. It is worthwhile to mention here that the estimation of latent segmentation based models using Quasi-Newton routines can be computationally unstable (see Bhat, 1997 for a discussion). The estimation of such models requires employing good starting values for the estimation procedure. Hence, for our analysis, the log-likelihood function and its corresponding gradient function were coded in the Gauss Matrix programming language. The coding of the gradient function ensures the reduction in instability associated with such an estimation process.

3.3 Data

3.3.1 Data Source

Data for our empirical analysis of this chapter is sourced from the Victoria crash database of Australia for the years 2006 through 2010. The data includes information reported by Victorian police officers for crashes involving at least one motor vehicle travelling on a roadway and resulting in property damage, injury or death, which are then compiled by VicRoads (a statutory body responsible for road transport in the state of Victoria). For the five years, the crash database has a record of 67,809 crashes involving 118,842 motor vehicles and 166,040 individuals, resulting in 1,550 fatalities and 87,855 injuries to the crash victims. A four point ordinal scale is used in the database to represent the injury severity of individuals involved in these crashes: 1) No injury; 2) Minor injury; 3) Serious injury and 4) Fatal injury.

3.3.2 Sample Formation and Description

This study is focused on the injury severity outcome of drivers, who are involved in either a single or a two passenger vehicle collisions. The crashes that involve more than two passenger vehicles are excluded from the analysis (about 9.9% of the sample). The crashes that involve commercial vehicles are also excluded to avoid the potential systematic differences between the crashes involving commercial and non-commercial driver groups. The final dataset, after removing records with missing information for essential attributes, consisted of 42,812 driver records. The final sample had a very small percentage of records that involved a fatally injured driver (about 1% of total crashes). Therefore, both the fatal and serious injury category levels are merged together in the current analysis.

From the dataset of 42,812 driver records, a sample of 5,132 records is randomly drawn for the purpose of estimating models and 37,680 records are set aside for the purpose of validation. In the final estimation sample, the distribution of driver injury severity levels is as follows: no injury 41.8%, minor injury 36.9% and serious/fatal injury 21.3%. Table 3.1 offers a summary of the sample characteristics of the exogenous factors in the estimation sample. From the descriptive analysis, we observe that a large portion of crashes involve short-side angular collisions (22.1%), and at locations with no traffic control (60.1%), in a medium speed zone location (66.5%), during the off peak period (33%), in clear weather (84.4%), in daylight (69.3%) and in the presence of at least one passenger in the vehicle (88.4%). The majority of drivers are adult (63.9%), use seat-belts (96.5%) and drive a sedan (71.4%). The drivers are somewhat more likely to be male than female (male 52.8% versus female 47.1%). It is also quite interesting to note that the share of vehicles that are more than 10 years old is quite large (43.4%).

3.4 Empirical Analysis

3.4.1 Variables Considered

The collision attributes considered in this empirical study can be grouped into six broad categories: crash characteristics, driver characteristics, vehicle characteristics, roadway design attributes, environmental factors and situational factors.

The crash characteristics examined were collision object (small object, large object, animal, and moving vehicle), trajectory of vehicle's motion (going straight or other movement), and manner of collision. As one would expect, the manner of collision, whether it is a head-on or a sideswipe, has significant implications for injury severity sustained. For example, the greater dissipation of kinetic energy associated with a head-on collision is likely to result in severe injuries compared to a side-swipe crash. Most of the earlier studies define the manner of collision as a crash level variable (rear-end, sideswipe, angular, and head-on) – by assigning one collision type for all vehicles involved in the same collision. But, depending on the initial point of impact it is possible that different vehicles involved in the same crash might have significantly different crash profiles. For example, in a rear-end collision involving two vehicles, one of the vehicle will be rear-ended and the other one will be the rear-ender. The driver of the rear-ended vehicle is likely to be pushed backward into the seat when struck by the rear-ender vehicle leading to a high probability of whiplash or neck injury due to the continuous movement of the neck at a different speed relative to the head and the rest of the body (Khattak, 2001; Chiou et al., 2013; Nordhoff, 2005). Due to the biomechanics of this type of crash, the driver in the rear-ended vehicle is likely to be more seriously injured in a rear-end crash compared to the driver in the rear-ender vehicle. Hence, it is incorrect to assign the same collision type variable to all vehicles involved in the same crash in analyzing vehicle occupant injury severity[6]. The study of the current chapter addresses this inconsistency and define a vehicle level manner of collision variable using a combination of collision type and the initial point of contact. A schematic diagram of the initial point of impact relative to the driver's seat position is shown in Figure 3.1 (the collision type and the initial point of impact are computed relative to the position of driver in the vehicle). Based on the collision type and the point of impact, we identified seven categories for the "manner of collision": Rear-ender (the rear vehicle that is involved in a rear-end collision), Rear-ended (the front vehicle that is involved in the rear-end collision), Near-sideswipe (sideswipe/near-side), Near-angular (angular/near-side), Short-side angular (angular/front and rear side), Far-side (angular and sideswipe/far-side) and Head-on (head-on/front side).

The driver characteristics included are driver gender, age and seat belt use information. Vehicle characteristics considered are vehicle type (characterized as sedan, station wagon, utility and panel van) and vehicle age. The roadway design attributes considered in the analysis are road surface type, presence of traffic control device, and presence of a speed zone (speed zone is a length or an area of road along which a signposted regulatory speed limit applies). The

[6] To be sure, Abdel-Aty and Abdelwahab (2003) examine the crash occurrence and Khattak (2001) examine the driver injury outcome by considering the manner of collision as a vehicle level variable, but only for rear-end collision.

environmental factors included are season, time of day, weather condition, and lighting condition. Finally, the situational factors included in the model are the number of passengers and whether or not the driver was ejected. The final specification of the model development was based on combining the variables when their effects were not statistically different and by removing the statistically insignificant variables in a systematic process based on statistical significance (90% significance level). For continuous variables, linear, polynomial and spline forms were tested.

3.4.2 Variable Considered for Segmentation of Crashes

The proposed modeling approach in this chapter theoretically can accommodate classification of segments based on the universal set of variables. However, in our analysis, we consider segmentation based only on traffic crash characteristics for two reasons. First, while it is plausible to consider all attribute sets in the latent segmentation consideration, the estimation of latent segmentation models with the entire attribute set is likely to result in convergence challenges as well as difficulty in interpreting the results (see Sobhani et al., 2013 and Eluru et al., 2012 for discussions on challenges associated with latent segmentation models). Second, in the safety literature, there has been substantial interest in exploring the impact of crash characteristics on injury severity. In fact, many previous injury severity studies have focused only on a specific type of crash, which is tantamount to specifying separate injury severity models for each crash type (as discussed in section 1.5.4). While these research attempts are very useful, the approach results in models where injury severity records are exclusively allocated to about various segments (defined by crash type) and analysed through separate severity models for each segment. However, doing so implies that the model estimation is undertaken on a relatively small sample of the accident records for at least some crash types. In this chapter, we offer an alternate approach by examining segmentation on the basis of crash characteristics (collision object, the trajectory of vehicle's motion, and manner of collision), and analyze driver-level injury severity within each segment using other crash attributes. The approach allows us to retain a smaller number of segments while assigning individuals probabilistically. In this manner, we ensure that the entire sample is utilized in model estimation for each segment. Thus, the latent segmentation based model provides an elegant and effective approach to study the influence of crash characteristics through segmentation, while acknowledging the need for separate injury severity models for each segment.

3.4.3 <u>Model Specification and Overall Measures of Fit</u>

The empirical analysis involves the estimation of four models: (1) the ordered logit (OL) model, (2) the generalized ordered logit (GOL) model, (3) the latent segmentation based ordered logit (LSOL) model, and (4) the latent segmentation based generalized ordered logit (LSGOL) model. Prior to discussing the estimation results, we compare the performance of these models in this section. The model comparisons are undertaken in two stages. First, we determine the appropriate latent segmentation scheme for the OL and GOL models. Second, we compare the traditional (unsegmented) OL and GOL models with the more general latent models (LSOL and LSGOL) obtained from the first step.

3.4.3.1 <u>Determining the Appropriate Latent Segmentation Model</u>

The estimation of the latent segmentation based model involves the probabilistic assignment of the drivers involved in collisions into a given number of segments (S) based on the available exogenous variables. In the application of these models, determining the appropriate number of segments is a critical issue with respect to interpretation and inferences. Therefore, we estimate these models with increasing numbers of segments ($S = 2, 3, 4, \dots$) until an addition of a segment does not add value to the model in terms of data fit. Many of the earlier studies suggest that the Bayesian Information Criterion (BIC) is the most consistent information criterion (IC) among all other traditionally used ICs (AIC, AICc, adjusted BIC) for segment analysis (Nylund et al., 2007; Bhat, 1997; Collins et al., 1993). The advantage of using the BIC is that it imposes substantially higher penalty than other ICs on over-fitting. Thus, in the current study context, the most appropriate number of segments in the LSOL and LSGOL models is determined based on the BIC measure.

We estimated the LSOL and LSGOL models with S = 2 (LSOL II and LSGOL II models) and 3 (LSOL III and LSGOL III models) segments and computed the BIC values for each of these models. The model with the *<u>lower</u>* BIC is the preferred model. For the LSOL model, the computed BIC values with 2 and 3 segments are 10049.72 (37 parameters) and 10257.93 (34 parameters), respectively. The BIC values for the LSGOL model with 2 and 3 segments are 10024.01 (41 parameters) and 10385.26 (31 parameters), respectively. Thus, we selected two segments as the appropriate number of segments for both the LSOL and LSGOL models in current study context.

3.4.3.2 Comparison across All Models - Non-nested Test

To evaluate the performance of the estimated OL, GOL, LSOL and LSGOL models, the BIC values are computed as shown in equation 2.18. Also, the AICc values are computed for each of the four models as shown in equation 2.19. Model with *lower* BIC and AICc values are preferred to models with higher values for these ICs. The BIC (AICc) values for the final specifications of the OL, GOL, LSOL and LSGOL models are 10086.40 (9929.59), 10048.06 (9769.44), 10049.72 (9808.17) and 10024.01 (9756.42), respectively. The comparison exercise clearly highlights the superiority of the LSGOL model in terms of data fit compared to all the other models.

3.4.4 Estimation Results

In presenting the effects of exogenous variables in the model specification, we will restrict ourselves to the discussion of the LSGOL model. Table 3.3 presents the estimation results. Following Bhat (1997), we first present some descriptive characteristics of the two segments in the LSGOL model, before proceeding to a discussion of the variables that impact segmentation and the injury severity levels of drivers within each segment.

3.4.4.1 Descriptive Characteristics of the Segments in the LSGOL Model

To delve into the characteristics that delineate the segments and to understand the characteristics of each segment, the model estimates are used to generate information on: (1) the population share of each of the two segments and (2) the overall injury severity level shares within each segment. These estimates are presented in Table 3.2. The population share or the size of each segment is computed as:

$$G_s = \frac{\sum_i P_{is}}{N} \tag{3.7}$$

where N is the total number of drivers in the estimation sample. From the first row of Table 3.2 labeled "Driver population share", it is evident that a driver is more likely to be assigned to segment 2 than to segment 1. Further, the driver injury severity outcome probabilities, conditional on assignment to a segment, are obtained using equation 3.3. The segment-specific injury outcome shares are then computed by taking the average (across all drivers) of the driver-specific probabilities associated with each injury outcome level. The results are presented in the second row panel of Table 3.2. It is clear that a driver, if allocated to segment 1, is likely to be involved in a more severe crash than if allocated to segment 2. Thus, we may label segment 1 as the "*high risk segment*" and segment 2 as the "*low risk segment*".

3.4.4.2 Latent Segmentation Component

The latent segmentation component determines the relative prevalence of each class, as well as the probability of a driver being assigned to one of the two latent segments based on the crash characteristics. In our empirical analysis, the crash characteristics that affect the allocation of drivers to segments include collision object, trajectory of vehicle's motion, and manner of collision. The results in Table 3.3 provide the effects of these crash characteristics, using the high risk segment (segment one) as the base segment. Thus, a positive (negative) sign for a variable indicates that crashes with the variable characteristic are more (less) likely to be assigned to the low risk segment relative to the high risk segment, compared to crashes that correspond to the characteristic represented by the base category for the variable. The positive sign on the constant term does not have any substantive interpretation, and simply reflects the larger size of the low risk segment compared to the high risk segment.

The results for the "collision object variables" indicate an increased likelihood of drivers being assigned to the high risk segment in case of a collision with stationary objects (small or large object) compared to a collision with another moving vehicle. In terms of the trajectory of the vehicle's motion, the driver of a vehicle traveling straight through just prior to a crash is at a higher risk of severe injury relative to drivers making other turning movements. This result is to be expected because straight-through drivers are likely to be travelling at higher speeds.

Consistent with several previous studies (Chiou et al., 2013; Khattak, 2001), our analysis also shows that being the driver of the rear-ended vehicle in a rear-end collision increases the probability of a high risk crash. The driver of the vehicle is likely to be pushed backward into the seat when struck by the following vehicle, which results in higher probability of whiplash or neck injury due to the continuous movement of the neck at a different speed than the head and the rest of the body (Khattak, 2001; Krafft et al., 2000; Nordhoff, 2005). Thus, the biomechanics of this type of collision explains the increased probability of a high risk crash.

The result associated with a head-on collision also reflects an increased likelihood of assigning the drivers involved in the crash to the high risk segment. Head-on collisions are often caused by drivers violating traffic rules, crossing the centerline by mistake and losing control of their vehicles (Zhang and Ivan, 2005). The pre-impact speed vectors of motor vehicles are directed in opposing directions during a head-on collision, resulting in greater dissipation of kinetic energy and heavier deformation of motor vehicle bodies (Prentkovskis et al., 2010), resulting in higher risk of injury (Tay and Rifaat, 2007; Gårder, 2006). The drivers who are involved in a near-angular collision also are likely to be assigned to the high risk segment. These crashes impose more risk on the driver due to the angle of impact (Jin et al., 2010) and the greater force of impact (Tay and Rifaat, 2007). Moreover, there is less collapsible structure between the striking force and the drivers, which might result in significant passenger compartment intrusion and the direct loading of impact resulting in serious chest and abdominal injury (Mackay et al., 1993; McLellan et al., 1996).

For the far-side manner of collision, the result indicates that this kind of collision reduces the propensity of drivers being in the high risk segment. The significant gap between the collision impact point and driver position might lessen the direct impact of force as a large amount of kinetic energy is absorbed by the vehicle (Sobhani et al., 2011), thereby reducing the risk of high injury severity.

3.4.4.3 <u>Injury Severity Component: High Risk Segment (Segment 1)</u>

The injury severity component within the high risk segment (segment 1) is discussed in this section. The two columns of the corresponding segment in Table 3.3 represent the latent injury risk propensity and the threshold demarcating the minor injury level from the serious/fatal injury level, respectively.

<u>Driver Characteristics:</u> The age of drivers involved in the collision has a significant influence on crash severity. The estimation results indicate a reduction in the risk propensity for young drivers (age less than 25). But the impact of driver age on the threshold demarcating the minor injury and serious/fatal injury levels indicates that the distance between these thresholds get contracted for young drivers relative to other adult drivers (age 25 to 64). The net implication is that young drivers in this first segment have a higher probability of sustaining no injury, and a lower overall probability of some kind of an injury (minor injury or serious/fatal injury). But the contraction of the distance between the thresholds implies that the effect of age on the minor injury and

serious/fatal injury categories is crash and driver-specific; for some contexts, the minor injury probability can increase with a concomitant decrease in the serious/fatal injury probability, while for other contexts the reverse can hold. This highlights the advantage of a GOL framework that allows for flexible exogenous variable impacts. The lower probability of injury among young adults may reflect the higher physiological strength of young drivers in withstanding crash impacts (Xie et al., 2012; O'Donnell and Connor, 1996; Castro et al., 2013), while the higher probability of serious/fatal injuries in some crashes may represent the lack of driving experience of young drivers because of which they do not take evasive maneuvers to reduce the impact of a crash in the making. Of course, other explanations are also possible. The parameter characterizing the effect of old age (age≥65) on driver injury severity suggests a higher injury risk propensity for this group of drivers relative to other adult individuals. As indicated in earlier studies (Bédard et al., 2002; Kim et al, 2013; Williams et al., 2003), older drivers tend to be slow in reacting to hazardous situations, may not be able to withstand crash impact forces well, and may suffer cognitive impairment and other medical conditions; all or some of these factors might contribute to their higher injury severity risk. It is interesting to note here that driver gender has no significant influence on crash severity outcome for segment 1. A plausible reason for this effect may be the additional physiological strength of male drivers (compared to female drivers) is less likely to lessen the effect of a more severe crash. Finally, in the category of driver characteristics, seat belt use significantly influences driver injury severity. The negative effect of this variable on the threshold separating the minor injury and serious/fatal injury levels indicates an increased likelihood of serious/fatal injuries for the drivers not wearing seat belts. The result can be explained by the reduction in restraint as well as possible high-risk driving behavior of those not using seatbelts (Obeng, 2008; Yau, 2004; Yasmin et al., 2012; Eluru and Bhat, 2007).

<u>Vehicle Characteristics:</u> The only vehicle characteristic influencing driver injury severity for the high risk segment is vehicle type. Table 3.3 shows that drivers in panel vans are associated with a lower injury risk propensity than drivers in other vehicle types, presumably because panel vans are larger and may offer more protection (Kockelman and Kweon, 2002; Xie et al., 2009; Eluru et al., 2010; Wang and Kockelman, 2005; Fredette et al., 2008).

Roadway Design Attributes: The roadway design attributes indicate a lower injury risk propensity for crashes occurring (a) on unpaved roads (perhaps because of very low speeds on such roads), (b) at intersections with some form of control for pedestrian movement and at roundabouts (relative to other types of intersections). The last result regarding roundabouts may be the consequence of moderated vehicle speeds and the angular movements at these locations, which can result in safer impact angles at the time of collision (Retting et al., 2001; Persaud et al., 2001; Chipman, 2004). On the other hand, crashes at stop-sign controlled intersections seem to increase injury severity risk relative to crashes at other intersections, attributable perhaps to non-compliance to stop signs and judgment problems (Chipman, 2004; Retting et al., 2003). Also, crashes occurring on very high speed roads, not surprisingly, lead to a high probability of serious/fatal injuries.

Environmental Factors: Time-of-day and lighting conditions are two of the environmental factors that significantly influence driver injury severity for the high risk segment. Injury risk reduces during the evening, but increases during the late night. The former effect may be a result of traffic congestion and slow driving speeds, because of which, when a crash does happen, the injury sustained tends to be rather mild. The latter result associated with late night crashes is well established in the literature; attributable to reduced visibility, fatigue, higher incidence of alcohol use, longer emergency response times, higher driver reaction time, and increased traffic speed (Plainis et al., 2006; Arnedt et al., 2001; Helai et al., 2008; Hu and Donnell, 2010; Kockelman and Kweon, 2002; de Lapparent, 2008). The lighting condition effect show a higher probabilty of no injury crashes during dark-lighted conditions, perhaps due to more cautious driving relative to broad daylight. As with the young driver effect, the impact of this variable on the other two injury severity categories is context-dependent.

Situational Factors: The presence of one or more passengers increases the probability of no injury, relative to the case of driving alone. This may be associated with public self-consciousness, where individuals behave and drive more responsibly with others around (Eluru et al., 2010). As expected, drivers who are ejected out of their vehicle during a crash have a high probability of sustaining serious/fatal injuries for the high risk segment.

3.4.4.4 Injury Severity Component: Low Risk Segment (Segment 2)

The injury severity component within the low risk segment (segment 2) is discussed in this section. The impact of exogenous variables within the low risk segment is different (for some variables) in magnitude as well as in sign from the impact of exogenous variables within the high risk segment. Also, the number of variables moderating the effect is different across the two segments.

Driver Characteristics: For the low risk segment, the influence of driver age on crash severity is along expected lines. We find that older drivers are associated with higher likelihood of severe crashes compared to other adult drivers as also seen in the other segment. Unlike the high risk segment, driver gender has a significant influence on driver injury severity outcome for low risk segment. The coefficient corresponding to driver gender of passenger vehicle reflects higher injury risk propensity for female drivers compared to male drivers perhaps because females are less capable of bearing physical and mental trauma compared to males (Evans, 2004; Sivak et al., 2010; Xie et al., 2009; Chen and Chen, 2011). As expected, our analysis showed an unequivocal benefit from employing seat belts. It is interesting to note that the seat belt variable affects the driver injury severity in different ways for the two segments.

Vehicle Characteristics: In the low risk segment, the results for the vehicle type reveal that the drivers of both station wagon and utility vehicles have a lower injury risk propensity, perhaps due to the larger weight of these vehicles. The vehicle age estimate demonstrates that drivers in older vehicles (Vehicle age 11 and above) have a higher risk propensity compared to the drivers in newer vehicles (vehicle age ≤ 10 years). The higher injury risk of drivers from older vehicles might be attributed to the absence of safety features, presence of mechanical defects, and/or the involvement of suspended and unlicensed drivers in these vehicles (Lécuyer and Chouinard, 2006, Kim et al, 2013; Islam and Mannering, 2006).

Roadway Design Attributes: The presence of traffic control devices significantly affect the severity of crashes. For both stop and yield sign variables, the corresponding latent propensity coefficients are negative indicating a lower injury risk; reduced travelling speed of drivers might be a plausible reason for such result. However, the effect of stop sign is strikingly different in the low risk segment compared to the impact of stop sign in the high risk segment. The different impacts in the two segments for stop sign highlight how the same variable can have distinct influence on injury severity based on the segment to which the driver is allocated.

The results for speed zone indicate that the injury propensity is higher for crashes occurring in zones with medium and higher speed limits relative to crashes occurring in lower speed limit zones. As is expected, within the two speed categories considered the higher speed category has a larger impact relative to the medium speed category. Such rapid increase in severity with progressive increase in speed limit has also been documented empirically by many earlier studies (Eluru et al., 2010; Chen et al., 2012; Tay and Rifaat, 2007).

Environmental Factors: The findings of the low risk segment indicate that if collisions occur in the winter season, the consequence is likely to be more injurious as compared to the accident in non-winter seasons (spring, summer and autumn). The prevalent adverse and damp weather conditions in winter might pose such risk on Victorian drivers. With respect to weather condition, the results presented in Table 3.3 indicate that the rainy/snowy weather condition results in more severe crashes compared to the clear weather, which may be attributed to the unfavourable driving conditions resulting from reduced visibility and reduced friction of the road surface. The results also reveal that injury propensity is higher for drivers in the presence of high wind compared to crashes occuring during clear weather. It is possible that under high wind conditions drivers suddenly lose vehicle control and sideswipe or run-off from their designated routes (Jung et al., 2011; Young and Liesman, 2007; Khattak and Knapp, 2001). With respect to lighting condition, the likelihood of driver injury risk propensity is found to be higher during dawn/dusk compared to other lighting conditions. This may be associated with sunglare during dawn/dusk period (Jurado-Piña et al., 2010; Gray and Regan,2007).

Situational factors: With respect to the situational factors, none of the variables are found to affect injury severity in the low risk segment.

3.5 Elasticity Effects

The parameter estimates of Table 3.3 do not directly provide the impact of exogenous variables on injury severity categories. On the other hand, the aggregate-level elasticity effects quantify the effects of these variables on driver injury severity outcomes. For this purpose, we compute the aggregate level "elasticity effects" for all independent variables and present the computed elasticities in Table 3.4. The effects are presented by injury severity categories for both the LSOL and LSGOL models for comparison purpose. The results in the table can be interpreted as the percentage change (increase for positive sign and decrease for negative sign) in the probability of the crash severity categories due to the change in that specific exogenous variable.

The following observations can be made based on the elasticity effects of the variables presented in Table 3.4. First, the most significant variables in terms of increase in serious/fatal injury (from both models) for drivers are driver age above 65, driver ejection, not wearing seat belts, and collision in high speed zone. In terms of serious/fatal injury reduction, the important factors are presence of pedestrian control, presence of roundabout, driving a panel van, unpaved road condition and presence of passengers. Second, the segmentation variables exhibit significant influence on injury severity profile with struck object collisions having the most significant contribution to increase in serious/fatal injury. Third, there are substantial differences in the elasticity effects of LSOL and LSGOL models. For instance, the LSOL model predicts an increase in minor injury for young driver while LSGOL model predicts a decrease in the same category. Such differences can also be observed for other variables − collision in dark-lighted condition, in the presence of one passenger and for pedestrian control.

3.6 Validation Analysis

We also carried out a validation experiment in order to ensure that the statistical results obtained above are not a manifestation of over fitting to data. 100 different data samples of about 2,500 records were generated randomly from the hold out validation sample consisting of 37,680 records to test the predictive performance of the estimated models. We evaluate both the aggregate and disaggregate measure of predicted fit by using these 100 different validation samples and present the average measure from the comparison, and also the confidence interval (C.I.), of the fit measures at 90% level.

At the disaggregate level we compute predictive log-likelihood (computed by calculating the log-likelihood for the predicted probabilities of the sample), predictive adjusted likelihood ratio index and probability of correct prediction (defined as an indicator if the observed outcome has the highest predicted probability). The results for disaggregate measures are presented in Table 3.5 (top row panel). At the aggregate level, root mean square error (RMSE) and mean absolute percentage error (MAPE) are computed by comparing the predicted and actual (observed) shares of injuries in each injury severity level for each set of full validation sample. The results for aggregate measure computation are also presented in Table 3.5 (bottom row panel). The comparison of LSOL and LSGOL model at the disaggregate level is far from conclusive with a slight edge to the LSGOL model. This is not surprising because the difference in BIC values for the two models is relatively small. The LSGOL model represents a clearly superior performance compared to that of the LSOL model at aggregate level.

3.7 Summary

This chapter formulates and estimates an econometric model for examining driver injury severity that accommodates systematic heterogeneity based on crash characteristics and relaxes the constant threshold assumption of traditional ordered logit model. The model is referred to as the latent segmentation based generalized ordered logit model. In traffic crash reporting, injury severity is typically characterized as an ordered variable resulting in application of the ordered response model for identifying the impact of exogenous variables. However, ordered systems impose a uni-directional impact of exogenous variables on injury severity alternatives. On the contrary, the generalized ordered logit model relaxes the restriction by allowing for the estimation of individual level thresholds as a function of exogenous variables. The widely employed ordered outcome model also restricts the impact of exogenous variables to be same across the entire population – homogeneity assumption. An alternative approach, referred to as latent segmentation approach, accommodates systematic heterogeneity by allocating the drivers probabilistically to different segments and by estimating segment specific models for each segment. The current chapter contributes to safety literature empirically by building on the GOL model − by formulating and estimating a latent segmentation based generalized ordered logit (LSGOL) model.

The empirical analysis was conducted using the Victoria crash database from Australia for the years 2006 through 2010. The model was estimated using a comprehensive set of exogenous variables - crash characteristics, driver characteristics, vehicle characteristics, roadway design attributes, environmental factors and situational factors. The empirical analysis involved the estimation of models using six different statistical frameworks: 1) OL, 2) GOL, 3) LSOL with two segments, 4) LSOL with three segments, 5) LSGOL with two segments and 6) LSGOL with three segments. The comparison exercise, based on information criterion metrics, highlighted the superiority of the LSGOL model with two segments on the estimation sample in terms of data fit compared to the other ordered outcome models. In the LSGOL approach, drivers were assigned probabilistically to two segments – high risk segment and low risk segment - based on a host of crash characteristics. In our empirical analysis, the crash characteristics that affected the allocation of drivers into segments include: collision object, trajectory of vehicle's motion and manner of collision. According to our results, the impact of exogenous variables in the low risk segment was different (for some variables) in magnitude as well as in sign from the impact of exogenous variables in the high risk segment.

In our research, to further understand the impact of various exogenous factors, elasticity effects were estimated for both the LSOL and LSGOL models for comparison purpose. The elasticity effects indicated that the most significant variables in terms of increase in serious/fatal injury (from both models) for drivers were driver age above 65, driver ejection, not wearing seat belts, and collision in high speed zone. In terms of serious/fatal injury reduction, the important factors were presence of pedestrian control, presence of roundabout, driving a panel van, unpaved road condition and presence of passengers. Further, the performance evaluation of these models on a validation sample revealed that the LSGOL model represents a clearly superior performance compared to that of the LSOL model at an aggregate level. But, the comparison of LSOL and LSGOL model at a disaggregate level was far from conclusive with a slight edge to the LSGOL model. In summary, the comparison exercise supports the hypothesis that LSGOL is a promising ordered outcome framework for accommodating population heterogeneity and for relaxing the fixed threshold assumption in the context of driver injury severity.

Figure 3.1: Schematic Diagram of Initial Point of Impact Relative to the Drivers' Seat Position

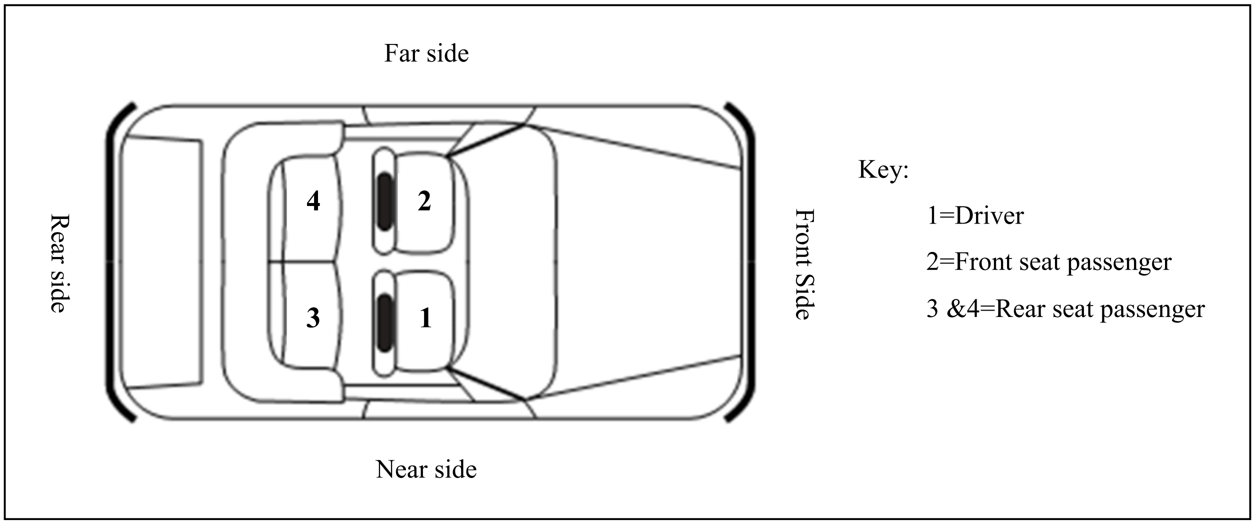


Table 3.1: Crash Database Sample Statistics

Variables	Sample Share	
	Frequency	Percentage
Crash characteristics		
Collision object		
Small Object	120	2.338
Large object	710	13.835
Collision with animals	207	4.034
Collision with Moving vehicle	4095	79.793
Trajectory of vehicle's motions		
Going Straight	2689	52.397
Other movement	2443	47.603
Manners of collision		
Rear-ender	469	9.139
Rear-ended	582	11.341
Near-sideswipe	96	1.871
Near-angular	729	14.205
Short-side angular	1133	22.077
Far-side	783	15.257
Head-on	303	5.904
Other collisions (Struck object)	1037	20.207
Driver characteristics		
Driver age		
Age less than 25	1350	26.306
Age 26 to 65	3282	63.952
Age above 65+	500	9.743
Driver gender		
Female	2420	47.155
Male	2712	52.845
Restraint system use		
Seat belt not used	179	3.488
Seat belt used	4953	96.512
Vehicle characteristics		
Vehicle Type		
Sedan	3666	71.434
Station wagon	900	17.537
Utility	447	8.710
Panel van	119	2.319
Vehicle age		
Vehicle age less than 11	2907	56.645
Vehicle age 11 and above	2225	43.355
Roadway design attributes		
Type of road surface		

Paved	4901	95.499
Unpaved	231	4.501
Traffic Control Device		
No control	3088	60.171
Signal	1107	21.571
Pedestrian control	34	0.663
Roundabout	181	3.527
Stop sign	171	3.332
Yield sign	473	9.217
Other traffic control	78	1.520
Speed zone		
Low speed zone	961	18.726
Medium speed zone	3412	66.485
High speed zone	759	14.790
Environmental factors		
Season		
Summer	1305	25.429
Autumn	1346	26.228
Winter	1265	24.649
Spring	1216	23.694
Time of day		
Morning peak	691	13.465
Off peak	1695	33.028
Evening peak	1250	24.357
Late evening	1160	22.603
Late night	336	6.547
Weather condition		
Clear	4332	84.412
Rain/snow	646	12.588
High wind	85	1.656
Other weather condition	69	1.345
Lighting condition		
Daylight	3558	69.330
Dusk/dawn	361	7.034
Dark-lighted	942	18.355
Dark-unlighted	271	5.281
Situation factors		
Presence of passengers		
No passenger	598	11.652
One passenger	2395	46.668
Two passengers	1136	22.136
More than two passengers	1003	19.544
Driver ejected out of the vehicle		
Ejected out	39	0.800
Did not Ejected out	5093	99.200

Table 3.2: Segment Characteristics and Mean Values of Segmentation Variables for LSGOL model

Segmentation Characteristics			
Components		**Segments**	
		Segment 1	**Segment 2**
Driver population share		0.418	0.582
Injury Severity	Property damage only	0.102	0.640
	Minor injury	0.578	0.239
	Serious Injury and Fatal Injury	0.320	0.121

Table 3.3: LSGOL Estimates

Segmentation Components								
Variables	**Segment 1**				**Segment 2**			
					Estimate		**t-stat**	
Constant	–		–		1.435		5.679	
Crash characteristics								
Collision object (Base: Other moving vehicle and animals)								
Small Object	–		–		-3.531		-4.552	
Large object	–		–		-4.035		-8.342	
Trajectory of vehicle's motions (Base: Other movement)								
Going Straight	–		–		-0.480		-3.423	
Manner of collision (Base: Rear-ender and short side-angular)								
Rear-ended	–		–		-1.574		-7.172	
Near-angular	–		–		-0.826		-4.427	
Head-on	–		–		-1.055		-4.402	
Far-side	–		–		0.530		2.197	
Injury Severity Components								
Variables	**Latent Propensity**		**Threshold**		**Latent Propensity**		**Threshold**	
	Estimates	**t-stat**	**Estimate**	**t-stat**	**Estimates**	**t-stat**	**Estimate**	**t-stat**
Constant	3.859	7.962	1.468	13.571	-1.746	-6.237	0.307	2.650
Driver characteristics								

Driver age (Base: Age 25 to 64)								
Age less than 25	-0.734	-2.478	-0.137	-1.643	–	–	–	–
Age above 65+	0.885	3.779	–	–	0.432	2.629	-0.498	-2.533
Driver gender (Base: Male)								
Female	–	–	–	–	1.189	8.003	0.316	2.740
Restraint system use (Base: seat belt used)								
Seat belt not used	–	–	-0.197	-2.161	0.717	2.472	–	–
Vehicle characteristics								
Vehicle Type (Base: Sedan)								
Station wagon	–	–	–	–	-0.744	-4.452	–	–
Utility	–	–	–	–	-0.939	-3.313	–	–
Panel van	-1.109	-2.495	–	–	–	–	–	–
Vehicle age (Base: Vehicle age less than 10)								
Vehicle age 11 and above	–	–	–	–	0.349	3.309	–	–
Roadway design attributes								
Type of road surface (Base: Paved)								
Unpaved	-1.338	-1.901	–	–	–	–	–	–
Traffic Control Device (Base: None traffic control and other control device)								
Pedestrian control	-2.135	-1.974	–	–	–	–	–	–
Roundabout	-1.292	-3.016	–	–	–	–	–	–
Stop sign	0.900	2.579	–	–	-1.326	-3.273	–	–
Yield sign	–	–	–	–	-0.471	-2.527	–	–

Speed zone (Base: Low speed ≤50 km/h)								
Medium speed (60-90 km/h)	–	–	–	–	0.356	2.469	–	–
High speed (≥100 km/h)	–	–	-0.189	-4.468	1.244	5.934	–	–
Environmental factors								
Season (Base: Spring, Summer, Fall)								
Winter	–	–	–	–	0.332	2.886	–	–
Time of day (Base: Morning peak, Off peak and Late evening)								
Evening peak	-0.751	-4.61	–	–	–	–	–	–
Late night	0.586	3.146	–	–	–	–	–	–
Weather condition (Base: Clear)								
Rain/snow	–	–	–	–	0.314	2.069	–	–
High wind	–	–	–	–	0.736	2.174	–	–
Lighting condition (Base: Daylight)								
Dusk/dawn	–	–	–	–	0.488	2.711	–	–
Dark-lighted	-0.897	-2.932	-0.408	-4.124	–	–	–	–
Situational factors								
Presence of passengers (Base: No passenger)								
One passenger	-0.532	-4.040	–	–	–	–	–	–
Two passengers	-2.201	-5.877	-0.426	-3.947	–	–	–	–
Driver ejected out of the vehicle	–	–	-0.929	-2.624	–	–	–	–

Table 3.4: Elasticity Effects

Variables	LSOL			LSGOL		
	No injury	Minor injury	Serious/Fatal injury	No injury	Minor injury	Serious/Fatal injury
Crash characteristics						
Rear-ended	-42.536	30.199	30.639	-38.906	28.750	26.138
Near-sideswipe	-18.587	13.139	13.488	–	–	–
Near-angular	-24.690	17.491	17.850	-19.507	14.382	13.162
Head-on	-31.219	22.105	22.591	-25.717	18.958	17.357
Far-side	–	–	–	11.630	-8.488	-7.998
Small Object	-68.129	48.024	49.677	-68.894	50.634	46.768
Large object	-82.773	56.013	64.419	-81.995	57.905	59.764
Going Straight	-12.860	9.054	9.396	-10.607	7.776	7.234
Driver characteristics						
Age less than 25	3.170	4.548	-14.126	12.339	-7.676	-10.800
Age above 65+	-20.643	-16.127	68.500	-17.328	-22.938	73.855
Female	-32.773	21.423	26.821	-33.961	27.437	18.742
Seat belt not used	-19.413	-4.348	45.571	-20.767	-4.876	49.151
Vehicle characteristics						
Station wagon	18.506	-12.480	-14.478	19.257	-12.312	-16.277
Utility	23.853	-16.822	-17.378	23.021	-15.277	-18.486
Panel van	13.191	6.629	-37.366	11.199	8.029	-35.904

Vehicle age 11 and above	-8.629	5.284	7.684	-9.667	5.566	9.241
Roadway design attributes						
Unpaved	15.147	5.989	-40.081	14.465	7.321	-41.066
Pedestrian control	22.987	2.511	-49.365	27.860	-0.111	-54.360
Roundabout	16.863	6.004	-43.463	13.554	8.069	-40.585
Stop sign	24.148	-35.585	14.733	24.694	-36.275	14.788
Yield sign	12.034	-7.910	-9.772	12.339	-7.676	-10.800
Medium speed zone	-8.934	5.619	7.697	-9.555	5.667	8.846
High speed zone	-39.029	4.574	68.420	-36.141	0.485	69.923
Environmental factors						
Winter	-7.986	4.794	7.279	-9.299	5.214	9.133
Evening peak	6.299	7.697	-25.740	6.238	8.905	-27.715
Late night	-4.434	-10.787	27.471	-3.696	-9.685	24.095
Rain/snow	-7.323	4.335	6.780	-8.852	4.861	8.873
High wind	-21.818	11.163	23.256	-21.333	10.121	24.155
Dusk/dawn	-15.040	8.381	14.835	-13.939	7.298	14.592
Dark-lighted	-2.359	-4.079	11.724	7.797	-17.138	14.563
Situation factors						
One passenger	-1.491	10.792	-15.884	4.042	7.282	-20.590
Two passenger	10.557	9.688	-37.540	24.627	-8.417	-33.573
Driver ejected out of the vehicle	-1.010	-32.988	16.292	0.000	-47.556	82.777

Table 3.5: Measures of Fit in Validation Sample

Disaggregate Level (100 Validation Samples)					
Summary statistic		**LSOL**		**LSGOL**	
		Predictions	**C.I.**	**Predictions**	**C.I.**
Number of observations		2547.540	–	2547.540	–
Number of parameters		37.000	–	41.000	–
Log-likelihood at zero		-2798.759	-2807.853 | -2789.665	-2798.759	-2807.853 | -2789.665
Log-likelihood at sample shares		-2696.838	-2706.066 | -2687.609	-2696.838	-2706.066 | -2687.609
Predictive Log-likelihood		-2438.553	-2447.559 | -2429.546	-2433.249	-2442.324 | -2424.174
Predictive adjusted likelihood ratio index		0.082	0.081 | 0.083	0.083	0.082 | 0.084
Average probability of correct prediction		0.531	0.529 | 0.533	0.531	0.529 | 0.532
Aggregate Level (100 Validation Samples)					
Injury categories/ Measures of fit	**Actual shares**	**LSOL**		**LSGOL**	
		Predictions	**C.I.**	**Predictions**	**C.I.**
No injury	42.522799	42.309	42.247 | 42.371	42.364	42.300 | 42.428
Non-incapacitating injury	36.518347	36.595	36.559 | 36.629	36.492	36.452 | 36.531
Incapacitating/Fatal injury	20.958854	21.096	21.049 | 21.142	21.144	21.099 | 21.189
RMSE	–	0.821	0.753 | 0.888	0.817	0.753 | 0.882
MAPE	–	2.352	2.351 | 2.354	2.343	2.341 | 2.345

CHAPTER 4 Examining Driver Injury Severity in Two Vehicle Crashes – A Copula Based Approach

4.1 Introduction

The most commonly identified exogenous factor that significantly affects traffic crash injury severity outcome is the collision type variable. Most of the earlier studies consider the collision type variable as an explanatory variable in modeling injury severity (except Ye et al., 2008 and Rana et al., 2010). In this approach, the analyst imposes the assumption that the injury severity profile for vehicle occupants in all types of crashes is the same and any potential differences between different collision types can be accurately captured by employing the collision type variable as an explanatory variable. However, it is possible that various collision types might lead to distinct vehicle occupant injury severity profiles i.e., the overall manifestation of injury severity is different by collision type. Thus, estimating a single injury severity model, when such distinct profiles of injury severity exist, will result in incorrect and biased estimates. Moreover, it is possible that the collision type and resulting injury severity are influenced by the same set of observed and unobserved factors. To resolve such an issue, this chapter endeavours to develop a closed form copula based framework to accommodate the impact of observed and unobserved effects on collision type and injury severity. The approach allows for flexible dependency structures across joint dimensions while retaining the closed form structure (see Bhat and Eluru, 2009). The proposed model is estimated using driver injury severity data for two vehicle crashes from the state of Victoria, Australia for the years 2006 through 2010 employing a comprehensive set of exogenous variables − driver characteristics, vehicle characteristics, roadway design attributes, environmental factors and crash characteristics. In summary, the current chapter contributes to safety literature on driver injury severity both methodologically and empirically. In terms of methodology, we formulate and estimate a copula-based multinomial logit-ordered logit framework to jointly analyze the collision type and injury severity outcome in a two-vehicle crash. Our study also accommodates the potential heterogeneity (across drivers) in the dependency effect of collision type and injury severity outcome within a closed form copula framework. In terms of empirical analysis, our study incorporates collision type as a vehicle level variable and addresses the inconsistency from earlier research while also examining the impact of a comprehensive set of exogenous variables on driver injury severity.

The rest of the chapter is organized as follows. Section 4.2 provides details of the econometric model framework used in the analysis. In Section 4.3, the data source and sample formation procedures are described. The model results and elasticity effects are presented in Section 4.4. Section 4.5 concludes the chapter and presents directions for future research.

4.2 Model Framework

The focus of the current chapter is to jointly model the collision type and injury severity outcome of drivers involved in a two vehicle collisions using a copula-based joint multinomial logit-ordered logit modeling framework. The analysis in this chapter focuses on driver injury severity in a crash. In this section, econometric formulation for the joint model is presented.

4.2.1 The Collision Type Outcome Model Component

Let q $(q = 1,2, \dots\dots, Q)$ and k $(k = 1,2, \dots\dots, K)$ be the indices to represent driver and collision type, respectively. Let j be the index for the discrete outcome that corresponds to the injury severity level j $(j = 1,2, \dots\dots, J)$ of driver q. In the joint framework, the modeling of collision type is undertaken using the multinomial logit structure. Thus, the propensity of a driver q involving in a collision of specific collision type k takes the form of:

$$u_{qk}^{*} = \beta_k x_{qk} + \xi_{qk} \tag{4.1}$$

where, x_{qk} is a column vector of exogenous variable, β_k is a row vector of unknown parameters specific to collision type k and ξ_{qk} is an idiosyncratic error term (assumed to be standard type-I extreme value distributed) capturing the effects of unobserved factors on the propensity associated with collision type k. A driver q is assumed to be involved in a collision type k if and only if the following condition holds:

$$u_{qk}^{*} > \max_{l=1,2,\dots,k,\ l\neq k} u_{ql}^{*} \tag{4.2}$$

The condition presented in equation 4.2 can be equivalently represented as a series of binary outcome models for each collision type, k (see Lee, 1983). For example, let η_{qk} be a dichotomous variable with $\eta_{qk} = 1$ if a driver q ends up in a collision type k and $\eta_{qk} = 0$ otherwise. Now, let us define v_{qk} as follows:

$$v_{qk} = \xi_{qk} - \left\{ \max_{l=1,2,\ldots,k,\ l \neq k} u_{ql}^{*} \right\} \tag{4.3}$$ [7]

By substituting the right side for u_{qk}^{*} from equation 4.1 in equation 4.2, we can write:

$$\eta_{qk} = 1 \ \ if \ \beta_k' x_{qk} + v_{qk} > 0 \tag{4.4}$$

The system in equation 4.4 represents the multinomial discrete outcome model of collision type as an equivalent series of binary outcome model formulation, one for each collision type k. In equation 4.4, the probability expression of collision type outcome is dependent on the distributional assumption of v_{qk}, which in turn depends on the distributional assumption of ξ_{qk}. Thus an assumption of independent and identical Type 1 Gumbel distribution for ξ_{qk} results in a logistic distributed v_{qk}. Consequently, the probability expression for the corresponding discrete outcome (collision type) model resembles the multinomial logit probability expression as follows:

$$\Lambda_k(\beta_k x_{qk}) = Pr(v_{qk} > -\beta_k x_{qk}) = \frac{\sum_{l \neq k} exp(\beta_k x_{ql})}{exp(\beta_k x_{qk}) + \sum_{l \neq k} exp(\beta_k x_{ql})} \tag{4.5}$$

4.2.2 <u>The Injury Severity Outcome Model Component</u>

In the joint model framework, the modeling of driver injury severity is undertaken using an ordered logit specification. In the ordered outcome model, the discrete injury severity levels (y_{qk}) are assumed to be associated with an underlying continuous latent variable (y_{qk}^{*}). This latent variable is typically specified as the following linear function:

$$y_{qk}^{*} = \alpha_k z_{qk} + \varepsilon_{qk}\ ,\ \ y_{qk} = j_k, if\ \tau_{k,j-1} < y_{qk}^{*} < \tau_{k,j} \tag{4.6}$$

where, y_{qk}^{*} is the latent injury risk propensity for driver q if he/she was involved in a collision type k, z_{qk} is a vector of exogenous variables, α_k is a row vector of unknown parameters and ε_{qk} is a random disturbance term assumed to be standard logistic. $\tau_{k,j}$ $(\tau_{k,0} = -\infty\ , \tau_{k,J} = \infty)$ represents

[7] The reader would note that the v_{qk} term applied here is different from the Lee's transformation. If one uses a symmetric distribution, that allows both positive and negative dependencies (such as the Gaussian copula proposed by Lee), then Lee's formulation would be adequate. However, when testing various copulas, some of which allow asymmetric and only positive dependencies, it is important to test our version as well as Lee's formulation to ensure we capture the dependencies in asymmetric copulas. We formulate the model in this form because we expect that the dependency for collision type and subsequent injury to be positively correlated (due to unobserved factors, see Portoghese et al., 2011 for a similar formulation in a different context).

the threshold associated with severity level j for collision type k, with the following ordering conditions: $\left(-\infty < \tau_{k,1} < \tau_{k,2} < \ldots\ldots\ldots < \tau_{k,J-1} < +\infty\right)$. Given these relationships across the different parameters, the resulting probability expressions for driver q sustaining an injury severity level j in a collision type k take the following form:

$$Pr(y_{qk} = j_k) = \Lambda_k\left(\tau_{k,j} - \alpha_k' z_{qk}\right) - \Lambda_k\left(\tau_{k,j-1} - \alpha_k' z_{qk}\right) \tag{4.7}$$

where, $\Lambda_k(.)$ is the standard logistic cumulative distribution function. The probability expression of equation 4.7 represents the independent injury severity model for a collision type k.

4.2.3 <u>The Joint Model: A Copula-based Approach</u>

The collision type and the injury severity component discussed in previous two subsections may be brought together in the following equation system:

$$\begin{aligned} &\eta_{qk} = 1 \;\; if\ \beta_k x_{qk} > v_{qk} \\ &y_{qk}^* = \alpha_k z_{qk} + \varepsilon_{qk}\ ,\;\; y_{qk} = 1\left[\eta_{qk} = 1\right] y_{qk}^* \end{aligned} \tag{4.8}$$

However, the level of dependency between the underlying collision type outcome and the injury severity level of driver depends on the type and extent of dependency between the stochastic terms v_{qk} and ε_{qk}. These dependencies (or correlations) are explored in the current study by using a copula-based approach. A copula is a mathematical device that identifies dependency among random variables with pre-specified marginal distribution (Bhat and Eluru (2009) and Trivedi and Zimmer (2007) provide a detailed description of the copula approach). In constructing the copula dependency, the random variables $(v_{qk}\ and\ \varepsilon_{qk})$ are transformed into uniform distributions by using their inverse cumulative distribution functions, which are then coupled or linked as a multivariate joint distribution function by applying the copula structure. Let us assume that $\Lambda_{vk}(.)$ and $\Lambda_{\varepsilon k}(.)$ are the marginal distribution of v_{qk} and ε_{qk}, respectively and $\Lambda_{vk,\varepsilon k}(.,.)$ is the joint distribution of v_{qk} and η_{qk}. Subsequently, a bivariate distribution $\Lambda_{vk,\varepsilon k}(v,\varepsilon)$ can be generated as a joint cumulative probability distribution of uniform [0, 1] marginal variables U_1 and U_2 as below:

$$\begin{aligned} &\Lambda_{vk,\varepsilon k}(v,\varepsilon) = Pr(v_{qk} < v,, \varepsilon_{qk} < \varepsilon) \\ &= [\Lambda_{vk}^{-1}(U_1) < v, \Lambda_{\varepsilon k}^{-1}(U_2) < \varepsilon\,] \end{aligned} \tag{4.9}$$

$= [U_1 < \Lambda_{vk}(v), U_2 < \Lambda_{\varepsilon k}(\varepsilon)\ ]$

The joint distribution (of uniform marginal variable) in equation 4.9 can be generated by a function $C_{\theta q}(.,.)$ (Sklar, 1973), such that:

$$\Lambda_{vk,\varepsilon k}(v, \delta_2) = C_{\theta q}(U_1 = \Lambda_{vk}(v), U_2 = \Lambda_{\varepsilon k}(\varepsilon)) \tag{4.10}$$

where $C_{\theta q}(.,.)$ is a copula function and θ_q the dependence parameter defining the link between v_{qk} and ε_{qk}. It is important to note here that, the level of dependence between collision type and injury severity level can vary across drivers. Therefore, in the current study, the dependence parameter θ_q is parameterized as a function of observed crash attributes as follows:

$$\theta_q = fn(\gamma_k s_{qk}) \tag{4.11}$$

where, s_{qk} is a column vector of exogenous variable, γ_k is a row vector of unknown parameters (including a constant) specific to collision type k and fn represents the functional form of parameterization. Based on the dependency parameter permissible ranges, alternate parameterization forms for the six copulas are considered in our analysis. For Normal, Farlie-Gumbel-Morgenstern (FGM) and Frank Copulas we use $\theta_q = \gamma_k s_{qk}$, for the Clayton copula we employ $\theta_q = exp(\gamma_k s_{qk})$, and for Joe and Gumbel copulas we employ $\theta_q = 1 + exp(\gamma_k s_{qk})$.

4.2.4 Estimation Procedure

The joint probability that the driver q gets involved in a collision type k and sustaining injury severity level j, from equation 4.5 and 4.7, can be written as:

$$\begin{aligned} &Pr(\eta_{qk} = 1, y_{qk} = j_k) \\ &= Pr\left\{(\beta_k x_{qk} > v_{qk}), \left((\tau_{k,j-1} - \alpha_k z_{qk}) < \varepsilon_{qk} < \ (\tau_{k,j} - \alpha_k z_{qk})\right)\right\} \\ &= Pr\left((\beta_k x_{qk} > v_{qk}), (\varepsilon_{qk} < \tau_{k,j} - \alpha_k z_{qk})\right) \\ &\qquad - \ Pr\left((\beta_k x_{qk} > v_{qk}), (\varepsilon_{qk} < \tau_{k,j-1} - \alpha_k z_{qk})\right) \end{aligned} \tag{4.12}$$

$$= \Lambda_{\varepsilon k}(\tau_{k,j} - \alpha_k z_{qk}) - \ \Lambda_{\varepsilon k}(\tau_{k,j-1} - \alpha_k z_{qk}) - \big(Pr[v_{qk} < -\beta_k x_{qk}\,, \varepsilon_{qk} < (\tau_{k,j} - \alpha_k z_{qk})] - Pr[v_{qk} < -\beta_k x_{qk}\,, \varepsilon_{qk} < \ (\tau_{k,j-1} - \alpha_k z_{qk})]\ \big)$$

The joint probability of equation 4.12 can be expressed by using the copula function in equation 4.10 as:

$$
\begin{aligned}
Pr(\eta_{qk} = 1, y_{qk} = j_k) \\
&= \Lambda_{\varepsilon k}(\tau_{k,j} - \alpha_k z_{qk}) - \Lambda_{\varepsilon k}(\tau_{k,j-1} - \alpha_k z_{qk}) \\
&- \left[C_{\theta q}\left(U_{q,j}^k, U_q^k\right) - C_{\theta q}\left(U_{q,j-1}^k, U_q^k\right)\right]
\end{aligned} \tag{4.13}
$$

where $U_{q,j}^k = \Lambda_{\varepsilon k}(\tau_{k,j} - \alpha_k z_{qk})$, $U_q^k = \Lambda_{vk}(-\beta_k x_{qk})$ (4.14)

Thus the likelihood function with the joint probability expression in equation 4.13 for collision type and driver injury severity outcomes can be expressed as:

$$
L = \prod_{q=1}^{Q}\left[\prod_{k=1}^{K}\prod_{j=1}^{J}\left\{Pr(\eta_{qk} = 1, y_{qk} = j_k)\right\}^{\omega_{qkj}}\right] \tag{4.15}
$$

where, ω_{qkj} is dummy with $\omega_{qkj} = 1$ if the driver q sustains collision type k and an injury severity level of j and 0 otherwise. All the parameters in the model are then consistently estimated by maximizing the logarithmic function of L. The parameters to be estimated in the model are: β_k in the MNL component, α_k and $\tau_{k,j}$ in OL component, and finally γ_k in the dependency component. In our analysis we employ six different copulas structure - the Gaussian copula, the Farlie-Gumbel-Morgenstern (FGM) copula, and set of Archimedean copulas including Frank, Clayton, Joe and Gumbel copulas (a detailed discussion of these copulas is available in Bhat and Eluru, 2009).

4.3 Data

4.3.1 Data Source

Data for our empirical analysis of the current chapter is sourced from the Victoria crash database of Australia for the years 2006 through 2010. The dataset has been briefly described in Section 3.3.1 of Chapter 3.

4.3.2 Sample Formation and the Dependent Variables

This study is confined to the injury severity outcome of drivers, who are involved in a two passenger vehicle collisions. Crashes involving only one vehicle or more than two vehicles are not included in the analysis. The crashes that involve commercial vehicles are also excluded to avoid the potential systematic differences between the crashes involving commercial and non-commercial driver groups.

In our analysis, the crash outcome is defined as the injury severity level sustained by the driver in each vehicle of the two vehicle collisions. The final dataset, after removing records with missing information for essential attributes consisted of about 34,278 driver records. In this final sample of drivers, the percentage of fatal crashes sustained by drivers is extremely small (0.40%). Therefore, both the fatal and serious injury categories are merged together. From this dataset, a sample of 8,509 driver records is randomly selected for the purpose of estimating models. In the final estimation sample, the distributions of the three driver injury severity levels are as follows: no injury 49.50%, minor injury 34.50% and serious/fatal injury 16.00%.

As discussed in section 3.4.1, depending on the initial point of impact it is possible that the different vehicles involved in the same crash might have significantly different crash profiles. Hence, it is incorrect to assign the same collision type variable to all vehicles involved in the same crash in analyzing vehicle occupant injury severity. Therefore, in estimating driver injury severity model, we compile the types of collision at a high level of disaggregation, and as a combination of collision type (rear-end, sideswipe, angular, and head-on) and the initial point of contact[8].

A schematic diagram of the initial point of impact relative to the driver's seat position is shown in Figure 3.1 of the preceding chapter. Based on the collision type and the point of impact, we identified eight categories for the "collision type": Rear-ender (the rear vehicle that is involved in rear-end collision), Rear-ended (the front vehicle that is involved in the rear-end collision), Near-sideswipe (sideswipe/near-side), Far-sideswipe (sideswipe/far-side), Near-angular (angular/near-side), Far-angular (angular/far-side), Short-side angular (angular/front and rear side) and Head-on (head-on/front side). In the final estimation sample, the distribution of collision type variable is as follows: rear-ender 11.91%, rear-ended 14.29%, near-sideswipe 2.49%, far-sideswipe 3.04%, near-angular 17.95%, far-angular 16.61%, short-side angular 26.80% and head-on 6.92%.

Table 4.1 offers a summary of the sample characteristics of collision type and injury severity level sustained by drivers. From the descriptive analysis, it is evident that the injury severity distributions vary substantially by collision type. More interestingly, we observe that for

[8] It is worthwhile to mention here that several previous studies (Tsui et al., 2009; Schiff and Cummings, 2004; Loo and Tsui, 2007) have examined the reliability of crash related factors documented in police-reported crash databases. The unreliability in reporting is mostly observed for casualty of crash, occupant position in the vehicle, demographics and seat-belt. Compiling crash details based on collision type and initial point of impact are less likely to be error prone. More importantly, the incompleteness of these variables in the Victorian crash database is approximately zero (zero for collision type and 0.3% for initial point of impact).

collision types within the same accident, rear-ender vs. rear-ended, near-sideswipe vs. far-sideswipe exhibit huge differences in the injury severity distribution. These observations highlight the need to define the collision type variable at a vehicle level rather than at the crash level. The descriptive analysis identifies head-on as the most serious collision type in terms of severe injuries while far-sideswipe crashes result in the least severe injuries. Further, Table 4.2 offers a summary of the sample characteristics of explanatory variables across different collision types. It can be observed from Table 4.2 that the proportions of different variables vary substantially across different collision types.

4.4 Empirical Analysis

4.4.1 Variables Considered

The collision attributes considered in the empirical study can be grouped into the following five broad categories:

- Driver characteristics including driver age, gender, seat belt use and local driver information;
- Vehicle characteristics including vehicle type (characterized as sedan, station wagon, utility and panel van) and vehicle age;
- Roadway design attributes including type of road surface, presence of traffic control device, speed zones and type of intersection;
- Environmental factors including time of day, day of week, weather condition, surface condition and lighting condition; and
- Crash characteristics including presence of passenger and trajectory of vehicle's motion.

The final specification of the model development was based on combining the variables when their effects were not statistically different and by removing the statistically insignificant variables in a systematic process based on statistical significance (90% confidence level). The coefficient estimates across different collision types were also restricted to be same when the effects were not significantly different.

4.4.2 Model Specification and Overall Measures of Fit

The empirical analysis involves estimation of models by using six different copula structures: 1) Gaussian, 2) FGM, 3) Clayton, 4) Gumbel, 5) Frank and 6) Joe (a detailed discussion of these copulas is available in Bhat and Eluru, 2009). The empirical analysis involved a series of model estimations. *First*, an independent copula model (separate MNL and OL models) was estimated to establish a benchmark for comparison. *Second*, 6 different models that restricted the dependency parameters across the eight collision types and injury severity models to be the same were estimated. *Third*, based on the copula parameter significance for each collision type, copula models that allow for different dependency structures for different collision type and injury severity combinations were estimated (for example Frank copula for the first three collision types Clayton copula for other collision types). Finally, to determine the most suitable copula model (including the independent copula model), a comparison exercise was undertaken. The alternative copula models estimated are non-nested and hence, cannot be tested using traditional log-likelihood ratio test. We employ the Bayesian Information Criterion (BIC) to determine the best model among all copula models (see Trivedi and Zimmer, 2007; Quinn, 2007; Eluru et al., 2010). The BIC values are computed as shown in equation 2.18. The model with the *lower* BIC is the preferred copula model. With exclusively a single copula dependency structure, the best model fit is obtained with Clayton. However, the lowest BIC value was obtained for a combination model of Frank-Clayton copulas (Frank copula structure for rear-ender and head-on collision and Clayton dependency structure with the remaining collision type). The copula model BIC comparisons confirm the importance of accommodating dependence between collision type and injury severity outcome in the analysis of driver injury severity.

4.4.3 Estimation Results

In presenting the effects of exogenous variables in the joint model specification, we will restrict ourselves to the discussion of the Frank-Clayton specification. For the ease of presentation, the collision type component (Table 4.3) and injury severity component (Table 4.4) are presented and discussed separately. The copula parameters are presented in the last row panel of Table 4.3.

4.4.3.1 Collision Type Component

The coefficients in Table 4.3 represent the effect of exogenous variables on each collision type category relative to the base category. In the following sections, the estimation results are discussed by variable groups.

Driver Characteristics: The impact of driver age on collision type indicates that young drivers are more likely to be the rear-ender and are less likely to be rear-ended in crashes relative to the adult drivers, perhaps reflecting a lack of driving experience and/or poor judgement and/or a greater risk-taking/aggressive driving propensity. The likelihood of being rear-ended or being involved in a far-sideswipe collision is lower for the older drivers. However, the older drivers are also more likely to be involved in angular collision (far- or near-angular) compared to the adult drivers, which might be a manifestation of longer time requirements for older drivers in complete turning movements (Alexander et al., 2002). Female drivers are more likely to be rear-ended or involved in a near-angular collision, while the odds of involving in a head-on collision is lower for female drivers compared to their male counterparts. The results also highlight that drivers who do not wear seat-belts are more likely to hit another vehicle from behind, a possible reflection of inherent aggressive personality of these drivers.

Vehicle Characteristics: The effects of the vehicle characteristics indicate that the drivers of utility and panel van are more likely to be the rear-ender, while the likelihood of being involved in head-on collisions are also higher for the driver of utility vehicle compared to other drivers. These results point towards aggressive attitude in driving and a false sense of security among large vehicle owners. The vehicle age variables suggest that compared to the drivers of newer vehicles (vehicle age less than 6), the drivers of older vehicles (vehicle age 6-10 or vehicle age 11 and above) are less likely to be rear-ended or involved in any form of sideswipe or angular collision (the effect of vehicle age 6-10 is insignificant for far-sideswipe collision).

Roadway Design Attributes: Among the roadway design attributes, the effect of roadway surface type is significant only for the head-on collision with positive coefficient for the gravel road surface compared to the paved and unpaved roads. Usually, gravel roads are associated with fewer lanes increasing the odds of head-on collisions as the lanes are unlikely to be median separated.

The estimation results corresponding to the presence of traffic control device highlight that the presence of traffic signal is associated with less sideswipe and head-on collision. Drivers are more likely to be rear-ended or involved in near-angular collision in the presence of roundabout. In the presence of a stop sign, the likelihood of rear-end, far-sideswipe and head-on collisions are lower, whereas the likelihood of near-angular collision is higher. The presence of yield sign has positive association with rear-ended and near-angular collision and negative association with sideswipe and head-on collision.

With respect to speed zone, the medium speed limit zone indicator increases the likelihood of rear-end collision; while the high speed limit zone indicators reveal increased likelihood of rear-ended, side-swipe and head-on collisions. The presence of T-intersection increases the odds of all collision types (except for short-side angular). Five or more legged intersection is positively correlated with the occurrence of rear-end and far-sideswipe collision. The variable representing the location as a non-intersection is associated with higher crash propensity for all collision types except for far- and short side-angular collision.

Environmental Factors: The effects of environmental factors indicate that the occurrence of far-angular collision is less at late night compared to the other times of day. Crashes occurring on wet surface are more likely to be head-on collision than those occurring on the dry surface condition. Far-sideswipe collision is less likely to occur at dawn/dusk period relative to the daylight period. Dark-lighted condition results in reduced likelihood of rear-end collision. However, dark-unlighted condition is associated with high risk of head-on collision. During weekend, drivers are less likely to be involved in rear-ended crashes, but are more likely to be involved in far-sideswipe and head-on crashes.

Crash Characteristics: Among the crash characteristic variables considered, none of the variables show significant impact on collision type occurrence.

4.4.3.2 Dependence Effects

As indicated earlier, the estimated Frank-Clayton copula based MNL-OL model provides the best fit in incorporating the correlation between the collision type and injury severity outcome. An examination of the copula parameters presented in the last row panel of Table 4.3 highlights the presence of common unobserved factors affecting collision type and injury severity. The Frank copula dependency structure is associated with the rear-ender and head-on collision types, while the Clayton dependency structure is associated with the rest of the six collision types. Further, except for far-angular collision type, all other copula dependencies are characterized by at least one additional exogenous variable. This provides support to our hypothesis that the dependency structures are not constant across the entire database. The various exogenous variables that contribute to the dependency include Female (rear-ender), medium speed limit (near-angular and head-on), yield sign (rear-ended), utility vehicle (near-sideswipe), late night (far-sideswipe) and high wind (near-angular and short-side angular). The Frank copula offers a symmetric dependency structure i.e. a positive coefficient represents a positive dependency while negative coefficient represents negative dependency. The exact nature of the dependency for the Frank copula is based on the realized coefficient for rear-ender and head-on crash types considering all significant variables. For the Clayton copula, the dependency is entirely positive and the coefficient sign and magnitude reflects whether a variable increases or reduces the dependency and by how much. The proposed framework by allowing for such parameterizations allows us to improve the model estimation results.

4.4.3.3 Injury Severity Component

The coefficients in Table 4.4 represent the effect of exogenous variables on injury severity outcome of drivers for each collision type category. The results suggest that the impact of exogenous variables vary (for some variables) in magnitude as well as in sign across collision types. The impacts of these variables are also substantially different from the estimates of the independent MNL-OL model (the results are not presented here to conserve on space). For instance, the differences in variable estimates (independent MNL-OL model and copula based MNL-OL model) are more than 20% in rear-ender for high wind and T-intersection; in rear-ended for high wind, presence of one passenger and two passenger; in far-angular for medium speed limit and high speed limit; in short-side angular for medium speed limit and high speed limit; and in head-on collision for weekend and morning peak-period.

In the following sections, the estimation results for injury severity component of the joint model are discussed by variable groups.

Driver Characteristics: The impacts of driver characteristics reveal significant variations based on driver age, gender, seat-belt use and driver knowledge of local conditions. The results indicate that the likelihood of being severely injured is lower for the young drivers compared to the adult drivers, particularly for rear-ended and short side-angular collisions, perhaps indicating the higher physiological strength of young drivers. Compared to the adult drivers, older drivers are more likely to sustain serious injury across a range of collision type, a result also observed in several previous studies (Bédard et al., 2002; Kim et al, 2013; Williams et al., 2003). Female drivers are consistently associated with higher injury risk propensity across all collision type presumably because of their lower physiological strength compared to their male counterparts. The negative impact of not using seat-belt is found significant only for near-angular collision type. The driver knowledge of local conditions characterized as local versus non-local drivers reveals that non-local drivers are likely to sustain serious injury for rear-ended, far-sideswipe and near-angular collisions. Driver unfamiliarity with the driving environment and road rules might contribute to such outcome.

Vehicle Characteristics: With respect to driver's vehicle type, the results indicate that drivers in station wagon are less likely to be severely injured compared to other drivers for seven of the eight collision types. The finding is consistent with the notion that heavier vehicles provide increased protection to drivers from severe injury. The positive effect of driving larger vehicles is significant in short side-angular and head-on collision for drivers of SUV and panel van. Consistent with several previous studies (Kim et al, 2013; Islam and Mannering, 2006) for most of the collision types, drivers in older vehicles (either vehicle age 6-10 or vehicle age 11 and above) have higher injury risk propensity compared to drivers in newer vehicles (vehicle age < 6 years); this can be attributed to the absence of advanced safety features in older vehicles.

Roadway Design Attributes: In terms of roadway design attributes, the estimates indicate that crashes on gravel road surface tend to be less severe compared to crashes on paved and unpaved surfaces for head-on collision. On gravel road surfaces, drivers are compelled to drive cautiously at a slower speed contributing to a reduction in the severity of crash outcomes. It is very interesting to note that the presence of signal decreases the injury propensity for rear-ender collision, and increases the injury propensity for both angular collisions (near- and far-angular). Injury propensity reductions are observed for the presence of pedestrian control (for rear-ended and short side-angular), roundabouts (for near- and short side-angular), stop sign (for short side-angular) and yield sign (for rear-ender and short side-angular).

The results for speed zones indicate that drivers are likely to sustain severe injuries for crashes occurring in zones with medium and higher speed limits highlighting that the probability of sustaining severe injuries increases with the increasing speed limits – a surrogate for vehicle speed at the time of crash. Among the type of intersection variables, T-intersection leads to higher injury propensity for rear-ender collisions and lower injury propensity for far-sideswipe collisions. Five or more legged intersection reflects reduced injury risk propensity for rear-ended collision. The reduction is also observed for non-intersection location in rear-ended and short side angular collision propensities.

Environmental Factors: In the category of environmental factors; time of day, weather condition and lighting condition have significant influence in moderating the driver injury severity across different collision types. With respect to the time of day, higher severity levels are associated with head-on collision during morning peak period. As expected, the injury severity for drivers is higher during late night. This is particularly so for rear-ender, short side-angular and head-on collision. The injury risk propensities of near-sideswipe and far-angular collision reflect higher severities for rainy/snowy/foggy weather. This may be due to unfavourable driving conditions resulting from the reduced visibility during adverse weather conditions. For high wind condition, rear-end collision propensities (rear-ender and rear-ended) indicate lower likelihood of severe injuries. The parameter characterizing the effect of weekend suggests lower injury severity level for head-on collision. The result is quite interesting and the reasons for the effect are not very clear. It is possibly a manifestation of unobserved information that is not considered in our analysis and warrants additional investigation in the future.

Crash Characteristics: Presence of passenger and trajectory of vehicle's motions are the crash characteristics that are found to affect driver injury severity. A higher injury risk propensity is observed for the presence of one passenger in the vehicle for the rear-ended and far-angular collision. However, the result associated with two passengers has a more uneven effect across different collision types indicating lower and higher likelihood of severe injury in the effect of rear-ender and rear-ended propensities, respectively. But presence of more than two passengers indicates lower likelihood of severe injury for rear-ender and short side-angular collision. Overall, the drivers with the presence of more passengers are less likely to be severely injured presumably a reflection of more responsible driving behavior in the presence of passengers (the same effect is observed in Eluru et al., 2010). Finally, the coefficients corresponding to the vehicle movement reveal that straight vehicle movement of the driver increases the injury risk propensity compared to other turning movements for far-sideswipe, near-, far- and short side-angular collisions. The result is expected because the drivers are likely to be travelling at a higher speed while travelling straight.

4.5 Elasticity Effects and Validation Analysis

The parameter estimates of Table 4.3 and 4.4 do not provide the magnitude of the effects of exogenous variables on the probability of involving in a specific type of collision or sustaining a specific injury severity category for drivers, respectively. For this purpose, we compute the aggregate level "elasticity effects" for all independent variables. The effects are computed for both the collision type and injury severity components and are presented in Table 4.5 and 4.6, respectively. However, to conserve on space, we present the elasticity effects only for the highest injury severity level (serious/fatal injury severity category) across all collision types.

The following observations can be made based on the results presented in Table 4.5 and 4.6. ***First***, the most significant variables in terms of collision type are: crashes at non-intersection location, crashes on gravel roads, presence of pedestrian control, driving a panel van, driver age less than 25, medium speed limit zone and not wearing seat-belt. ***Second***, the most significant variables in terms of increase in serious/fatal injury for drivers are crashes in high speed limit zone and driver age 65 and above. In terms of serious/fatal injury reduction, the important factors are driving a station wagon, presence of roundabout and presence of pedestrian control. ***Third***, the impacts, in magnitude, are substantially different in injury severity for several variables (driver age 65+, non-local driver, high speed limit road and collision during late-night) across different collision types. The effects are also different in direction (sign) for presence of signal and collision at T intersection. These differences clearly highlight that each collision type has a fundamentally distinct injury severity profile underscoring the importance of examining the effect of various exogenous variables on driver injury severity outcome by different collision types.

In an effort to further assess the performance of the joint model, a validation experiment is also carried out. For testing the predictive performance of the models, 50 data samples, of about 5,000 records each, are randomly generated from the hold out validation sample consisting of 25,769 records. For these samples, we present the average measures of predictive log-likelihood and BIC values along with the 95% level confidence band. The average predictive log-likelihood measure for the copula model and independent model are -13,277.24 [(-13326.17) — (-13228.30)] and -13280.37 [(-13329.306) — (-13231.438)], respectively. The BIC values for the copula model and independent model are 27714.26 [27615.130 — 27813.394] and 27720.13 [27621.79 — 27818.47], respectively, further highlighting the enhanced performance of the copula model.

4.6 Summary

The focus of the current chapter was to jointly model the collision type and injury severity outcome of drivers involved in a two vehicle collisions using a copula-based joint multinomial logit-ordered logit modeling framework. The closed form copula based framework was formulated to accommodate the impact of observed and unobserved effects on collision type and injury severity while also incorporating parameterization of dependency profile in an unordered and ordered joint structure. The proposed model was estimated using driver injury severity data for two vehicle crashes from the state of Victoria, Australia for the year 2006 through 2010 employing a comprehensive set of exogenous variables − driver characteristics, vehicle characteristics, roadway design attributes, environmental factors and crash characteristics.

The empirical analysis involved estimation of models by using six different copula structures: 1) Gaussian, 2) FGM, 3) Clayton, 4) Gumbel, 5) Frank and 6) Joe. The most suitable copula model was obtained for a combination model of Frank-Clayton copulas (Frank copula structure for rear-ender and head-on collision and Clayton dependency structure with the remaining collision type). Further, the comparison between copula and the independent models confirmed the importance of accommodating dependence between collision type and injury severity outcome in the analysis of driver injury severity. The model estimation results presented in the current chapter suggested that the impact of exogenous variables vary (for some variables) in magnitude as well as in sign across collision types. The variables in moderating the effect of different collision types also reveal varying effects.

In this research, to further understand the impact of various exogenous factors, elasticity effects were estimated for both the collision type and injury severity components. The elasticity effects clearly highlighted that each collision type has a fundamentally distinct injury severity profile underscoring the importance of examining the effect of various exogenous variables on driver injury severity outcome by different collision types. In summary, the findings of this chapter provided a more complete picture of injury severity profile associated with different collision type, thus target based countermeasures could be devised to address the entire profile of collision mechanism.

Table 4.1: Sample Characteristics of Collision Type and Injury Severity Level Sustained by Drivers

Injury Severity	Collision Type							
	Rear-ender	Rear-ended	Near-sideswipe	Far-sideswipe	Near-angular	Far-angular	Short-side angular	Head-on
No injury	612 (60.41%)*	422 (34.70%)	101 (47.64%)	183 (70.66%)	701 (45.91%)	803 (56.83%)	1211 (53.11%)	175 (29.71%)
Minor injury	261 (25.77%)	659 (54.19%)	72 (33.96%)	59 (22.78%)	526 (34.45%)	432 (30.57%)	718 (31.49%)	210 (35.65%)
Serious/Fatal injury	140 (13.82%)	135 (11.10%)	39 (18.40%)	17 (6.56%)	300 (19.65%)	178 (12.60%)	351 (15.39%)	204 (34.63%)
Total	1013	1216	212	259	1527	1413	2280	589

*The numbers in parenthesis correspond to column percentages

Table 4.2: Sample Characteristics of Explanatory Variables across Different Collision Types

Variables	Rear-ender	Rear-ended	Near-sideswipe	Far-sideswipe	Near-angular	Far-angular	Short-side angular	Head-on
Driver characteristics								
Driver age								
Age less than 25	306 (30.21)*	212 (17.43)	57 (26.89)	55 (21.24)	343 (22.46)	332 (23.50)	556 (24.39)	131(22.24)
Age 25 to 64	612 (60.81)	913 (75.09)	133(62.73)	190 (26.65)	952 (62.35)	885 (62.63)	1473 (64.6)	401 (68.08)
Age above 65+	91(8.98)	91 (7.48)	22 (10.38)	14 (5.41)	232 (15.19)	196 (13.87)	251 (11.01)	57 (9.68)
Driver gender								
Female	439 (43.34)	693 (56.99)	92 (43.40)	113 (43.63)	790 (51.74)	648 (45.86)	1049 (46.01)	205 (34.80)
Male	574 (56.66)	523 (43.01)	120 (56.60)	146 (56.37)	737 (48.26)	765 (54.14)	1231 (53.99)	384 (65.20)
Restraint system use								
Seat belt not used	38 (3.75)	32 (2.63)	9 (4.25)	11 (4.25)	35 (2.29)	31 (2.19)	61 (2.68)	22 (3.74)
Seat belt used	975 (96.25)	1184 (97.37)	203 (95.75)	248 (95.75)	1492 (97.71)	1382 (97.81)	2219 (97.32)	567 (96.26)
Locality of driver								
Non-local driver	119 (11.75)	147 (12.09)	33 (15.57)	36 (13.90)	145 (9.50)	121 (8.56)	199 (8.73)	109 (18.51)
Local Driver	894 (88.25)	1069 (87.91)	179 (84.43)	223 (86.10)	1382 (90.50)	1292 (91.44)	2081 (91.27)	480 (81.49)
Vehicle characteristics								
Vehicle Type								
Car	688 (67.92)	887 (72.94)	154 (72.64)	182 (70.27)	1099 (71.97)	1013 (71.69)	1684 (73.86)	378 (64.18)
Station wagon	177 (17.47)	219 (18.01)	34 (16.04)	50 (19.31)	285 (18.66)	248 (17.55)	395 (17.32)	118 (20.03)
Utility	108 (10.66)	85 (6.99)	17 (8.02)	21 (8.11)	108 (7.07)	118 (8.35)	159 (6.97)	80 (13.58)
Panel van	40 (3.95)	25 (2.06)	7 (3.30)	6 (2.32)	35 (2.29)	34 (2.41)	42 (1.84)	13 (2.21)
Vehicle age								
Vehicle age less than 6	282 (27.84)	404 (33.22)	75 (35.38)	88 (33.98)	496 (32.48)	439 (31.07)	636 (27.89)	172 (29.20)
Vehicle age 6-10	297 (29.32)	333 (27.38)	57 (26.89)	87 (33.59)	373 (24.43)	383 (27.11)	651 (28.55)	158 (26.83)
Vehicle age 11 and above	434 (42.84)	479 (39.39)	80 (37.74)	84 (32.43)	658 (43.09)	591 (41.83)	993 (43.55)	259 (43.97)
Roadway design attributes								
Type of road surface (Base: Paved)								
Paved	990 (97.73)	1189 (97.78)	206 (97.17)	254 (98.07)	1483 (97.12)	1385 (98.02)	2231 (97.85)	531 (90.15)
Unpaved	1 (0.10)	1 (0.08)	0 (0.00)	1 (0.39)	1 (0.07)	0 (0.00)	2 (0.09)	6 (1.02)
Gravel	22 (2.17)	26 (2.14)	6 (2.83)	4 (1.54)	43 (2.82)	28 (1.98)	47 (2.06)	52 (8.83)
Traffic Control Device								

No Control	622 (61.40)	730 (60.03)	177 (83.49)	211 (81.47)	587 (38.44)	638 (45.15)	957 (41.97)	566 (96.10)
Signal	237 (23.40)	304 (25.00)	23 (10.85)	34 (13.13)	319 (20.89)	437 (30.93)	764 (33.51)	4 (0.68)
Other traffic control	13 (1.28)	29 (2.38)	0 (0.00)	4 (1.54)	26 (1.70)	24 (1.70)	40 (1.75)	8 (1.36)
Pedestrian control	11 (1.09)	12 (0.99)	1 (0.47)	0 (0.00)	4 (0.26)	2 (0.14)	7 (0.31)	0 (0.00)
Roundabout	30 (2.96)	44 (3.62)	8 (3.77)	6 (2.32)	70 (4.58)	65 (4.60)	86 (3.77)	4 (0.68)
Stop sign	14 (1.38)	13 (1.07)	0 (0.00)	1 (0.39)	157 (10.28)	60 (4.25)	125 (5.48)	1 (0.17)
Yield sign	86 (8.49)	84 (6.91)	3 (1.42)	3 (1.16)	364 (23.84)	187 (13.23)	301 (13.20)	6 (1.02)
Speed zone								
Low speed (≤50 km/h)	118 (11.65)	141 (11.60)	30 (14.15)	42 (16.22)	326 (21.35)	298 (21.09)	461 (20.22)	74 (12.56)
Medium speed (60-90 km/h)	783 (77.30)	952 (78.29)	141 (66.51)	164 (63.32)	1057 (69.22)	1016 (71.90)	1664 (72.98)	314 (53.31)
High speed (≥100 km/h)	112 (11.06)	123 (10.12)	41 (19.34)	53 (20.46)	144 (9.43)	99 (7.01)	155 (6.80)	201 (34.13)
Type of intersection								
Cross intersection	282 (27.84)	293 (24.10)	34 (16.04)	43 (16.60)	654 (42.83)	701 (49.61)	1108 (48.60)	16 (2.72)
T intersection	263 (25.96)	361 (29.69)	52 (24.53)	66 (25.48)	569 (37.26)	471 (33.33)	832 (36.49)	69 (11.71)
Y intersection	5 (0.49)	4 (0.33)	0 (0.00)	0 (0.00)	6 (0.39)	6 (0.42)	10 (0.44)	2 (0.34)
Five and more legged intersection	28 (2.76)	43 (3.54)	2 (0.94)	9 (3.47)	39 (2.55)	43 (3.04)	78 (3.42)	1 (0.17)
Non-intersection	435 (42.94)	515 (42.35)	124 (58.49)	141 (54.44)	259 (16.96)	192 (13.59)	251 (11.01)	501 (85.06)
Environmental factors								
Time of day								
Morning peak	138 (13.62)	168 (13.82)	36 (16.98)	44 (16.99)	229 (15.00)	194 (13.73)	311 (13.64)	82 (13.92)
Off peak	358 (35.34)	440 (36.18)	69 (32.55)	97 (37.45)	514 (33.66)	489 (34.61)	774 (33.95)	183 (31.07)
Evening peak	291 (28.73)	352 (28.95)	54 (25.47)	52 (20.08)	419 (27.44)	368 (26.04)	588 (25.79)	148 (25.13)
Late evening	197 (19.45)	233 (19.16)	49 (23.11)	54 (20.85)	324 (21.22)	329 (23.28)	532 (23.33)	144 (24.45)
Late night	29 (2.86)	23 (1.89)	4 (1.89)	12 (4.63)	41 (2.69)	33 (2.34)	75 (3.29)	32 (5.43)
Weather condition								
Clear	863 (85.19)	1058 (87.01)	183 (86.32)	228 (88.03)	1343 (87.95)	1270 (89.88)	1981 (86.89)	446 (75.72)
Rainy/Snowy/Foggy	131 (12.93)	138 (11.35)	24 (11.32)	26 (10.04)	168 (11.00)	128 (9.06)	271 (11.89)	127 (21.56)
High wind	19 (1.88)	20 (1.64)	5 (2.36)	5 (1.93)	16 (1.05)	15 (1.06)	28 (1.23)	16 (2.72)
Surface condition								
Dry	837 (82.63)	1025 (84.29)	178 (83.96)	226 (87.26)	1297 (84.94)	1230 (87.05)	1906 (83.60)	403 (68.42)
Wet	170 (16.78)	187 (15.38)	34 (16.04)	33 (12.74)	226 (14.80)	182 (12.88)	371 (16.27)	175 (29.71)
Muddy	3 (0.30)	1 (0.08)	0 (0.00)	0 (0.00)	0 (0.00)	1 (0.07)	1 (0.04)	6 (1.02)
Snowy	3 (0.30)	3 (0.25)	0 (0.00)	0 (0.00)	4 (0.26)	0 (0.00)	2 (0.09)	5 (0.85)
Lighting condition								

Day	760 (75.02)	933 (76.73)	153 (72.17)	197 (76.06)	1139 (74.59)	1033 (73.11)	1617 (70.92)	412 (69.95)
Dusk/dawn	79 (7.80)	78 (6.41)	18 (8.49)	10 (3.86)	92 (6.02)	88 (6.23)	166 (7.28)	29 (4.92)
Dark-lighted	137 (13.52)	173 (14.23)	38 (17.92)	43 (16.60)	267 (17.49)	269 (19.04)	446 (19.56)	74 (12.56)
Dark-unlighted	30 (2.96)	27 (2.22)	3 (1.42)	8 (3.09)	27 (1.77)	16 (1.13)	43 (1.89)	72 (12.22)
Other lighting condition	7 (0.69)	5 (0.41)	0 (0.00)	1 (0.39)	2 (0.13)	7 (0.50)	8 (0.35)	2 (0.34)
Days of Week								
Weekend	221 (21.82)	222 (18.26)	52 (24.53)	80 (30.89)	351 (22.99)	359 (25.41)	548 (24.04)	198 (33.62)
Weekday	792 (78.18)	994 (81.74)	160 (75.47)	179 (69.11)	1176 (77.01)	1054 (74.59)	1732 (75.96)	391 (66.38)
Crash characteristics								
Trajectory of vehicle's motions								
Going straight	742 (73.25)	311 (25.58)	75 (35.38)	110 (42.47)	687 (44.99)	702 (49.68)	1655 (72.59)	418 (70.97)
Other movement	271 (26.75)	905 (74.42)	137 (64.62)	149 (57.53)	840 (55.01)	711 (50.32)	625 (27.41)	171 (29.03)
Presence of passenger								
No passenger	2 (0.20)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	1 (0.17)
One passenger	559 (55.18)	689 (56.66)	121 (57.08)	139 (53.67)	797 (52.19)	676 (47.84)	1081 (47.41)	258 (43.80)
Two passenger	256 (25.27)	305 (25.08)	53 (25.00)	64 (24.71)	363 (23.77)	393 (27.81)	655 (28.73)	157 (26.66)
More than two passengers	196 (19.35)	222 (18.26)	38 (17.92)	56 (21.62)	367 (24.03)	344 (24.35)	544 (23.86)	173 (29.37)

*The numbers in parenthesis correspond to column percentages within each category

Table 4.3: MNL (Collision Type) Model Estimates and Copula Parameters

Variables	Rear-ender	Rear-ended	Near-sideswipe	Far-sideswipe	Near-angular	Far-angular	Short-side angular	Head-on
Constant	—	-1.957(0.090)‡	-1.617(0.106)	-2.750(0.153)	-2.705(0.132)	-0.846(0.072)	-0.352(0.056)	-3.846(0.276)
Driver characteristics								
Driver age (Base: Age 25 to 64)								
Age less than 25	0.361(0.077)	-0.424(0.085)	—	—	—	—	—	—
Age above 65+	—	-0.416(0.125)	—	-0.711(0.286)	0.311(0.076)	0.311(0.076)	—	—
Driver gender (Base: male)								
Female	—	0.463(0.068)	—	—	0.239(0.060)	—	—	-0.264(0.106)
Restraint system use (Base: seat belt used)								
Seat belt not used	0.378(0.180)	—	—	—	—	—	—	—
Vehicle characteristics								
Vehicle Type (Base: Sedan)								
Utility	0.310(0.100)	—	—	—	—	—	—	0.310(0.100)
Panel van	0.609(0.184)	—	—	—	—	—	—	—
Vehicle age (Base: Vehicle age less than 6)								
Vehicle age 6-10	—	-0.181(0.064)	-0.341(0.074)	—	-0.341(0.074)	-0.181(0.064)	—	—
Vehicle age 11 and above	—	-0.169(0.076)	-0.391(0.151)	-0.474(0.142)	-0.140(0.058)	-0.140(0.058)	—	—
Roadway design attributes								
Type of road surface (Base: Paved)								
Gravel	—	—	—	—	—	—	—	1.361(0.250)
Traffic Control Device (Base: No traffic control and other control device)								
Signal	—	—	-0.7095(0.163)	-0.709(0.163)	—	—	—	-2.8209(0.547)
Pedestrian control	1.381(0.356)	1.381(0.356)	—	—	—	—	—	—
Roundabout	—	0.309(0.182)	—	—	0.494(0.148)	—	—	—
Stop sign	-0.894(0.282)	-1.261(0.301)	—	-2.406(1.017)	1.070(0.118)	—	—	-2.657(1.162)
Yield sign	—	0.314(0.126)	-2.094(0.426)	-2.094(0.426)	0.937(0.80)	—	—	-1.692(0.442)
Speed zone (Base: Low speed zone ≤50 km/h)								
Medium speed (60-90 km/h)	0.616(0.073)	0.616(0.073)	—	—	—	—	—	—
High speed (≥100 km/h)	—	0.575(0.116)	0.575(0.116)	0.671(0.186)	—	—	—	0.934(0.133)
Type of intersection (Base: Cross intersection)								
T intersection	0.217(0.092)	0.438(0.082)	0.4378(0.082)	0.634(0.182)	0.143(0.075)	-0.136(0.071)	—	1.384(0.296)

Five and more legged intersection	0.449(0.215)	0.708(0.181)	—	1.184(0.375)	—	—	—	—
Non-intersection	1.807(0.077)	1.807(0.077)	2.141(0.137)	2.141(0.137)	0.768(0.093)	—	—	3.864(0.277)
Environmental factors								
Time of day (Base: Morning peak, Off peak and Late evening)								
Late night	—	—	—	—	—	-0.3816(0.196)	—	—
Surface condition (Base: Dry)								
Wet	—	—	—	—	—	—	—	0.933(0.113)
Lighting condition (Base: Daylight)								
Dusk/dawn	—	—	—	-0.557(0.339)	—	—	—	—
Dark-lighted	-0.407(0.101)	-0.243(0.091)	—	—	—	—	—	—
Dark-unlighted	—	—	—	—	—	—	—	0.969(0.176)
Days of Week								
Weekend	—	-0.278(0.083)	—	0.406(0.145)	—	—	—	0.504(0.111)
Copula Parameters								
	Frank	Clayton	Clayton	Clayton	Clayton	Clayton	Clayton	Frank
Constant	3.047(1.667)	1.423(0.383)	0.495(0.625)	0.636(0.602)	0.772(0.511)	2.661(0.582)	1.473(0.414)	1.783(1.046)
Female Driver	0.971(0.540)	—	—	—	—	—	—	—
Medium Speed limit	—	—	—	—	1.482(0.228)	—	—	0.943(0.527)
Yield Sign	—	1.651(0.285)	—	—	—	—	—	—
Utility Vehicle	—	—	3.978(0.737)	—	—	—	—	—
Late night	—	—	—	5.079(0.904)	—	—	—	—
High wind	—	—	—	—	2.783(0.599)	—	2.318(0.497)	—

ǂStandard errors are presented in parenthesis

Table 4.4: OL (Injury Severity) Model Estimates

Variables	Rear-ender	Rear-ended	Near-sideswipe	Far-sideswipe	Near-angular	Far-angular	Short-side angular	Head-on
Threshold 1	1.970(0.406)ǂ	1.405(0.207)	1.405(0.207)	2.122(0.192)	1.405(0.207)	3.010(0.343)	2.122(0.192)	0.332(0.400)
Threshold 2	3.413(0.347)	4.021(0.163)	2.947(0.172)	4.021(0.163)	2.947(0.172)	4.593(0.294)	3.685(0.164)	1.904(0.319)
Driver characteristics								
Driver age (Base: Age 25 to 64)								
Age less than 25	—	-0.437(0.131)	—	—	—	—	-0.195(0.086)	—
Age above 65+	0.454(0.107)	—	1.182(0.385)	—	0.569(0.080)	0.454(0.107)	0.569(0.080)	—
Driver gender (Base: Male)								
Female	0.7714(0.046)	0.7714(0.046)	0.7714(0.046)	0.7714(0.046)	0.7714(0.046)	0.7714(0.046)	0.7714(0.046)	0.7714(0.046)
Restraint system use (Base: seat belt used)								
Seat belt not used	—	—	—	—	0.616(0.288)	—	—	—
Locality of driver (Base: Local driver)								
Non-local driver	—	0.272(0.103)	—	0.718(0.384)	0.272(0.103)	—	—	—
Vehicle characteristics								
Vehicle Type (Base: Sedan)								
Station wagon	-0.4827(0.071)	-0.237(0.079)	—	-1.100(0.508)	-0.483(0.071)	-0.237(0.079)	-0.483(0.071)	-0.237(0.079)
Utility	—	—	—	—	—	—	-0.690(0.178)	—
Panel van	—	—	—	—	—	—	—	-0.846(0.491)
Vehicle age (Base: Vehicle age less than 6)								
Vehicle age 6-10	0.214(0.059)	—	0.214(0.059)	—	—	0.214(0.059)	0.214(0.059)	—
Vehicle age 11 and above	0.297(0.047)	0.297(0.047)	0.297(0.047)	0.297(0.047)	0.297(0.047)	—	0.297(0.047)	—
Roadway design attributes								
Type of road surface (Base: Paved)								
Gravel	—	—	—	—	—	—	—	-0.558(0.332)
Traffic Control Device (Base: None traffic control and other control device)								
Signal	-0.392(0.148)	—	—	—	0.228(0.113)	0.572(0.101)	—	—
Pedestrian control	—	-0.969(0.440)	—	—	—	—	-0.969(0.440)	—
Roundabout	—	—	—	—	-1.227(0.188)	—	-1.227(0.188)	—
Stop sign	—	—	—	—	—	—	-0.409(0.165)	—
Yield sign	-0.942(0.275)	—	—	—	—	—	-0.317(0.113)	—
Speed zone (Base: Low speed zone ≤50 km/h)								

Medium speed (60-90 km/h)	—	—	—	—	—	0.343(0.117)	0.419(0.096)	—
High speed (≥100 km/h)	0.844(0.102)	—	0.844(0.102)	0.844(0.102)	0.844(0.102)	1.187(0.132)	1.187(0.132)	0.844(0.102)
Type of intersection								
T intersection	0.248(0.137)	—	—	-1.231(0.423)	—	—	—	—
Five or more legged intersection	-1.007(0.510)	—	—	—	—	—	—	—
Non-intersection	—	-0.254(0.079)	—	—	—	—	-0.254(0.079)	—
Environmental factors								
Time of day (Base: Morning peak, Off peak and Late evening)								
Morning peak	—	—	—	—	—	—	—	0.694(0.211)
Late night	1.202(0.231)	—	—	—	—	—	0.504(0.184)	1.202(0.231)
Weather condition (Base: Clear)								
Rainy/Snowy/Foggy	—	—	0.727(0.139)	—	—	0.727(0.139)	—	—
High wind	-0.807(0.339)	-0.807(0.339)	—	—	—	—	—	—
Lighting condition (Base: Daylight)								
Dusk/dawn	—	—	—	—	—	—	0.281(0.141)	0.621(0.359)
Dark-lighted	—	—	—	—	—	—	0.307(0.094)	—
Dark-unlighted	—	—	—	—	—	-1.005(0.518)	—	—
Days of Week								
Weekend	—	—	—	—	—	—	—	-0.442(0.172)
Crash Characteristics								
Presence of passenger (Base: No passenger)								
One passenger	—	0.876(0.137)	—	—	—	0.300(0.091)	—	—
Two passenger	-0.342(0.122)	0.421(0.144)	—	—	—	—	—	—
More than two passengers	-0.342(0.122)	—	—	—	—	—	-0.265(0.086)	—
Trajectory of vehicle's motions (Base: Other movement)								
Going Straight	—	—	—	0.505(0.063)	0.312(0.089)	0.505(0.063)	0.505(0.063)	—

ǂStandard errors are presented in parenthesis

Table 4.5: Elasticity Effects for Collision Type Component

Variables	Rear-ender	Rear-ended	Near-sideswipe	Far-sideswipe	Near-angular	Far-angular	Short-side angular	Head-on
Driver characteristics								
Driver age (Base: Age 25 to 64)								
Age less than 25	40.052	-36.538	0.811	0.888	0.653	0.667	0.731	0.124
Age above 65+	-2.373	-36.715	-0.452	-53.674	25.389	26.211	-7.004	2.607
Driver gender (Base: Male)								
Female	-9.030	36.468	-8.155	-7.794	13.883	-9.601	-9.606	-29.141
Restraint system use (Base: seat belt used)								
Seat belt not used	36.388	-5.688	-6.026	-6.029	-4.373	-4.443	-4.466	-6.411
Vehicle characteristics								
Vehicle Type (Base: Sedan)								
Utility	25.514	-6.913	-8.598	-8.698	-4.668	-4.571	-4.588	18.495
Panel van	63.287	-9.834	-10.378	-10.395	-7.638	-7.763	-7.805	-11.037
Vehicle age (Base: Vehicle age less than 6)								
Vehicle age 6-10	11.740	-6.122	-21.807	11.030	-19.612	-5.395	12.873	9.223
Vehicle age 11 and above	9.808	-6.815	-28.154	-35.500	-4.457	-4.685	9.346	9.712
Roadway design attributes								
Type of road surface (Base: Paved)								
Gravel	-15.978	-14.490	-21.988	-22.344	-7.502	-6.705	-6.714	137.276
Traffic Control Device (Base: None traffic control and other control device)								
Signal	17.724	15.759	-43.360	-43.319	8.339	7.798	7.796	-98.954
Pedestrian control	114.393	111.707	-46.430	-46.886	-37.835	-38.792	-39.048	-46.820
Roundabout	-13.374	17.187	-13.015	-12.459	37.679	-14.206	-14.138	-10.784
Stop sign	-54.852	-70.433	24.128	-92.304	157.516	-7.493	-7.133	-90.032
Yield sign	0.846	-28.630	-95.608	-95.711	110.263	-11.532	-11.315	-78.830
Speed zone (Base: Low speed ≤50 km/h)								
Medium speed (60-90 km/h)	39.964	39.521	-18.763	-18.811	-12.982	-13.048	-13.139	-19.165

High speed (≥100 km/h)	27.923	28.489	22.347	33.058	-19.898	-19.922	-20.066	47.918
Type of intersection (Base: Cross intersection)								
T intersection	-6.852	14.867	5.415	24.146	-2.274	-27.260	-14.652	96.935
Five and more legged intersection	18.932	51.882	-25.865	136.337	-17.955	-18.533	-18.744	-27.104
Non-intersection	76.774	74.714	103.364	96.499	-24.854	-83.150	-82.944	220.728
Environmental factors								
Time of day (Base: Morning peak, Off peak and Late evening)								
Late night	4.898	4.921	4.290	4.265	6.109	-27.427	6.525	2.695
Surface condition (Base: Dry)								
Wet	-9.591	-8.540	-13.453	-13.741	-4.305	-3.780	-3.765	80.177
Lighting condition (Base: Daylight)								
Dusk/dawn	1.660	1.652	2.153	-42.425	0.984	1.043	1.067	2.577
Dark-lighted	-28.355	-14.310	9.538	9.565	6.756	6.863	6.928	9.638
Dark-unlighted	-10.604	-9.515	-14.745	-15.028	-4.866	-4.309	-4.304	89.813
Days of Week (Base: Weekdays)								
Weekend	-2.347	-26.895	-4.597	39.239	0.062	0.337	0.357	41.697

Table 4.6: Elasticity Effects for Serious/Fatal Injury Severity Category

Variables	Rear-ender	Rear-ended	Near-sideswipe	Far-sideswipe	Near-angular	Far-angular	Short-side angular	Head-on
Driver characteristics								
Driver age (Base: Age 25 to 64)								
Age less than 25	—	-35.757	—	—	—	—	-16.532	—
Age above 65+	48.599	—	128.737	—	54.896	48.825	59.236	—
Driver gender (Base: Male)								
Female	72.080	65.571	64.029	71.814	63.669	72.314	68.000	58.566
Restraint system use (Base: seat belt used)								
Seat belt not used	—	—	—	—	63.399	—	—	—
Non-local driver	—	27.496	—	81.777	24.752	—	—	—
Vehicle characteristics								
Vehicle Type (Base: Sedan)								
Station wagon	-38.650	-20.505	—	-71.215	-36.310	-20.673	-37.388	-16.409
Utility	—	—	—	—	—	—	-47.841	—
Panel van	—	—	—	—	—	—	—	-48.777
Vehicle age (Base: Vehicle age less than 6)								
Vehicle age 6-10	20.437	—	18.066	—	—	20.804	19.551	—
Vehicle age 11 and above	27.753	28.177	24.822	28.376	25.345	—	26.529	—
Roadway design attributes								
Type of road surface (Base: Paved)								
Gravel	—	—	—	—	—	—	—	-35.466
Traffic Control Device (Base: None traffic control and other control device)								
Signal	-32.734	—	—	—	20.103	59.027	—	—
Pedestrian control	—	-60.601	—	—	—	—	-59.060	—
Roundabout	—	—	—	—	-68.876	—	-69.494	—
Stop sign	—	—	—	—	—	—	-31.320	—
Yield sign	-61.729	—	—	—	—	—	-25.443	—

Speed zone (Base: Low speed ≤50 km/h)								
Medium speed (60-90 km/h)	—	—	—	—	—	30.204	34.311	—
High speed (≥100 km/h)	—	—	80.545	89.029	89.666	174.732	153.362	64.888
Type of intersection (Base: Cross intersection)								
T intersection	23.907	—	—	-79.906	—	—	—	—
Five and more legged intersection	-62.106	—	—	—	—	—	—	—
Non-intersection	—	-22.965	—	—	—	—	-20.703	—
Environmental factors								
Time of day (Base: Morning peak, Off peak and Late evening)								
Morning peak	—	—	—	—	—	—	—	55.937
Late night	177.750	—	—	—	—	—	53.022	106.305
Weather condition (Base: Clear)								
Rain/snow/FOG/Smoke/Dust	—	—	70.554	—	—	87.743	—	—
High wind	-53.838	-53.830	—	—	—	—	—	—
Lighting condition (Base: Daylight)								
Dusk/dawn	—	—	—	—	—	—	27.162	50.379
Dark-lighted	—	—	—	—	—	—	28.979	—
Dark-unlighted	—	—	—	—	—	-62.290	—	—
Crash Characteristics								
Presence of passenger (Base: No passenger)								
One passenger	—	75.635	—	—	—	28.133	—	—
Two passenger	—	43.947	—	—	—	—	—	—
More than two passengers	—	—	—	—	—	—	-22.103	
Days of Week (Base: Weekdays)								
Weekend	—	—	—	—	—	—	—	-30.567
Trajectory of vehicle's motions (Base: Other movement)								
Going Straight	—	—	—	46.105	26.585	47.230	40.339	—

CHAPTER 5 Analyzing the Continuum of Fatal Crashes: A Generalized Ordered Approach

5.1 Introduction

A number of research efforts have examined the impact of exogenous characteristics (such as driver characteristics, vehicle characteristics, roadway design attributes, environmental factors and crash characteristics) associated with fatal crashes employing crash data with at least one fatality. These studies employed two broad dependent variable categorizations – (1) fatal/non-fatal or (2) fatal/serious injury. The binary categorization was analyzed employing descriptive analysis or logistic regression methods for identifying the critical factors affecting fatal crashes (for example see Zhang et al., 2013; Al-Ghamdi, 2002; Helai et al., 2008; Travis et al., 2012). Several studies have also investigated the factors affecting the involvement in a fatal crash as a function of individual characteristics. The important individual behavioral determinants of fatal crashes include excessive speed, violation of traffic rules and lack of seat belt use (Siskind et al., 2011; Valent et Al., 2002; Sivak et al., 2010; Viano et al., 2010). Other driver attributes such as aggressive driving behavior, unlicensed driving and distraction during driving are identified to be the most significant contributors of fatal crashes for young drivers (Lambert-Bélanger et al., 2012; Hanna et al., 2012, Preusser et al., 1998b; Chen et al., 2000; Williams, 1985). Studies have also examined the effect of race/ethnicity in fatal crashes (Braver, 2003; Romano et al., 2006; Campos-Outcalt et al., 2003; Harper et al., 2000). On the other hand, most critical factors identified from earlier research for older drivers in fatal crashes are frailty and reduced driving ability (Baker et al., 2003; Preusser et al., 1998a; Lyman et al., 2002, Thompson et al., 2013). Gates et al., (2013) investigate the influence of stimulants (such as amphetamine, methamphetamine and cocaine) on unsafe driving actions in fatal crashes. Stübig et al., (2012) investigate the effect of alcohol consumption on preclinical mortality of traffic crash victims (see also Fabbri et al., 2002; Tulloh and Collopy, 1994).

Many of the earlier studies also focused on the vehicular characteristics of fatal crashes (Evans and Frick, 1994; Fredette et al., 2008) and demonstrated that the relative risk of fatality is much higher for the driver of lighter vehicle (sedan, compact car) compared to those in the heavier vehicle (SUV, Vans, Pickups). Among the environmental factors, it was found that collision during night time (Arditi et al., 2007) has the most significant negative impact on fatality risk in a crash.

In terms of crash characteristics, head-on crash and crashes on high speed limit road locations increased the probability of fatalities in a crash (Fredette et al., 2008; Shibata and Fukuda, 1994; Bédard et al., 2002).

These studies offer many useful insights on what factors affect crash related fatality, particularly in the context of fatal vs. non-fatal injury categorization. However, there is one aspect of fatal crashes that has received scarce attention in the traditional safety analysis. The studies that dichotomize crashes into fatal versus non-fatal groups assume that all fatal crashes are similar. Keeping all else same, a fatal crash that results in an immediate fatality is clearly much more severe than another crash that leads to fatality after several days. To address this issue, the objective of the current chapter is to identify the associated risk factors of driver fatalities while recognizing that fatality is not a single state but rather is made up of a timeline between dying instantly to dying within thirty days of crash.

The data for the current study is sourced from the Fatality Analysis Reporting System (FARS) database for the year 2010. FARS database compiles crashes if at least one person involved in the crash dies within thirty consecutive days from the time of crash. Further, FARS database reports the exact timeline of the fatal occurrence within thirty days from the time to crash. The detailed information available in FARS provides us a continuous timeline of the fatal occurrences from the time of crash to death. This allows for an analysis of the survival time of victims before their death. This chapter builds on existing fatality analysis research by developing disaggregate level model for the discrete representation of the continuous fatality timeline using the FARS dataset. The fatality timeline information obtained through FARS is categorized as an ordered variable ranging from death in thirty days to instantaneous death in seven categories as follows: died between 6th-30 days of crash, died between 2nd-5 days of crash, died between 7th-24 hours of crash, died between 1st-6 hours of crash, died between 31st-60 minutes of crash, died between 1st-30 minutes of crash and died instantly. We employ the mixed generalized ordered logit (MGOL) framework to examine driver fatalities characterized as an ordinal discrete variable of an underlying severity continuum of fatal injuries.

In modeling the discretized fatality timeline, the EMS response time variable is an important determinant. However, it is possible that the EMS response time and fatality timeline are influenced by the same set of observed and unobserved factors, generating endogeneity in the outcome model of interest. In fact, it was identified that EMS response time are affected by several external environmental and regional factors (Brodsky, 1992; Meng and Weng, 2013). Such correlations impose challenges in using the EMS response variable as an explanatory variable in examining fatality outcome of crashes. For example, consider two potential crash scenarios. In scenario 1 a relatively major crash occurs and in scenario 2 a minor crash occurs. When the information of a crash is provided the urgency with which the EMS teams are deployed for the first scenario is likely to be higher than the urgency for the second scenario. So, we potentially have a case where EMS time for arrival is lower for scenario 1 but potentially the consequences of the crash for scenario 1 are much severe i.e. survival time is much smaller. So, in a traditional modeling approach one would conclude that lower EMS arrival times are associated with smaller survival times. This is a classic case of data endogeneity affecting the modeling results. Hence, it is necessary to account for this endogeneity in the modeling process. In this chapter, we propose to apply an econometric approach to accommodate for this. Specifically, we propose to estimate a driver-level fatal injury severity model while also accounting for endogeneity bias of EMS arrival time using ordered outcome modeling framework with endogeneity treatment. In doing so, the correction for endogeneity bias is pinned down in the ordered outcome models by employing a two-stage residual inclusion (2SRI) approach.

In summary, the current chapter makes a three-fold contribution to the literature on vehicle-occupant injury severity analysis. ***First***, our study is the first attempt to analyze the fatal injury from a new perspective and examine fatality as a continuous spectrum based on survival time ranging from dying within thirty days of crash to dying instantly. ***Second***, we propose and estimate a two equation model that comprises of regression for EMS response time and MGOL with residuals from the EMS model to correct for endogeneity bias on the effect of exogenous factors on the timeline of death. ***Finally***, we compute elasticity measures to identify important factors affecting survival time after motor vehicle crash.

The rest of the chapter is organized as follows. Section 5.2 provides details of the econometric model framework used in the analysis. In Section 5.3, the data source and sample formation procedures are described. The model estimation results and elasticity effects are presented in Section 5.4 and 5.5, respectively. Section 5.6 summarize the major findings of the chapter.

5.2 MODEL FRAMEWORK

The focus of this chapter is on examining the driver-level fatal injury at a disaggregate level while also accounting for endogeneity bias of EMS arrival time by using a MGOL model framework with endogeneity treatment. In doing so, the correction for endogeneity bias is pinned down in MGOL model by employing a 2SRI approach[9] (as opposed to the two-stage predictor substitution approach). The framework used for MGOL model with endogenous treatment consists of a two-stage procedure. In the first stage, the residuals are computed from the linear regression estimates of the endogenous variable (EMS arrival time). In the second stage, MGOL model is estimated by including the first-stage residuals as additional regressor along with the endogenous variable in examining the outcome of interest. In this section, econometric formulation for MGOL model with the 2SRI treatment is presented.

5.2.1 First Stage

Let i $(i = 1,2, \dots \dots, I)$ and j $(j = 1,2, \dots \dots, J)$ be the indices to represent driver and the time between crash occurrence and time of death for each fatally injured driver i. In this chapter, index j takes the values of: died between 6th to 30 days of crash $(j = 1)$, died between 2nd to 5 days of crash $(j = 2)$, died between 7th to 24 hours of crash $(j = 3)$, died between 2nd to 6 hours of crash $(j = 4)$, died between 31st to 60 minutes of crash $(j = 5)$, died between 1st to 30 minutes of crash $(j = 6)$ and died instantly $(j = 7)$ for all fatally injured drivers. Let us also assume that y_i represents the discrete levels of time to death, $\boldsymbol{x}_\mathrm{i}$ is a column vector of observable exogenous variables, $\boldsymbol{u}_i$ is a set of e $(e = 1,2, \dots \dots, E)$ endogenous variables and $\boldsymbol{q}_i$ is a $1 \times E$ set of unobservable endogenous variables possibly correlated with both the outcome and the endogenous variables, generating endogeneity bias in the outcome model. In our analysis, we hypothesize that EMS arrival time may be correlated with the unobservable determinants of fatal injury severity of drivers, thus we have $e = 1$ in the current study context. Following Terza et al. (2008), we present the endogeneity of $\boldsymbol{u}_i$ by assuming an idiosyncratic influence of the same latent variables $\boldsymbol{q}_i$ on both the outcome and endogenous variables as a linear regression model as:

$$L_i = \rho \boldsymbol{w}_i + \boldsymbol{q}_i \tag{5.1}$$

[9] The reader is referred to Terza et al. (2008) for a detailed discussion of why the two stage residual inclusion method provides consistent estimates in non-linear models, while the two stage predictor substitution method does not.

where,

$\boldsymbol{w}_i = [\boldsymbol{x}_i\ \boldsymbol{v}_i]$ and $\boldsymbol{v}_i$ is a set of at least E instrumental variables

ρ is a corresponding row vector of parameter estimates

The residuals of endogenous variables can be computed as:

$$q_i^R = \boldsymbol{u}_i - Pr(\boldsymbol{u}_i|\boldsymbol{w}_i) \tag{5.2}$$

where, $Pr(\boldsymbol{u}_i|\boldsymbol{w}_i)$ is the predictor of $\boldsymbol{u}_i$.

5.2.2 Second Stage

In the proposed two-stage model, the modeling of discrete levels of fatal crashes is undertaken using MGOL specification. The MGOL accommodates unobserved heterogeneity in the effect of exogenous variable on injury severity levels in both the latent injury risk propensity function and the threshold functions (Srinivasan, 2002, Eluru et al., 2008). In the MGOL model, the discrete levels of time to death (y_i) are assumed to be a mapping (or partitioning) of an underlying continuous latent variable (y_i^*) as follows:

$$y_i^* = (\beta + \alpha_i)\boldsymbol{x}_i + \sigma\boldsymbol{u}_i + \lambda\boldsymbol{q}_i + \varepsilon_i\ ,\ \ y_i = j, if\ \tau_{i,j-1} < \ y_i^* < \tau_{i,j} \tag{5.3}$$

where,

β, σ and λ are corresponding row vectors of associated parameters for $\boldsymbol{x}_i$, $\boldsymbol{u}_i$ and $\boldsymbol{q}_i$, respectively.

α_i is a row vector representing the unobserved factors specific to driver i and his/her trip environments

ε_i is a random disturbance term assumed to be standard logistic

$\tau_{i,j}$ represents the thresholds

Once the linear regression for the endogenous variable is estimated, we can insert the computed residuals of equation 5.2 as additional regressors in equation 5.3 for the outcome of interest. Thus, substituting the residuals for the unobservable latent factors, we can re-write equation 5.3 as:

$$y_i^* = (\beta + \alpha_i)\boldsymbol{x}_i + \sigma\boldsymbol{u}_i + \lambda q_i^R + \varepsilon_i\ ,\ \ y_i = j, if\ \tau_{i,j-1} < \ y_i^* < \tau_{i,j} \tag{5.4}$$

In the above setting, the endogeneity of $\boldsymbol{u}_i$ will be absent if λ turns out to be zero. Moreover, in equation 6.4, $\tau_{i,j}$ $(\tau_{i,0} = -\infty, \tau_{i,J} = \infty)$ represents the upper threshold associated with driver i and time scale j, with the following ordering conditions: $\left(-\infty < \tau_{i,1} < \tau_{i,2} < \ldots\ldots\ldots < \tau_{i,J-1} < +\infty\right)$. To maintain the ordering conditions and allow the thresholds to vary across drivers, Eluru et al., (2008) propose the following non-linear parameterization of the thresholds as a function of exogenous variables:

$$\tau_{i,j} = \tau_{i,j-1} + exp[(\delta_j + \gamma_{i,j})\boldsymbol{z}_{i,j}] \tag{5.5}$$

where, $\boldsymbol{z}_{ij}$ is a set of exogenous variable associated with j th threshold; δ_j is a time to death-specific row vector of parameters to be estimated (we need to restrict δ_1 to be a row vector of zero values for identification reason) and γ_{ij} is another row vector representing the unobserved factors specific to driver i and his/her trip environments. The traditional OL model assumes that the thresholds $\tau_{i,j}$ remain fixed across drivers ($\tau_{i,j} = \tau_j \ \forall \ i$); that is, it assumes that δ_j has all zero elements for all j values (except for the constant). Thus, the model will collapse to a simple OL model if α_i turns out to be zero in equation 5.4 and $\tau_{i,j}$ remain fixed across driver in equation 5. On the other hand, if α_i and $\gamma_{i,j}$ terms of equation 5.4 and 5.5 are found to be zero in model estimation, then the model will collapse to simple GOL model.

In equations 5.4 and 5.5, we assume that α_i and γ_{ij} are independent realizations from normal distribution for this study. Thus, conditional on α_i and γ_{ij}, the probability expression for individual i and alternative j in MGOL model with the 2SRI treatment take the following form:

$$\begin{aligned}\pi_{ij} &= Pr(y_i = j | \alpha_i, \gamma_{ij}) \\ &= \Lambda[\tau_{i,j-1} + \exp((\delta_j + \gamma_{i,j})\, \boldsymbol{z}_{i,j}) - \{(\beta + \alpha_i)\boldsymbol{x}_i + \sigma\boldsymbol{u}_i + \lambda q_i^R\}] - \\ &\quad \Lambda[\tau_{i,j-2} + \exp((\delta_{j-1} + \gamma_{i,j-1})\, \boldsymbol{z}_{i,j}) - \{(\beta + \alpha_i)\boldsymbol{x}_i + \sigma\boldsymbol{u}_i + \lambda q_i^R\}]\end{aligned} \tag{5.6}$$

The unconditional probability can subsequently be obtained as:

$$P_{ij} = \int_{\alpha_i, \gamma_{ij}} [Pr(y_i = j | \alpha_i, \gamma_{ij})] * dF(\alpha_i, \gamma_{ij}) d(\alpha_i, \gamma_{ij}) \tag{5.7}$$

The parameters to be estimated in the MGOL model with the 2SRI treatment are: the parameters corresponding to the linear regression (ρ), the parameters corresponding to the propensity ($\beta, \sigma, \lambda\ and\ \alpha_i$) and the parameters corresponding to thresholds ($\delta_j\ and\ \gamma_{i,j}$). In this study, we use a quasi-Monte Carlo (QMC) method proposed by Bhat (2001) for discrete outcome model to draw realization from its population multivariate distribution. Within the broad framework of QMC sequences, we specifically use the Halton sequence (4,000 Halton draws) in the current analysis (see Eluru et al., 2008 for a similar estimation process).

5.3 DATA

5.3.1 Data Source

The data for the current chapter is sourced from the FARS database for the year 2010. FARS data is a census of all fatal crashes in the US and compiles crashes if at least one person involved in the crash dies within thirty consecutive days from the time of crash. The FARS database has a record of 30,196 fatal crashes with 32,885 numbers of fatalities for the year 2010. This database is obtained from the US Department of Transportation, National Highway Traffic Safety Administration's National Center for Statistics and Analysis (ftp://ftp.nhtsa.dot.gov). The FARS dataset provides a continuous timeline of the fatal occurrences from the time to crash until thirty days. It also provides information on a multitude of factors (driver characteristics, vehicle characteristics, roadway design and operational attributes, environmental factors, crash characteristics and situational variables) representing the crash situation and events.

5.3.2 Sample Formation and Description

This study is focused on fatality outcome of passenger vehicles' drivers who were involved in either a single or two vehicle crashes. The crashes that involve more than two vehicles are excluded from the dataset. Commercial vehicles involved collisions are also excluded to avoid the potential systematic differences between commercial and non-commercial driver groups. From the dataset, only the drivers who were fatally injured are considered for the current study. The final FARS dataset, after removing records with missing information for essential attributes consisted of about 5,102 driver records. The continuous timeline (computed as the difference between declared death time and crash time) provided in FARS was then discretized as a seven point discrete ordinal variable to represent the scale of fatal injury severity of drivers involved in these crashes - from least severe to most severe fatal crashes as follows: 1) Died between 6th to 30 days of crash, 2) Died between 2nd to 5 days of crash, 3) Died between 7th to 24 hours of crash, 4) Died between

2nd to 6 hours of crash, 5) Died between 31st to 60 minutes of crash, 6) Died between 1st to 30 minutes of crash and 7) Died instantly. The distributions of driver fatalities over the fatality scale in our final estimation sample are presented in Table 5.1. We adopted a seven alternative discrete spectrum for our analysis based on observed frequencies and time to death groupings of policy interest. It is important to note that, within an ordered outcome structure, it would be relatively easy to incorporate a larger number of alternative categories, if needed, while still retaining a parsimonious specification. From Table 5.1 we can see that more that 60% drivers died within one hour of crash and almost one third of these crash victims are reported to die instantly. Also, only 5.9% of the drivers can evade mortality more than five days of crashes.

Table 5.2 offers a summary of the sample characteristics of the exogenous factors in the estimation dataset. From the descriptive analysis, we observe that a large portion of crashes occur on high speed limit road (54.5%), on rural road (62.8 %), during dry weather condition (70.6%) and at non-intersection location (75.9%). The majority of drivers are aged between 25 and 64 (57.1%). In addition to the variables describing the crash situation and events presented in Table 5.2, FARS database also provides information on crash notification time, EMS response time and time of EMS arrival at hospital. From this information, it is possible to compute EMS response time (as the difference between EMS arrival time at the crash scene and crash time) and hospital arrival time (as the difference between EMS arrival at hospital and EMS arrival at crash scene). However, EMS arrival time at hospital is available only for the crash victim who arrived first at hospital among all other crash victims (if present) for that specific crash. Therefore, hospital arrival time is not available for all fatal records of driver, and, hence is not considered in our final estimation sample. On the other hand, the sample we use in the current study provides information about the EMS response time. From the descriptive statistics of this variable we observe that EMS response time exceeds one hour – most popularly referred to as the "golden hour" – only for 3.1% of records. The median EMS response time is about 11 minutes, with a range of 0 minute to approximately 9.5 hours.

5.4 Empirical Analysis

5.4.1 <u>Variables Considered</u>

In our analysis, we selected a host of variables from six broad categories: <u>driver characteristics</u> (including driver age, alcohol consumption and previous driving conviction records), <u>vehicle</u>

characteristics (including vehicle age), roadway design and operational attributes (including speed limit, traffic control device, roadway functional class and land use), environmental factors (including time of day, lighting condition and weather condition), crash characteristics (including manner of collision and collision location) and situational variable (including driver ejection, number of passengers and EMS response time). The final specification of the model development was based on combining the variables when their effects were not statistically different and by removing the statistically insignificant variables in a systematic process based on statistical significance (95% confidence level). For continuous variables, linear, polynomial and spline forms were tested.

5.4.2 Model Specification and Overall Measures of Fit

In the research effort, initially we estimated three different models: 1) OL, 2) GOL and 3) MGOL, by considering EMS response time as an explanatory variable in our empirical analysis. In our initial specifications of all the three aforementioned models we obtained a counterintuitive result for EMS response time indicating that the likelihood of early death decreases with an increase in EMS response time. Therefore, to further explore the effect of this indicator variable, several specifications (log transformation, dummy categories) of EMS response time have been explored in OL, GOL and MGOL frameworks. However, for all the aforesaid specifications, we observe that a longer EMS response time has negative impact on the survival probability of drivers in the current study context. The result could be a manifestation of endogeneity between crash seriousness and EMS response time i.e. severe crashes are likely to have shorter EMS times while less severe crashes are likely to have longer EMS times. So, in such scenarios the early arrival of EMS coincides with early death causing a non-intuitive parameter estimate. Thus, to control for the endogeneity of EMS response time with fatal crash outcomes, we include a residual variable through 2SRI method in examining the fatality spectrum. To that extent, we have further estimated the following three ordered outcome models with endogenous treatment: 1) OL with the 2SRI treatment, 2) GOL with the 2SRI treatment and 3) MGOL model with the 2SRI treatment. After controlling for the endogeneity, the coefficient on the logarithm of EMS response time is found out to be positive in all three model specifications indicating that the likelihood of early death increases with an increase in EMS response time.

Prior to discussing the estimation results, we compare the performance of these models in this section. At first, the exogeneity of regressors $\boldsymbol{u}_i$ in equation 5.4 is tested for $\lambda = 0$ by using likelihood ratio (LR) test within each set of models. The equation for LR test statistic is presented in equation 2.17. The computed value of the LR test is compared with the χ^2 value for the corresponding degrees of freedom. These estimates are presented in Table 5.3. From the first three rows of LR test values in table 5.3 we can see that all three models with 2SRI treatment outperform the corresponding models without 2SRI treatments at any significance level. The LR test comparisons confirm the importance of accommodating endogoneity between EMS response time and fatal injury outcome in the analysis of driver fatalities. Further, we also compare the estimated ordered models with 2SRI treatments by using LR test for selecting the preferred model among those. The results are presented in last three rows of Table 5.3. The LR test values indicate that MGOL model with 2SRI treatment outperforms the OL model with 2SRI treatment at any level of statistical significance. The MGOL model with 2SRI treatment outperforms the GOL model with 2SRI treatment at the 0.05 significance level. The comparison exercise clearly highlights the superiority of the MGOL model with 2SRI treatment in terms of data fit compared to all the other ordered models.

5.4.3 <u>Estimation Results</u>

In presenting the effects of exogenous variables in the model specification, we will restrict ourselves to the discussion of the MGOL model with 2SRI treatment. Table 5.4 presents the estimation results. To reiterate, the dependent variable under consideration is the 7 point ordinal variable defined as: died between 6th-30 days of crash, died between 2nd-5 days of crash, died between 7th-24 hours of crash, died between 1st-6 hours of crash, died between 31st-60 minutes of crash, died between 1st-30 minutes of crash and died instantly. Estimation results of Table 5.4 has six different columns. The first column corresponds to the propensity and represents the estimates of the parameters of equation 5.4. From second to sixth columns of estimation results in Table 5.4 corresponds to the thresholds and represent parameters of equation 5.5. In MGOL model, when the threshold parameter is positive (negative) the result implies that the threshold is bound to increase (decrease); the actual effect on the probability is quite non-linear and can only be judged in conjunction with the influence of the variable on propensity and other thresholds. In the following sections, the estimation results are discussed by variable groups.

Driver Characteristics: The effect of driver age is found to have significant impact on the length of hospital stay before death. The parameter characterizing the effect of young driver (age 24 & less) suggests that the likelihood of dying earlier is lower for young driver compared to middle-aged (age 25-64) driver. The negative sign of latent propensity associated with old driver (age 65 & above) suggests that the likelihood of dying earlier is lower for older driver compared to middle aged driver. On the other hand, the impacts of old driver on both of the fourth and fifth thresholds are negative. The results suggest an increased probability of dying within 6th-30 days of crash and, also in general, a decreased possibility of instant death, presumably due to the declined wound healing and immune competence of drivers with advancing age after surviving the early phase of trauma (Tohira et al., 2012).

As expected, MGOL model estimates related to alcohol impairment indicate a higher likelihood of early mortality risk of alcohol impaired drivers compared to the sober drivers. At the same time, the positive values of the second threshold of alcohol impaired driver reflects an increase in the probability of dying within 2nd-5 days of crash. Intoxicated drivers are identified to be less immune to post traumatic response and suffer from more severe abdominal injuries (Zeckey et al., 2011; Stübig et al., 2012). Furthermore, higher impact speed differential due to the risk taking disposition of alcohol intoxicated driver presumably reduces the time to death of this group of drivers (Soderstrom et al., 2001).

Previous driving records also have significant influence on time to death after crash. The results associated with previous recorded suspension and revocation of driving licence indicates that an increase in number of previous recorded suspension and revocation deceases the likelihood of early mortality. The result is perhaps indicating more cautious driving of this group of driver to avoid any further conviction while driving. Also, the result indicates that drivers are less likely to evade early mortality with an increasing record of other previous record of harmful motor vehicle convictions (other than previous recorded suspension and revocation of driving licence, previous recorded crashes, previous drinking convictions and previous speeding convictions). However, the effect of other previous record of harmful motor vehicle convictions variable results in an estimate that is normally distributed with mean 0.104 and standard deviation of 0.208 implying that almost 71% of the drivers with higher records of earlier harmful motor vehicle convictions involved in the collision sustain early death.

Vehicle Characteristics: Among different vehicle characteristics explored in this study, only vehicle age is significant in the final model specification. Vehicle age result does not have any effect on the propensity of time to death after crash, but demonstrates a higher likelihood of death within 1st-30 minutes of crash for the driver of old vehicles (vehicle age≥11 years) and in general, a higher probability of instant death in a crash. The result highlights the advantages of newer vehicle fleet – presence of advanced safety technologies (electronic stability control, improvement in air bag design, crash cage, energy-absorbing steering columns, crash-resistant door locks and high-penetration-resistant windshields) and designs of newer vehicle with improved crash worthiness (O'Neill, 2009; Ryb et al., 2011).

Roadway Design and Operational Attributes: The results for speed limit indicate that the propensities to die earlier are higher for crashes occurring on roads with medium or higher speed limit roads relative to crashes on lower speed limit roads. As is expected, within the two speed categories considered, the higher speed category has a larger impact relative to the medium speed category, which underscores the fact that the probability of early mortality risk increases with the increasing speed limits of roadways. MGOL model estimates for higher speed limit results in a parameter that is normally distributed with a mean 0.359 and standard deviation 0.447, which indicates that almost 78% of the drivers cannot evade early death for the crashes occurring on higher speed limit roads. Higher speed, representing average driving speed, significantly increases the kinetic energy of crashes (Elvik, 2004; Sobhani et al., 2011) resulting in medical complications with multiple injuries and traumatic brain injury to the victims (Weninger and Hertz, 2007). Further, the cabin intrusion caused by high mechanical force of such crash might also increase the extrication time of victims from the damaged vehicle (Weninger and Hertz, 2007). Crashes at stop-sign controlled or other traffic controlled (such as warning sign, regulatory sign, railway crossing sign) intersections seem to increase the likelihood of early death relative to crashes at other locations, possibly suggesting non-compliance with these traffic control devices and judgment problems (Chipman, 2004; Retting et al., 2003)..

Environmental Factors: With respect to time of day, the latent propensities for off peak and evening peak periods (related to morning peak and nigh-time) are found negative, indicating lower likelihood of early mortality, may be a result of traffic congestion and slow driving speeds during these periods. At the same time, the effect of off peak period on the threshold indicates a lower probability of dying between 1st-30 minutes after crash. The weather condition effects simplified to a simple binary representation of cloudy condition. The result indicates that if collisions occur during cloudy weather (relative to those during other weather conditions) the drivers are less likely to evade early death, perhaps because of the reduced visibility, which presumably results in reduced perception-reaction and reduced ability to take evasive actions at the crash incident (Tay et al., 2011). The effect of cloudy weather condition on the threshold also indicates increased likelihood of death between 2nd-5 days of crash.

<u>Crash Characteristics</u>: With respect to manner of collision, the time to death propensity is observed to be lower for front-to-rear collision relative to other manners of collision. The results associated with a head-on collision reflect a higher probability of death between 1st-6 hours of crash and in general indicate the anticipated increased likelihood of early death. Head-on collisions are often caused by drivers violating traffic rules, crossing the centerline by mistake and losing control of their vehicles (Zhang and Ivan, 2005). The pre-impact speed vectors of motor vehicles are directed in opposing directions during a head-on collision, resulting in greater dissipation of kinetic energy and heavier deformation of motor vehicle bodies (Prentkovskis et al., 2010), resulting in higher risk of injury.

As observed in several previous studies (Al-Ghamdi, 2002), the results related to crash location of our study reflect an increased injury risk propensity for collision at non-intersection location (related to crashes at intersection and other locations). However, the effects of "non-intersection location" indicator in threshold parameterization are relatively complex. It has a positive impact on the threshold between 1st-6 hours and 31st-60 minutes crash outcome categories; while it has a negative impact on the threshold between 31st-60 minutes and 1st-30 minutes categories. In general, the net implication is that collision at non-intersection location has a higher probability of sustaining early death (the specific impact of other fatal crash categories on driver fatalities are context-specific).

<u>Situational Variables</u>: As identified in several previous studies (Palanca et al., 2003), the result related to driver ejection indicate an increased early death propensity. Number of passenger in

vehicle at the time of collision is also found to have significant impact on the time to death of driver. The results related to presence of more passengers reflect an increased early death propensity, perhaps indicating inattentiveness to the driving task due to distraction caused by in vehicle interactions among occupants.

The last two rows of estimation results in Table 5.4 represent the associated results of: (1) the logarithm of EMS response time and (2) the residual obtained from regressing the logarithm of EMS response time variable on morning peak, late-night, dark-not lighted, rain, snowy, rural, principle arterial and minor arterial indicator variables[10]. The role of the residual variable is to control for the endogeneity of the EMS response time variable in examining the time to death. From Table 5.4, we can see that after controlling for endogeneity, the coefficient on the logarithm of EMS response time is positive and statistically significant indicating that EMS response time has the expected impact on severity once we control for the endogeneity bias. Specifically, as can be observed from the coefficient of the residual term, the non-intuitive impact of EMS time was a result of the correlation between EMS time and unobserved determinants. Through our approach, by accounting for the endogeneity we were able to differentiate between the observed impact of EMS time and the spurious effect due to the unobserved factors.

5.5 Elasticity Effects

The parameter effects of the exogenous variables in Table 5.4 do not provide the magnitude of the effects on time to death of drivers. For this purpose, we compute the aggregate level "elasticity effects" for all categories of independent variable (see Eluru and Bhat, (2007) for a discussion on the methodology for computing elasticities) and present the computed elasticities in Table 5.6. The effects are computed for all categories of fatal crashes. The results in the table can be interpreted as the percentage change (increase for positive sign and decrease for negative sign) in the probability of the fatal severity categories due to the change in that specific exogenous variable.

The following observations can be made based on the elasticity effects of the variables presented in Table 5.6. <u>First</u>, the results in Table 5.6 indicate that there are considerable differences in the elasticity effects across different fatal crash categories, suggesting that fatality is not a single state but rather is made up of multiple discrete states from dying instantly to dying within thirty

[10] The estimation results for the linear regression model are presented in Table 5.5.

days of crash. Second, the most significant variables in terms of lower survival probability for drivers are crashes on high speed limit roads, crashes on medium speed limit roads and head-on crashes. Third, in terms of longer survival probability, the important factors are old driver, front-to-rear crash and crashes during off peak period. Fourth, elasticity estimates of EMS response time in Table 5.6 emphasize the importance of early EMS response. Finally, the elasticity analysis assists in providing a clear picture of attribute impact on driver time-to-death variables. The elasticity analysis conducted provides an illustration of how the proposed model can be applied to determine the critical factors contributing to reducing the survival time.

5.6 Summary

The focus of this chapter was to identify the associated risk factors of driver fatalities while recognizing that fatality is not a single state but rather is made up of a timeline between dying instantly to dying within thirty days by using Fatality Analysis Reporting System (FARS) database for the year 2010. In the US, safety researchers have focused on examining fatal crashes (involving at least one fatally injured vehicle occupant) by using FARS dataset. FARS database compiles crashes if at least one person involved in the crash dies within thirty consecutive days from the time of crash along with the exact timeline of the fatal occurrence. Previous studies using FARS dataset offer many useful insights on what factors affect crash related fatality, particularly in the context of fatal vs. non-fatal injury categorization. However, there is one aspect of fatal crashes that has received scarce attention in the traditional safety analysis. The studies that dichotomize crashes into fatal versus non-fatal groups assume that all fatal crashes in the FARS dataset are similar. Keeping all else same, a fatal crash that results in an immediate fatality is clearly much more severe than another crash that leads to fatality after several days. Research attempts to discern such differences are useful in determining what factors affect the time between crash occurrence and time of death so that countermeasures can be implemented to improve safety situation and to reduce crash related fatalities.

To that extent, the current chapter makes a three-fold contribution to the literature on vehicle occupant injury severity analysis. First, our study is the first attempt to analyze the fatal injury from a new perspective and examine fatality as a continuous spectrum based on survival time ranging from dying within thirty days of crash to dying instantly. For the empirical analysis, the fatality timeline information obtained through FARS was categorized as an ordered variable ranging from death in thirty days to instantaneous death in seven categories as follows: died within 6th-30 days of crash, died within 2nd-5 days of crash, died within 7th-24 hours of crash, died within 1st-6 hours of crash, died within 31st-60 minutes of crash, died within 1st-30 minutes of crash and died instantly. Second, we estimated two-equation model that comprises of regression for EMS response time and ordered outcome model with residuals from the EMS model to correct for endogeneity bias on the effect of exogenous factors on the timeline of death. In doing so, the correction for endogeneity bias was pinned down in the ordered outcome models by employing a two-stage residual inclusion (2SRI) approach. In the research effort, we estimated the following three ordered outcome models with endogenous treatment: 1) OL with the 2SRI treatment, 2) GOL with the 2SRI treatment and 3) MGOL model with the 2SRI treatment while employing a comprehensive set of exogenous variables (driver characteristics, vehicle characteristics, roadway design and operational attributes, environmental factors, crash characteristics and situational variables). The comparison exercise highlighted the superiority of the MGOL model with the 2SRI treatment on the sample in terms of data fit compared to the other ordered outcome models in the current study context. Finally, we computed elasticity measures to identify important factors affecting survival time after motor vehicle crash. In our research, to further understand the impact of various exogenous factors, elasticity effects were estimated. The elasticity effects indicated that there were considerable differences in the elasticity effects across different fatal crash categories, suggesting that fatality is not a single state but rather is made up of multiple discrete states from dying instantly to dying within thirty days of crash. The most significant variables in terms of lower survival probability for drivers were crashes on high speed limit roads, crashes on medium speed limit roads and head-on crashes. In terms of longer survival probability, the important factors were old driver, front-to-rear crash and crashes during off-peak period. Moreover, the elasticity analysis assisted in providing a clear picture of attribute impact on driver time-to-death variables.

Table 5.1: Distribution of Fatal Injury Severity Categories

Fatal Crash Categories	Frequency	Percentage
Died between 6th to 30 days of crash	302	5.9%
Died between 2nd to 5 days of crash	270	5.3%
Died between 7th to 24 hours of crash	233	4.6%
Died between 2nd to 6 hours of crash	1175	23.0%
Died between 31st to 60 minutes of crash	824	16.1%
Died between 1st to 30 minutes of crash	1086	21.3%
Died instantly	1212	23.8%
Total	5102	100.0%

Table 5.2: Crash Database Sample Statistics

Categorical Explanatory Variables	Sample Share	
	Frequency	Percentage
Driver Characteristics		
Driver age		
Age 24 & less	1144	22.423
Age 25-64	2915	57.134
Age 65 & above	1043	20.443
Under the influence of alcohol	1778	34.849
Vehicle Characteristics		
Vehicle age		
Vehicle age<11 years	2822	55.312
Vehicle age≥11 years	2280	44.688
Roadway Design and Operational Attributes		
Speed limit		
Speed limit less than 26 mph	261	5.116
Speed limit 26 to 50 mph	2059	40.357
Speed limit above 50mph	2782	54.528
Traffic control device		
No traffic control, traffic signal and yield sign	4271	83.712
Stop sign	401	7.860
Other traffic control device	430	8.428
Roadway functional class		
Principal Arterial	1680	32.928
Minor Arterial	997	19.541
Collector	1208	23.677
Local Road	1217	23.853
Land use		
Rural	3206	62.838
Urban	1896	37.162
Environmental Factors		
Time of day		
Morning Peak	548	10.741
Off-peak	1266	24.814
Evening peak	828	16.229
Late evening	1311	25.696
Late night	1149	22.521
Lighting condition		
Daylight and other lighting condition	2910	57.036
Dark-not lighted	1430	28.028
Dark-lighted	762	14.935
Weather condition		

Dry	3601	70.580
Rain	422	8.271
Snowy	210	4.116
Cloudy	850	16.660
Other weather condition	19	0.372
Crash Characteristics		
Manner of collision		
Front to rear	124	2.430
Head-on	897	17.581
Other type of collision	4081	79.988
Collision location		
Non-Intersection	75.931	75.931
Intersection	15.759	15.759
Other Location	8.310	8.310
Situational Variables		
Driver ejection		
Ejected	1197	23.461
Not ejected	3905	76.539
Ordinal/Continuous Explanatory Variables	**Mean**	
Previous Recorded suspensions and revocations	0.444	
Previous record of other harmful motor vehicle convictions	0.323	
Number of passengers	0.400	
Logarithm of EMS response time (in minute)	2.473	

Table 5.3: Measures of Fit in Estimation Sample

Summary Statistic	OL	GOL	MGOL
Log-likelihood at zero	-9928.0	-9928.0	-9928.0
Log-likelihood at sample shares	-9016.3	-9016.3	-9016.3
Number of observations	5102	5102	5102
Summary Statistic	**Without 2SRI Treatment**		
Log-likelihood at convergence	-8844.8	-8794.9	-8793.7
Number of parameters	18	28	30
Summary Statistic	**With 2SRI Treatment**		
Log-likelihood at convergence	-8839.8	-8790.8	-8787.4
Number of parameters	19	29	31
Log-likelihood (LR) test	**LR Test Values**		
OL without 2SRI/OL with 2SRI	9.9 (1 degree of freedom)		
GOL without 2SRI/GOL with 2SRI	8.2 (1 degree of freedom)		
MGOL without 2SRI/MGOL with 2SRI	12.6 (1 degree of freedom)		
OL with 2SRI/GOL with 2SRI	98.1 (10 degrees of freedom)		
OL with 2SRI/MGOL with 2SRI	104.8 (12 degrees of freedom)		
GOL with 2SRI/MGOL with 2SRI	6.8 (2 degrees of freedom)		

Table 5.4: MGOL Estimates

Variables	Latent Propensity	τ_2	τ_3	τ_4	τ_5	τ_6
Constant	-1.712(-5.229)	-0.441(-6.101)	-0.854(-13.236)	0.141(2.296)	-0.104(-1.386)	0.069(1.915)
Driver Characteristics						
Driver age (Base: Age 25-64)						
Age 24 & less	-0.147(-2.207)*	–	–	–	–	–
Age 65 & above	-1.015(-10.966)	–	–	-0.281(-4.334)	-0.182(-2.071)	–
Under the influence of alcohol	0.488(3.488)	0.434(3.261)	–	–	–	–
Previous Recorded suspensions and revocations	-0.068(-3.264)	–	–	–	–	–
Previous record of other harmful motor vehicle convictions	0.104(2.598)	–	–	–	–	–
SD	0.208(3.596)	–	–	–	–	–
Vehicle Characteristics						
Vehicle age (Base: Vehicle age<11 years)						
Vehicle age≥11 years	–	–	–	–	-0.157(-2.689)	–
Roadway Design and Operational Attributes						
Speed limit (Base: Speed limit<26 mph)						
Speed limit 26 to 50 mph	0.251(2.117)	–	–	–	–	–
Speed limit above 50mph	0.359(2.981)	–	–	–	–	–
SD	0.447(2.707)	–	–	–	–	–
Traffic control device (Base: No traffic control, traffic signal and yield sign)						
Stop sign	0.223(1.975)	–	–	–	–	–
Other traffic control device	0.171(2.148)	–	–	–	–	–
Environmental Factors						
Time of day (Base: Morning Peak, Late evening and Late Night)						
Off peak	-0.218(-3.157)	–	–	–	–	-0.161(-2.323)

Evening peak	-0.151(-2.012)	–	–	–	–	–
Weather condition (Base: Dry, Rain, Snowy and Other weather condition)						
Cloudy	0.467(2.987)	0.276(1.872)	–	–	–	–
Crash Characteristics						
Manner of collision (Base: Other type of collision)						
Front to rear	-0.317(-1.765)	–	–	–	–	–
Head-on	0.661(5.312)	–	–	0.261(3.781)	–	–
Collision location (Base: Intersection and Other location)						
Non-intersection	0.362(3.741)	–	–	0.217(3.346)	-0.234(-3.168)	–
Situational Variables						
Driver ejection (Base: Not ejected)						
Ejected	0.267(3.651)	–	–	–	0.145(1.963)	–
Number of passengers	0.159(4.874)	–	–	–	–	–
EMS response time						
Logarithm of EMS response time *(in minute)*	0.247(1.993)					
Residual from regression of Logarithm of EMS arrival time (in minute) on morning peak, late night, dark-not lighted, rain, snowy, rural, principle arterial and minor arterial	-0.363(-2.929)	–	–	–	–	–

τ_2= Threshold between 1st-5 days/ 7th-24 hours; τ_3= Threshold between 7th-24 hours/ 1st-6 hours; τ_4 = Threshold between 1st-6 hours/ 31st-60 minutes; τ_5 = Threshold between 31st-60 minutes/ 1st-30 minutes; τ_6 = Threshold between 1st-30 minutes/ Died Instantly

* t-stats are presented in parenthesis

Table 5.5: Linear Regression Estimates

Variables	Coefficient	t-stat
Constant	2.190	80.670
Roadway functional class (Base: Collector and Local road		
Principal Arterial	-0.074	-2.709
Minor Arterial	-0.118	-3.699
Land use (Base: Urban)		
Rural	0.363	14.183
Time of day (Base: Off-peak, Evening peak and Late evening)		
Morning Peak	0.070	1.794
Late night	0.213	6.877
Lighting condition (Base: Daylight and other lighting condition and Dark-lighted)		
Dark-not lighted	0.120	4.161
Weather condition (Base: Dry, Cloudy and Other weather condition)		
Rain	0.088	2.052
Snowy	0.145	2.419

Table 5.6: Elasticity Effects

Variables	Died between 6-30 days	Died between 2-5 days	Died between 7-24 hours	Died between 2-6 hours	Died between 31-60 minutes	Died between 1-30 minutes	Died instantly
Driver Characteristics							
Driver age (Base: Age 25-64)							
Age 24 & less	13.694	11.328	9.375	6.150	0.648	-4.202	-10.384
Age 65 & above	108.781	93.747	74.488	2.099	-18.729	-17.964	-35.615
Under the influence of alcohol	-39.798	22.427	-7.732	-5.153	-0.658	3.421	8.611
Previous Recorded suspensions and revocations	6.270	5.246	4.367	2.871	0.312	-1.950	-4.823
Previous record of other harmful motor vehicle convictions	-6.784	-6.263	-5.725	-4.674	-2.074	1.511	8.782
Vehicle Characteristics							
Vehicle age (Base: Vehicle age<11 years)							
Vehicle age≥11 years	0.000	0.000	0.000	0.000	-15.654	3.016	7.977
Roadway Design and Operational Attributes							
Speed limit (Base: Speed limit<26 mph)							
Speed limit 26 to 50 mph	-22.181	-18.747	-15.721	-10.576	-1.590	6.632	18.132
Speed limit above 50mph	-25.631	-23.452	-21.341	-16.467	-5.703	7.403	28.948
Traffic control device (Base: No traffic control, traffic signal and yield sign)							
Stop sign	-18.668	-15.959	-13.637	-9.645	-2.112	5.457	16.739
Other traffic control device	-14.440	-12.444	-10.636	-7.420	-1.487	4.369	12.715
Environmental Factor							
Time of day (Base: Morning Peak, Late evening and Late Night)							
Off peak	20.164	17.115	14.241	9.234	0.793	-18.852	-4.193
Evening peak	14.059	11.726	9.714	6.322	0.590	-4.386	-10.591
Weather condition (Base: Dry, Rain, Snowy and Other weather condition)							
Cloudy	-36.654	1.822	-14.118	-9.567	-1.671	5.906	16.468
Crash Characteristics							

Manner of collision (Base: Other type of collision)							
Front to rear	31.998	25.907	20.893	12.738	-0.027	-9.992	-21.155
Head-on	-49.427	-44.114	-39.224	4.763	-1.926	7.301	19.919
Collision location (Base: Intersection and Other location)							
Non-intersection	-34.151	-29.006	-24.066	12.236	-24.944	7.878	17.758
Situational Variables							
Driver ejection (Base: Not ejected)							
Ejected	-22.368	-19.547	-16.760	-11.686	12.997	4.252	11.816
Number of passenger	-13.339	-11.499	-9.842	-6.893	-1.428	4.007	11.851
EMS response time	-2.410	-2.001	-1.706	-1.157	-0.227	0.690	2.033

CHAPTER 6 Pooling Data from Fatality Analysis Reporting System (FARS) and Generalized Estimates System (GES) to Explore the Continuum of Injury Severity Spectrum

6.1 Introduction

A number of studies have explored the impact of various factors on vehicle occupant injury severity at disaggregate level (see Bédard et al., 2002; Fredette et al., 2008 for a detailed review). These studies can broadly be categorized as: a) studies that focus exclusively on crashes involving only fatalities (employing a sample of crashes involving fatalities) and b) studies that examine crashes that involve all levels of injury severity – ranging from no injury to fatality (employing a random sample of traffic crashes that compile different levels of injury severity). In the US, the former category of studies predominantly use the Fatality Analysis Reporting System (FARS) database (see Evans and Frick, 1988; Preusser et al., 1998a; Zador et al., 2000; Gates et al., 2013) while the latter group of studies primarily employ the General Estimates System (GES) database (see Kockelman and Kweon, 2002; Eluru and Bhat, 2007).

The FARS database is a census (not a sample) of all fatal crashes in the US; i.e., crashes that lead to at least one fatality within thirty consecutive days from the time of crash. The GES database, on the other hand, comprises a sample of road crashes across the US involving at least one motor vehicle travelling on a roadway and resulting in property damage, injury or death to the road users. The two datasets employed in the safety literature have their own advantages and limitations. The FARS focuses exclusively on fatal crashes. Therefore, one cannot reliably use this data to analyze the factors that increase or decrease the probability of fatality (because the data does not include crashes that do not lead to fatalities). The GES fills this gap by compiling data on a sample of roadway crashes involving all possible severity consequences (no injury, possible injury, non-incapacitating injury, incapacitating injury and fatality) providing a more representative sample of traffic crashes in the US. One of the advantages of FARS, however, is that the collected information includes the date and time of occurrence of the fatalities resulting within a 30-day time period from the crash. This detailed information provides us a continuous timeline of the fatal occurrences from the time to crash (instead of considering all fatalities to be the same). This allows for an analysis of the survival time of victims before their death. The GES, on the other hand, does not offer such detailed information except identifying who died in the crash.

Examining the impact of various exogenous factors on all levels of injury severity as well as on the survival time of fatalities can potentially play a critical role in field triage - screening process to determine the more severe cases. Preclinical trauma care is one of the most important factors affecting the outcome of motor vehicle crash (MVC) victims (Chalya et al., 2012; Palanca et al., 2003). In pre-hospital setting, along with the anatomic and physiological conditions of MVC victims, different mechanism-of-injuries (vehicle intrusion, occupant ejection, vehicle telemetry and death in same passenger compartment) are also considered by emergency medical service (EMS) personnel as conditions for trauma triage of victims (Sasser et al., 2012; Isenberg et al., 2011). In fact, it is evident from previous studies (Stewart, 1990) that prolonging survival beyond the first hour can potentially help avoid fatality with proper preclinical care. Hence, a refined specification of fatality might allow us to identify potential survivors that might benefit by providing emergency treatment.

In an effort to identifying exogenous factors that help in prolonging survival time, using detailed information available in FARS data, we examined fatal crashes from a new perspective in the preceding chapter. We identified that fatality is an aggregation of a continuous spectrum ranging from dying instantly to dying within thirty days of crash (as reported in the FARS data). Keeping all else same, a fatal crash that results in an immediate fatality is clearly much more severe than another crash that leads to fatality after several days. Therefore, it is useful to explicitly recognize the different levels of severity among fatal crashes. Such refined definition of fatal crashes, as opposed to lumping all fatal crashes into a single category, allows one to differentiate fatal crashes based on the survival time and to derive insights on factors that can prolong survival time. While using the FARS data is very helpful for understanding the differences across different fatal crashes, it inherently excludes crashes with other possible, non-fatal injury severity outcomes. This makes it difficult to generalize the findings to the overall crash population. Besides, while analyzing the survival time of only fatal crash victims (using FARS data) helps in deriving the influence of various exogenous factors on survival time conditional upon the occurrence of a fatality, it doesn't allow the analyst to derive the influence of those factors in increasing the chances of survival. This is because the FARS data doesn't provide a representative sample of non-fatal crashes.

One way to address this issue is combining information from both the FARS and GES datasets into a single, disaggregate crash-level database[11]. This will bring together the strengths of both datasets – the representativeness of crashes with all injury severity outcomes from the GES data and the detailed information on fatal crashes from the FARS data. The challenge, however, lies in combining the two datasets in a statistically appropriate way. Since FARS is a census of all fatal traffic crashes in the US, all fatal crashes in the GES sample for a year should be available in the FARS data for that year. Now, if one could identify these crashes directly, it would be easy to augment the fatal crash records in GES with the detailed information from FARS. However, there is no mechanism to easily link crashes across these two databases because the datasets do not have a common identifier. Hence, an alternative, statistically valid method needs to be used for fusing information from both the datasets.

The approach is a proof of concept investigation of data pooling from two datasets while ensuring statistical validity. While, there could be various other alternative datasets for such investigation, given the extensive use of GES and FARS datasets in safety literature, they serve as good candidates for the research exercise. In this context, this chapter is geared towards addressing the challenge of pooling data from GES and FARS. While several approaches exist in the literature to fuse information from different data sources without a common identifier (Konduri et al., 2011; Sivakumar and Polak, 2013), a simple approach is to replace fatal crashes from the GES sample by a random sample from the FARS census of crashes. We conduct statistical tests to assess if this approach suffices for the purpose of developing a database that allows us to examine the whole spectrum of injury severity ranging from no injury to fatality, along with differentiating fatal crashes based on survival time. Moreover, the simultaneous interpretation of information would allow researchers to provide recommendations using a single modeling framework, rather than making inferences from the results of separate econometric models from different datasets.

In summary, the current chapter makes a three-fold contribution to the literature on driver injury severity analysis. First, we propose and test the efficacy of a simple yet statistically valid approach to fuse the FARS and GES datasets into a single, disaggregate crash level database that combines information from both the datasets. Second, the Generalized Ordered Logit (GOL)

[11] To be sure, the reader would note that there have been compilation of GES and FARS datasets to obtain the Annual Traffic Safety Facts (see NHTSA, 2010). However, in these efforts, there is no attempt to pool data from the two sources. The report provides trends separately for FARS and GES datasets. Further, in our research, we examine the effect of exogenous variables on severity in pooled and un-pooled data.

model (also referred to as Partial Proportional Odds model) is employed on the pooled dataset to analyze the influence of a variety of exogenous factors on traffic crash injury severity, while considering a very refined characterization of fatal crashes along with other, non-fatal injury severity outcomes. Finally, we compute elasticity measures to identify important factors affecting driver injury severity outcomes.

The rest of the chapter is organized as follows. The data source and sample formation are presented in Section 6.2. Section 6.3 provides details of the approach used for pooling data from FARS and GES. Section 6.4 presents the empirical analysis along with a statistical assessment of the proposed approach to fuse information from both data sources. The estimation results of the GOL model are described in section 6.5. The elasticity effects are presented in section 6.6 and section 6.7 concludes the chapter.

6.2 Data Source and Sample Formation

The data for the current study is sourced from the FARS and GES databases for the year 2010. The datasets are briefly described in Section 2.3.1 and Section 5.2.1, respectively. The reader would note that the exogenous variable information available in FARS and GES datasets are very similar making it relatively easier to fuse the fatality information from FARS into the GES data.

This chapter is focused on injury severity outcome of passenger vehicles' drivers who were involved in either a single or two vehicle crashes. The crashes that involve more than two vehicles are excluded from both FARS and GES datasets. Commercial vehicles involved collisions are also excluded in order to avoid the potential systematic differences between commercial and non-commercial driver groups. In order to prepare the final FARS dataset, crash records involving non-motorized road users (19,670 records), commercial vehicles (17,795 records), records with passenger and more than two vehicles (18,073 records), non-fatal crash records of drivers (8,012 records) and records with missing information for essential attributes (2,468 records) are deleted. Thus, the final FARS dataset consisted of 8,845 records. From the continuous timeline of the fatal occurrences, a seven point discrete ordinal variable is created to represent the scale of fatal injury severity of drivers involved in these crashes - from least severe to most severe fatal crashes (and their proportions): 1) Died between 6th-30 days of crash (6.0%), 2) Died between 2nd-5 days of crash (5.2%), 3) Died between 7th-24 hours of crash (4.4%), 4) Died between 2nd-6 hours of crash (21.6%), 5) Died between 31st-60 minutes of crash (14.5%), 6) Died between 1st-30 minutes of crash (20.1%) and 7) Died instantly (28.3%). In order to prepare the final GES dataset, crash records involving non-motorized road users and commercial vehicles (34,808 records), records with passenger and more than two vehicles (32,824 records), and records with missing information for essential attributes (23,094 records) are deleted. Thus, the final GES dataset consisted of about 25,294 records. From this dataset, a sample of 6,062 records is randomly sampled out for the purpose of estimating models. The reader would note that the sampling process was employed to reduce the computational time necessary to estimate and compare the models described subsequently. A five point ordinal scale is used in the database to represent the injury severity of individuals involved in these crashes. In the estimation sample, the distributions of driver injury severities are as follows: No injury 63.7%, Possible injury 14.0%, Non-incapacitating injury 13.1%, Incapacitating injury 8.2% and Fatal injury 1.0%.

6.3 Research Framework

In the current research effort, we employ the Generalized Ordered Logit (GOL) or the partial proportional odds logit model (in Chapter 2 we provide a detailed description of the econometric framework) to examine the driver injury severity by using pooled dataset from FARS and GES. The injury severity variable is analyzed using the ordered outcome framework to recognize the inherent ordinality of the injury severity levels. However, the prerequisites for any data pooling exercise are that different sources to be pooled are comparable (Verma et al., 2009) and share a common data generation process (Louviere et al., 1999). This section presents an approach to pool information from both the data sources and the tests used to assess if the pooled data represents a common data generation process for the individual data sources. A conceptual diagram of the research methodology employed is provided in Figure 6.1.

6.3.1 Testing Data Pooling Exercise

The GES dataset has a five point ordinal scale to represent injury severity while a seven point ordinal scale is defined to distinguish the severity of different fatal crashes based on the survival time. In this chapter, we form the pooled dataset by replacing the fatal crash records in GES with a random sample of crashes in FARS. In the pooled dataset we can generate an eleven point ordinal representation of injury severity, with 4 categories for non-fatal crashes and 7 categories for fatal crashes (5 + 7 – 1). Prior to developing models to analyze the newly generated injury severity scale, it is imperative that we validate the pooled dataset. As the actual data generation process is latent we have to resort to comparing the pooled dataset with the un-pooled dataset. In our pooling exercise, the records from FARS are being added to the GES data, the evaluation would be geared towards comparing the pooled data with the original GES data (un-pooled data). Specifically, we undertake comparison of the pooled sample with the un-pooled sample in two ways: (1) uni-variate sample comparison, by simply comparing the distributions of the variables in the two samples and (2) econometric model estimate comparison.

While the descriptive comparison of pooled and original samples is relatively straight forward, the more challenging task is to perform a more statistically rigorous analysis to examine if the crash records from FARS can replace those in the GES data. For this purpose, as a *first step*, we estimate the injury severity model using the original GES data and compare the model estimates with the injury severity model estimated from the pooled dataset – while maintaining the same number of injury severity categories in the GES and pooled datasets. To do so, all the fatal records pooled from FARS into the GES sample were categorized as fatal (i.e., a single category) regardless of the survival time of the victims. The pooled data sample is obtained by removing the 59 fatal records in the GES sample of 6,062 records.

To statistically ensure the validity of our comparison results and to ensure that the statistical results obtained from the pooled samples are stable, we consider multiple samples of fatal crash data from FARS to replace fatalities in GES. Specifically, for testing the validity of the pooled data, 15 data samples – 5 samples of about 2,000 records, 5 samples of about 3,000 records and 5 samples of about 5,000 records – are randomly generated from the 8,845 records of FARS database and combined with the GES data to form pooled data. These 15 data samples along with the full sample (of 8,845 records) from FARS dataset are used to generate 16 different sets of pooled databases. The fatal records replaced in GES by the FARS fatal records in these 16 samples are presented in Table 6.1. GOL models of injury severity are estimated for these 16 pooled samples under the five point ordinal scale system and compared with the GOL model parameters obtained using un-pooled GES data to ensure that the estimates have not been altered significantly due to the newly added records.

6.3.2 Weight Variable for Pooling

The reader would note, from Table 6.1, that the GES (un-pooled) database has a very small percentage of fatalities. This is because the percentage of fatal crashes is small compared to all other crashes. As our primary objective is examining the impact of exogenous variables on seven categories of the fatality spectrum (based on survival time) it is useful to oversample the fatal crashes from FARS. Otherwise, we are likely to have very small number of records for each of the fatal injury severity alternatives. Of course, the oversampling of fatalities from FARS to replace GES fatalities necessitates creating an appropriate weight variable to weight the pooled data. This approach ensures that the distribution of the injury severity variable in the pooled data is the same as that in the GES data. Therefore, to generate the pooled sample, we remove the fatal crashes (m_i) from the GES sample and replace it with fatal cases (n_i) from the FARS along with a specific weight ω_{FG} computed as $\frac{m_i}{n_i}$. Specifically, a weight of ω_{FG} is assigned to the FARS crash records (that replace the GES fatalities) in the pooled samples while the other non-fatal crash records (from GES) were weighted by 1. The associated weights for 16 different pooled samples are shown in Table 6.1.

6.3.3 Severity Parameters Comparison Exercise

The 16 pooled data samples created with appropriate weights are employed to generate injury severity parameter estimates. The parameter estimates obtained using the pooled data are compared with that of the original GES parameter estimates obtained using un-pooled data (i.e., the original GES data) by computing the percentage error (considering parameter estimates from un-pooled data as the base case). Then, a hypothesis test that the parameters are obtained from the same distribution ($i.e., \beta_P = \beta_{UP}$[12] where *P*=Pooled and *UP*=Un-pooled) is carried out to examine the differences between parameter estimates. If this hypothesis is rejected, the estimates from pooled model represent estimates from a dissimilar latent data generation process (Bass and Wittink, 1975). On the contrary, if the hypothesis is not rejected, it will provide support that the

[12] To test the hypothesis that $\beta_P = \beta_{UP}$, we need to obtain the distribution of $(\beta_P - \beta_{UP})$. The standard error for the distribution is obtained as $\sqrt{SE_P^2 + SE_{UP}^2}$ where SE_P and SE_{UP} represent standard errors of the parameters obtained using pooled and un-pooled data respectively. Then, one can simply do a t-test on $(\beta_P - \beta_{UP})$. That is, if the ratio of the estimate of $(\beta_P - \beta_{UP})$ to its standard error is less than the critical t-value at a chosen confidence level, then one cannot reject the hypothesis that $\beta_P = \beta_{UP}$.

proposed pooling of GES and FARS datasets has not altered the distribution of the parameters and that the pooling process is statistically valid. The percentage error in parameter estimates and the hypothesis tests are first computed separately for each of the 16 pooled data samples. Subsequently, for ease of presentation, we present and discuss the average measures from each sample type – 1 pooled sample with 8,845 records from FARS (sample 16th of Table 6.1); 5 pooled samples with about 5,000 records from FARS (samples 11-15th of Table 6.1); 5 pooled samples with about 3,000 records from FARS (samples 6-10th of Table 6.1); and 5 pooled samples with about 2,000 records from FARS (samples 1-5th of Table 6.1).

6.3.4 Eleven Point Pooled Model

After we confirm that the differences in model estimates from the five point ordinal models are within an acceptable margin, we can employ the pooled data to estimate an injury severity model with an eleven point severity scale with 4 categories of injury severity for non-fatal crashes and 7 categories for fatal crashes. For our analysis we chose one sample from the 16 different pooled data samples for the purpose of estimating the best specified eleven point GOL model. The chosen sample has 2,967 randomly sampled records from the FARS data to replace the 59 fatal records from GES and the remaining 6,003 records from the GES data.

6.4 EMPIRICAL ANALYSIS

6.4.1 Variables Considered

In our analysis, to estimate models using pooled data, we prepared the datasets such that both GES and FARS datasets have exactly the same set of independent variables. We selected a host of variables from five broad categories: Driver characteristics (including driver gender, driver age, restraint system use, alcohol consumption and physical impairment), Vehicle characteristics (including vehicle type and vehicle age), Roadway design and operational attributes (including roadway class, speed limit, types of intersection and traffic control device), Environmental factors (including time of day and road surface condition) and Crash characteristics (including collision object, manner of collision, collision location and trajectory of vehicle's motion). It should be noted here that several variables such as presence of shoulder, shoulder width, point of impact, number of lanes, lighting condition could not be considered in our analysis because either the information was entirely unavailable or there was a large fraction of missing data for these attributes in the dataset. To be sure, we employ the manner of collision and time of day variables as surrogates for point of impact and lighting condition, respectively. In the final specification of the model, statistically insignificant variables were removed. The reader would note that the pooling exercise was undertaken using the variables that are common to both datasets. Hence, variables such as emergency crew arrival times were not considered in our models as they are unavailable in GES data.

6.4.2 Validation Exercise of Pooled Data

The first step in the validation exercise was to examine the similarities and dissimilarities in independent variables across the pooled and un-pooled samples. In the comparison, we found that the exogenous factor distributions of all pooled datasets (16 datasets) are almost the same. For the sake of brevity we chose to present the results for one sample only. The sample characteristics of the exogenous factors of un-pooled and one pooled (weighted) dataset are presented in Table 6.2. Overall, we find that the characteristics of the pooled and un-pooled samples across the entire sample (in columns 2 and 3) and across fatal crashes (columns 4 and 5) are very similar. We observe that there are slightly higher proportions of driving under the influence of alcohol and negotiating curves among the fatal crashes in the pooled data than those in the un-pooled data. Also, the proportions for fatal crashes in the pooled dataset are marginally lower for two way traffic-with median and for vehicle age 6-10 years. It is not unanticipated that pooling would introduce such minor differences between the datasets.

In the second step of our validation, a comparison exercise between the parameter estimates obtained using un-pooled and pooled data is also carried out by using 16 different pooled samples. The reader would note that a direct comparison of parameter estimates is considered only for illustrative purposes. A more rigorous statistical approach is also undertaken. The percentage errors in injury severity parameter estimates obtained using pooled datasets compared to parameter estimates obtained using un-pooled data are presented in Figure 6.2 for all the variables (variable numbers are defined in Table 6.4 along with the injury severity estimates obtained using the un-pooled model). From this plot, we can see that, among 44 variables in the final models, 32 variables have an error percentage lower than 10%, 8 variables have an error percentage between 10 and

25% and 4 variables have an error percentage higher that 25%. Overall, for such highly non-linear models such as GOL, estimated using two datasets, these are reasonably small differences.

To undertake a more rigorous statistical comparison, we test the hypothesis that the parameter estimates obtained using the pooled and un-pooled datasets are not systematically different and the observed numerical differences can be accounted by the randomness in data samples. The test values of the homogeneity hypothesis test ($\beta_P = \beta_{UP}$) between parameter estimates obtained using un-pooled and pooled datasets are plotted against the variable numbers and is presented in Figure 6.3. From this plot, we can clearly see that the test statistics lie within the bounds +1.96 and -1.96 (critical t-stats at 95% confidence level). In fact, the largest difference is less than 1 indicating that there is no systematic difference in the estimates from pooled and un-pooled models. This same trend can be observed for all types of pooled data samples with different numbers of FARS records in the pooled data. Thus, we can find no evidence to reject the hypothesis that the severity parameter estimates obtained using pooled data and the severity parameter estimates obtained using un-pooled data follow different distribution. Based on our comparison of descriptive statistics and severity parameter estimates, we can argue that there is no evidence to suggest that the data pooled from GES and FARS results from a distinct latent data generation process than that in GES.

6.4.3 Metric for Comparing Eleven Point Model with Five Point Model

Another issue that needs to be addressed before estimating the eleven point scale ordinal model is developing a statistical approach to determine if the eleven point ordinal model is an improvement on the five point ordinal model (with all fatal crashes lumped into a single category). Due to the nature of the log-likelihood measure employed in model estimation, increasing the resolution will lead to deterioration of model log-likelihood. Hence, comparing log-likelihoods between a five alternative model and eleven alternative model is not statistically valid. Interestingly, we could not find any method in literature to make a meaningful comparison of models with different resolutions of dependent variable definitions. Hence, we developed an approach based on first principles to address this issue. In a five point ordinal scale model all fatalities are treated equally i.e. there is no distinction across fatal crashes. So in a five alternative model, we implicitly assume that the seven fatality groups considered in the eleven alternative model are all equally likely. Recognizing this assumption, one could generate an equivalent eleven alternative log-likelihood based on the five alternative model log-likelihood value. This can be compared with the log-likelihood of the eleven alternative model that differentiates between the various fatality classes.

The exact equation for the computation of log-likelihood takes the following form:

$$L = \sum_{i=1}^{N}\left[\left(\sum_{j=1}^{4}\left(logP_i(j)\right)^{d_{ij}}\right) + \frac{1}{7} * \omega_{FG} * \left(logP_i(J)\right)^{d_{i5}}\right] \tag{6.1}$$

where, ω_{FG} is the weight, i be the index for drivers ($i = 1,2,\dots,N$), j be the index for driver injury severity levels ($j = 1,2,\dots\dots\dots,J$), $P_i(j)$ represents the probability of injury severity level j, and d_{ij} is a dummy variable taking the value 1 if the driver i sustains an injury of level j and 0 otherwise. Once the equivalent log-likelihood is generated based on the above equation, one could easily employ the likelihood ratio (LR) test to check if the eleven point ordinal scale model offers additional improvement. The LR test statistic is defined as 2 * (LL_{11} – LL_5) where LL_{11} and LL_5 represent log-likelihood values at convergence of the eleven point and equivalent five point ordinal models, respectively. The LR test statistic thus computed is compared with the chi-square distribution value of k degrees of freedom where k corresponds to the additional parameters in the unrestricted model. In our case, for all samples, the additional number of parameters is 6. Hence, if the LR test statistic is larger than the ψ^2 value for 6 degrees of freedom, we can conclude that the considering fatality as multiple states enhances the data fit.

The log-likelihood values along with the LR test statistic for the equivalent and the actual eleven point models for various samples are presented in Table 6.3. The resulting LR test values for the comparison of equivalent/actual eleven point models for all sample types are more than 23 indicating the actual eleven point model outperforms the equivalent eleven point model at any reasonable level of statistical significance. The consistent improvement offered by the pooled model clearly indicates that the refined categorization of fatal injury crashes improves the model fit and provides more information to the model for examining the injury severity outcome. This is of particular relevance to this empirical exercise because fatal crashes comprise a very small portion of our sample (only 1%) – thus by introducing further disaggregation of an alternative with such a small sample share, there was a risk of worsening the model.

6.5 Estimation Results

The best specification model for the eleven point ordinal injury severity categorization is discussed in this section. To reiterate, the dependent variable under consideration is the eleven point ordinal variable defined as: no injury, possible injury, non-incapacitating injury, incapacitating injury, and 7 categories within fatal crashes - died between 6th-30 days of crash, died between 2nd-5 days of crash, died between 7th-24 hours of crash, died between 2nd-6 hours of crash, died between 31st-60 minutes of crash, died between 1st-30 minutes of crash and died instantly. The estimation results for the sample of 2,936 records from FARS are presented in Table 6.5. In GOL model, when the threshold parameter is positive (negative), the result implies that the threshold is bound to increase (decrease); the actual effect on the probability is quite non-linear and can only be judged in conjunction with the influence of the variable on propensity and other thresholds. In the following sections, the estimation results are discussed by variable groups.

Driver Characteristics: In the category of driver characteristics, the result for driver gender indicates higher injury risk propensity for female drivers compared to male drivers. The effect of this variable is also significant for the threshold demarcating possible and non-incapacitating injury. The positive sign of the coefficient in the threshold indicates higher likelihood of possible injury for the female drivers. The result is perhaps indicative of the lower physiological strength of female drivers (compared to male drivers) in withstanding the impact of a crash (Xie et al., 2009; Chen and Chen, 2011). The age of drivers involved in the collision also has a significant influence on injury severity. The parameter characterizing the effect of young driver (age<25) suggests a reduction in the likelihood of severe injuries compared to middle-aged drivers (age 25 to 64), perhaps indicating the higher physiological strength of young drivers in withstanding crash impacts (see Xie et al., 2012; O'Donnell and Connor, 1996; Castro et al., 2013 for similar result). However, the estimation result indicates that compared to the middle aged driver, the latent injury propensity is higher for older drivers (age≥65), while the negative sign of threshold demarcating the possible and non-incapacitating injury indicates a lower likelihood of possible injuries and, in general, a higher likelihood of dying instantly for older drivers.

The result related to drunk driving indicates that alcohol impairment leads to higher injury risk propensity of drivers compared to sober drivers. The negative effect of this variable on the threshold separating the possible injury and non-incapacitating injury level indicates a lower likelihood of possible injury for the alcohol impaired drivers. The net implications of these effects is that alcohol impaired drivers have a lower likelihood of no injury and a higher likelihood of dying instantly in a crash compared to sober drivers. A crash involving physically impaired drivers is associated with an overall higher injury risk propensity. The result may be reflecting increased reaction times for physically impaired drivers. As expected, injury risk propensity is higher for the drivers not wearing seat belts relative to the buckled up drivers (see Obeng, 2008; Yau, 2004; Yasmin et al., 2012; Eluru and Bhat, 2007 for a similar result).

Vehicle Characteristics: With respect to driver's vehicle type, the estimation results show that latent injury risk propensities are lower for the drivers of pickups and vans compared to the drivers of other passenger vehicles (passenger cars and SUV), presumably because pickups and vans have huge mass which offer more protection to the occupants of these vehicles (Kockelman and Kweon, 2002; Xie et al., 2009; Eluru et al., 2010; Fredette et al., 2008). The vehicle age results demonstrate that latent injury propensities are higher for drivers in older vehicles (vehicle age 6-10 years and vehicle age $\geq$ 11 years) relative to drivers in newer vehicles (vehicle age $\leq$ 5 years). As is expected, within the vehicle age categories considered the oldest vehicle age category has a larger impact relative to the moderately older vehicle age category. The effect of vehicle age variable on the threshold also indicates increased likelihood of possible injury for 6-10 years old vehicle. The higher injury risk of older vehicle's driver may be attributable to the absence of advanced safety features and/or the involvement of suspended and unlicensed drivers in older vehicles (Lécuyer and Chouinard, 2006, Kim et al., 2013; Islam and Mannering, 2006).

Roadway Design Attributes: Several roadway design attributes considered are found to be significant determinants of driver injury severity. Among those, the injury risk propensities are higher for crashes occurring on medium (26 to 50 mph) and high (above 50 mph) speed limit locations (with larger impact for high speed limit locations) compared to lower (less than 26 mph) speed limit locations (see Eluru et al., 2010; Chen et al., 2012; Tay and Rifaat, 2007 for similar results). The presence of traffic control device is also found to have significant effect on the severity of crashes. The influence of traffic control device reveals that the presence of other traffic control devices (such as warning sign, regulatory sign, railway crossing sign) increases the likelihood of injury risk propensity of the drivers, possibly suggesting non-compliance with these traffic control devices. With respect to intersection type, crashes occurring at T-intersection (relative to non- and all other types of intersection) have a lower injury risk propensity. This is perhaps a consequence of lower approaching speed of vehicles at a T-intersection.

Environmental Factors: Among different environmental factors explored in this study, only time-of-day and surface condition are significant in the final model specification. Compared to crashes during daytime and late evening (6.00 a.m. to 11.59 p.m.), the likelihood of injury risk propensity is found to be higher for late night (12.00 a.m. to 5.59 a.m.) period. This finding is consistent with several previous studies; attributable to reduced visibility, fatigue, longer emergency response times, higher driver reaction time and/or increased traffic speed (Plainis et al., 2006; Helai et al., 2008; Hu and Donnell, 2010; Kockelman and Kweon, 2002; de Lapparent, 2008). The surface condition effects simplified to a simple binary representation of presence/absence of snowy road surafce condition. The result indicates that if collisions occur on a snowy road surface (relative to those on a dry surface), the drivers are more likely to evade serious injury, perhaps due to reduced speeding possibility and/or could be related to more cautious driving (Edwards, 1998; Mao et al., 1997; Eluru and Bhat, 2007).

Crash Characteristics: As observed in several previous studies (Yamamoto et. al., 2004; Holdridge et al., 2005), the results related to collision object of our study reflect an increased injury risk propensity for collision with large object (related to collision with small object and moving vehicle). However, the effects of "collision with large object" (building, concrete traffic barrier, wall, tree, bridge, snow bunk) indicator in threshold parameterization are relatively complex. It has a negative impact on the threshold between possible and non-incapacitating injury; while it has a positive impact on the threshold between non-incapacitating and incapacitating injury. In general, the net implication is that collision with large object has a lower probability of sustaining no injury (the specific impact of other injury severity categories on driver injury severity are context-specific). The result also suggests that collision with other object (animal, non-fixed object) has a lower injury risk propensity. The negative effect of this variable on the threshold separating the possible injury and non-incapacitating injury level indicates a lower likelihood of possible injury for the drivers.

The results related to collision type reflect lower injury risk propensities for both rear-end and sideswipe-same direction collisions compared to angular collision. Rear-end collision also has a significant impact on the threshold between non-incapacitating and incapacitating injury; and implies higher probability of incapacitating injury. Head-on collision reflects the anticipated higher injury risk propensity compared to angular collision. This is perhaps a consequence of greater dissipation of kinetic energy. Crashes in driveway access location lead to an overall reduced injury risk propensity (relative to collision at non-intersection location). The impacts of driveway access on both of the first two thresholds are positive, which implies that the effects of driveway access on different injury categories are crash and driver-specific. However, the results suggest an increased probability of no injury category and, in general, a decreased possibility of instant death category, perhaps indicating driving at lower speed or more watchful driving at these locations (Rifaat and Tay, 2009). The results in Table 6.5 underscore that crashes at intersections and intersection related crashes do not have any effect on injury risk propensity. But the effect of the indicator variable on the threshold indicates a higher probability of possible injury and an overall lower probability of instant death in a crash (relative to the crashes at non-intersection location).

The effects of the trajectory of vehicle's motions underscore an overall higher injury risk propensity for the driver whose vehicle was stopped in a traffic lane compared to the one who was going straight at the time of collision. Also, the effect of stopped in a traffic lane variable on threshold between non-incapacitating and incapacitating injury indicates a higher likelihood of non-incapacitating injury. Both turning manoeuvres (left and right) of drivers have lower injury risk propensities compared to going straight. This may be reflecting more watchful driving as well as lower speeds while turning. Negotiating a curve does not have any effect on the risk propensity, but the indicator variable has a positive impact on the threshold between possible and non-incapacitating injury. This effect implies a higher probability of possible injury and an overall reduced probability for instant death (relative to going straight).

6.6 Elasticity Effects

The parameter effects of the exogenous variables in Table 6.5 do not provide the magnitude of the variable effects on the injury severity of drivers. To quantify the effects of these variables on driver injury severity outcomes, we compute the aggregate level “elasticity effects” for a selected set of independent variables – female driver, driver age≥65, driving under the influence of alcohol, vehicle age ≥ 11 years, high speed limit road, late night, collision with large stationary object and head-on collision. The elasticity estimates are presented in Table 6.6. For the ease of presentation, we focus only on the elasticity effects for the seven fatal crash injury severity categories. The numbers in the table can be interpreted as the percentage change (increase for positive sign and decrease for negative sign) in the probability of the crash severity categories due to the change in that specific exogenous variable.

The following observations can be made based on the elasticity effects presented in Table 6.6. First, the results in Table 6.6 indicate that there are considerable differences in the elasticity effects across different fatal crash categories. Specifically, the differences are substantial for collision on high speed limit road and head-on collision. This supports our hypothesis that the severity of fatal crashes is not a single, un-separable category but rather is a continuum ranging from dying instantly to dying within thirty days of crash. These results also suggest that considering a fine resolution categorization of fatal crashes in examining the crash injury severity outcome offers the potential to provide useful information for policy makers in developing the EMS system and trauma triage.

Second, the most important variables in terms of early death are collision on a high speed limit road, head-on collision and driving under the influence of alcohol. Finally, the elasticity analysis conducted provides an illustration of how the proposed pooled model can be applied to determine the critical factors contributing to reducing the survival time. For example, based on crash characteristic elasticities computed, if EMS services can identify *critical* crashes with likelihood for survival on the field it might assist in determining the appropriate mode of patient transfer (by road or air lifting depending on the crash characteristics) and also providing appropriate medical supervision at the hospital.

6.7 Summary

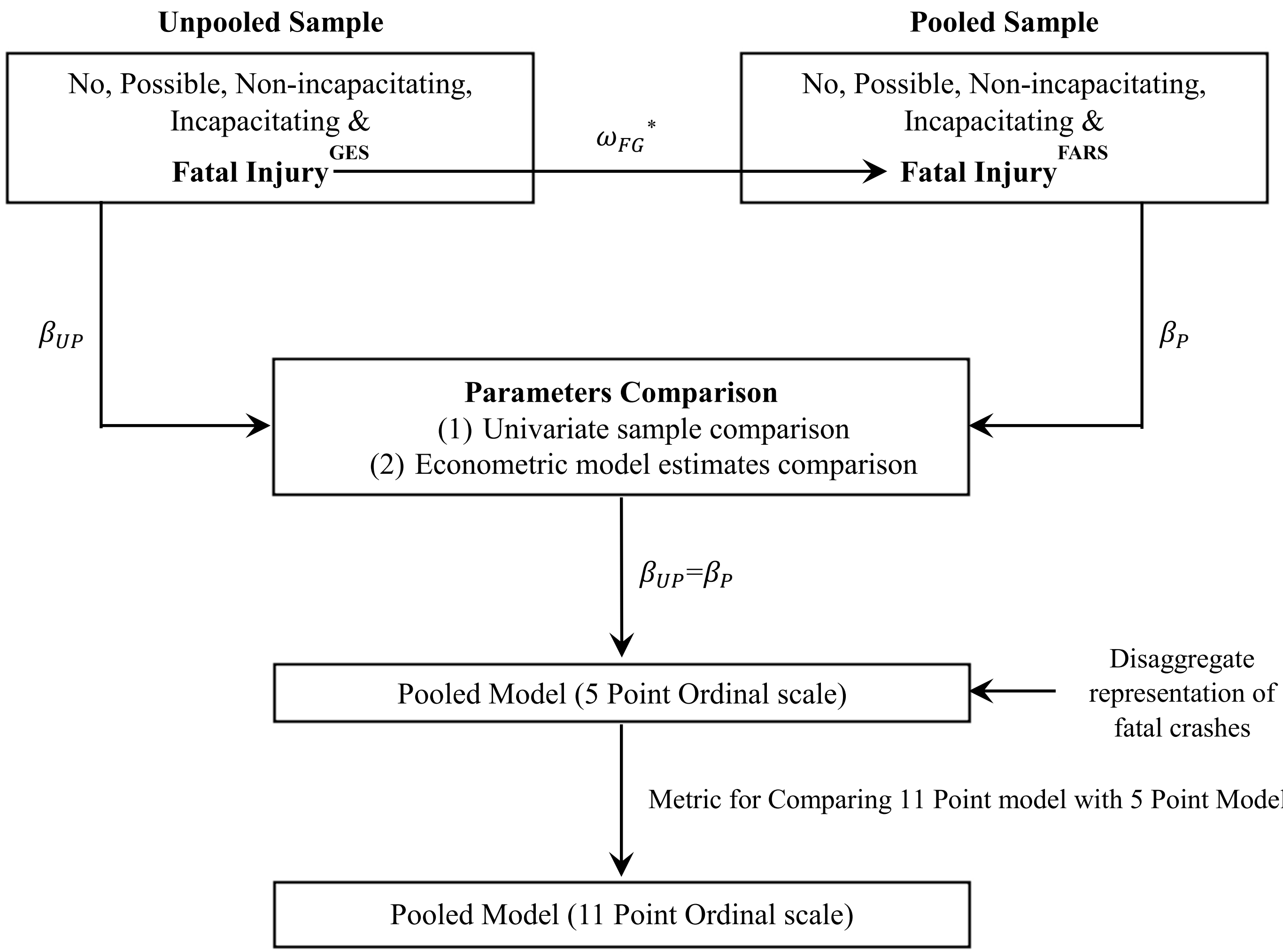
The focus of this chapter was to develop a framework for pooling of data from Fatality Analysis Reporting System (FARS) and Generalized Estimates System (GES) data. The current chapter makes three important contributions to literature on vehicle occupant injury severity analysis. First, we developed and tested a simple approach to combine information from FARS and GES databases toward a pooled database that brings together the strengths of individual databases.

Second, after demonstrating the validity of the approach, the pooled data set was employed to undertake injury severity analysis with a very refined characterization of fatality along with other injury severity levels. Specifically, a Generalized Ordered Logit model (also referred to as Partial Proportional Odds model) was estimated on an eleven-alternative ordinal categorization of injury severity. Finally, using the empirical model results, we identified important factors affecting vehicle occupant severity levels by evaluating elasticities of a selected set of exogenous variables.

The empirical analysis involved the validation of the five point ordinal (no injury, possible injury, non-incapacitating injury, incapacitating injury and fatal injury) pooled sample against the original GES sample (un-pooled sample) through two methods: (1) univariate sample comparison and (2) econometric model estimate comparison. The validation exercise confirmed that there was no evidence to suggest that the data pooled from GES and FARS resulted from distinct latent data generation process than the GES sample - the severity parameter estimates obtained using the pooled data closely resembled the severity parameter estimates obtained using the un-pooled GES data. After we confirmed that the differences in parameter estimates obtained using pooled and un-pooled data from the five point ordinal models were within the acceptable margins, we employed the pooled data to estimate models of fine resolution of injury severity with an eleven point ordinal scale defined as: no injury, possible injury, non-incapacitating injury, incapacitating injury, died between 6th-30 days of crash, died between 2nd-5 days of crash, died between 7th-24 hours of crash, died between 2nd-6 hours of crash, died between 31st-60 minutes of crash, died between 1st-30 minutes of crash and died instantly. To compare the model with the five-alternative model estimated using the un-pooled data, we generated an equivalent eleven alternative log-likelihood based on the five alternative model. The consistent improvement offered by the model estimated using the pooled data clearly indicated that inclusion of multiple discrete states of fatal injury category improves the model fit and provides more information in examining the injury severity outcome.

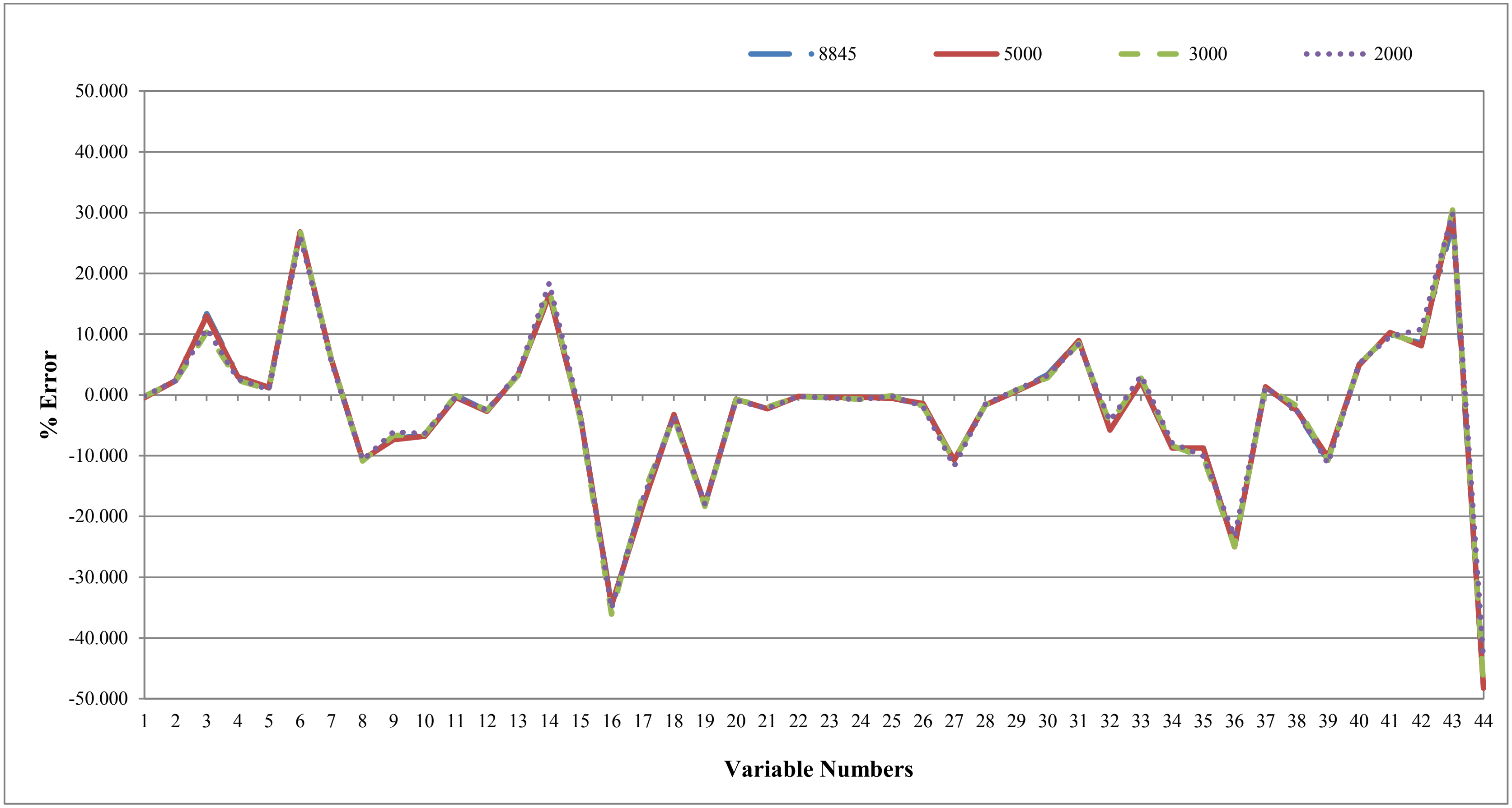
In this research, to further understand the impact of various exogenous factors, elasticity effects were estimated for the seven fatal crash injury severity categories. The elasticity effects indicated that there were considerable differences in the elasticity effects across different fatal crash categories, which signify the importance of considering the fine resolution of fatal crashes in examining the crash injury severity outcome. The most important variables in terms of early death were collision on the high speed limit road, head-on collision and driving under the influence of alcohol. In terms of most important factor affecting late MVC death (non-instantaneous) was collision during late night. In summary, the pooling of fatal crashes with high resolution information from FARS dataset and replacing the fatal crashes in GES data allowed us to examine the impact of various attributes on all levels of injury severity and in turn allowed us to draw on the strengths of FARS and GES datasets to generate a single, potentially more beneficial sample for analysis. Finally, through the elasticity exercise, we demonstrated how our approach can be employed to identify factors affecting potentially fatal crashes (non-instantaneous) and improving the chances of survival of motor vehicle occupants involved.

Figure 6.1: % Flow Chart Showing Research Framework



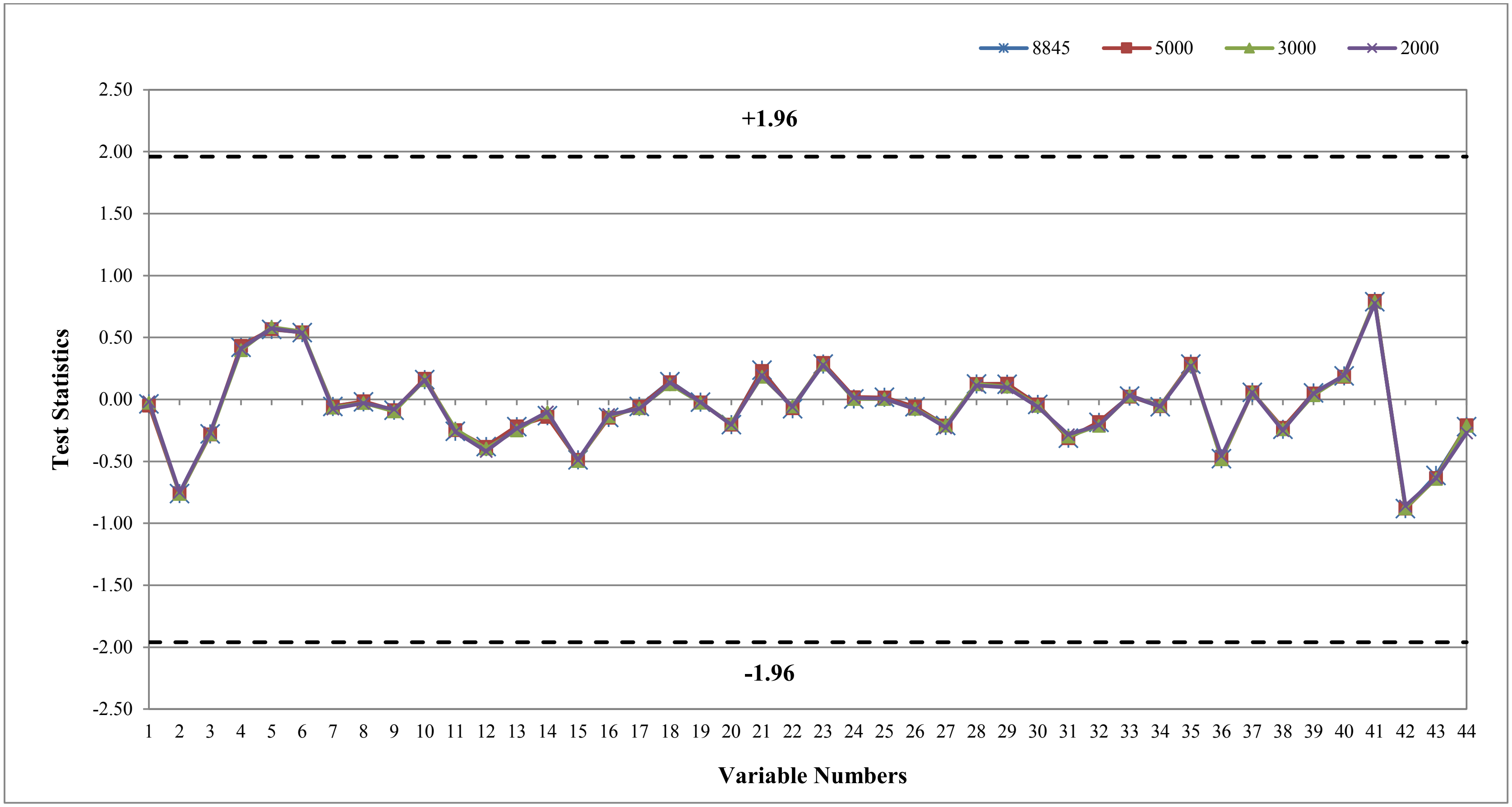
*Specific weights for FARS crash records in pooled dataset

Figure 6.2: % Error in Parameter Estimates obtained using Pooled Model Plotted against Variable Numbers



Note - Pooled data is obtained by replacing 59 fatality records from GES with 8,845 records from the FARS data for the 8,845 sample. The same process is applied to other sample sizes.

Figure 6.3: Test Statistics for Parameter Estimates Plotted against Variable Numbers



Note - Pooled data is obtained by replacing 59 fatality records from GES with 8,845 records from the FARS data for the 8,845 sample. The same process is applied to other sample sizes.

Table 6.1: Fatal Cases and Weight of Data Samples

Datasets	Samples	Fatal Cases	Weight
Un-pooled	---	59	---
Pooled	1	1956	59/1956
	2	1945	59/1945
	3	2010	59/2010
	4	1921	59/1921
	5	1983	59/1983
	6	2967	59/2967
	7	3101	59/3101
	8	3062	59/3062
	9	2980	59/2980
	10	2983	59/2983
	11	4976	59/4976
	12	4939	59/4939
	13	4921	59/4921
	14	4931	59/4931
	15	5004	59/5004
	16	8845	59/8845

Table 6.2: Sample Characteristics of "Driver Injury Severity"

Variables	Sample: Un-pooled Data	Sample: Pooled Data (With Weight)	Fatal Crashes: Un-pooled Data	Fatal Crashes: Pooled Data (With Weight)
	Frequency			
Driver Characteristics				
Driver gender (Base: Male)				
Female	2786	2786	18	18
Driver age (Base: Age 25 to 64)				
Age less than 25	1671	1666	19	14
Age above 65 & above	514	514	11	11
Restraint system use (Base: Restrained)				
Unrestrained	230	233	28	31
Under the influence of alcohol	312	325	11	24
Other physical impairment	195	197	6	8
Vehicle Characteristics				
Vehicle Type (Base: SUV, Passenger car)				
Pickup	1010	1019	4	13
Vans	413	414	2	3
Vehicle age (Base: Vehicle age ≤ 5 years)				
Vehicle age 6-10 years	2077	2068	28	19
Vehicle age ≥ 11 years	1897	1903	21	27
Roadway Design and Operational Attributes				
Speed limit (Base: Speed limit less than 26 mph)				
Speed limit 26-50 mph	3948	3940	32	24
Speed limit>50mph	1445	1452	25	32
Traffic Control Device				

Other traffic control device	145	148	1	4
Type of intersection				
T intersection	729	731	2	4
Traffic-way description				
Two way-with median	1398	1387	17	6
Environmental Factor				
Time of Day (Base: 6.00 a.m. to 11.59 p.m.)				
Late night (12.00 a.m. to 5.59 a.m.)	473	472	16	15
Surface condition				
Snowy	262	263	1	2
Crash Characteristics				
Collision object (Base: Another moving vehicle)				
Collision with large stationary object	525	517	25	17
Collision with other object	205	206	0	1
Manner of collision (Base: Angular collision)				
Head-on	347	346	10	9
Side swipe-same direction	342	342	1	1
Front to rear	1858	1858	1	1
Collision location (Base: Non-intersection)				
Driveway access	625	626	0	1
Intersection	2641	2646	6	11
Trajectory of vehicle's motions (Base: Going straight)				
Stopped in Traffic Lane	584	583	1	0
Turning right	155	155	0	0
Turning Left	680	683	0	3
Negotiating a curve	318	325	10	17

Table 6.3: Log-likelihood Values for Equivalent and Actual Eleven Point Ordinal "Driver Injury Severity" Models

Samples	Average log-likelihood		Log-likelihood Ratio Test Statistic
	Equivalent eleven point ordinal model	Actual eleven point ordinal model	
2000 (5 random samples)	-5976.897	-5964.741	24.312
3000 (5 random samples)	-5977.325	-5965.546	23.558
5000 (5 random samples)	-5977.418	-5965.788	23.260
8845 (1 sample)	-5977.790	-5966.164	23.252

Note - Pooled data is obtained by replacing 59 fatality records from GES with 8,845 records from the FARS data for the 8,845 sample. The same process is applied to other sample sizes.

Table 6.4: Estimation Results of "Driver Injury Severity" by using Un-pooled (GES) Data Sample

Variables	Latent Propensity	τ_2	τ_3	τ_4
Constant	1.253[1*] (0.122) ǂ	-0.534[28] (0.071)	0.166[37] (0.038)	0.971[41] (0.075)
Driver Characteristics				
Driver gender (Base: Male)				
Female	0.537 (0.060)[2]	0.245 (0.066)[29]	– –	– –
Driver age (Base: Age 25 to 64)				
Age less than 25	-0.117 (0.064)[3]	– –	– –	– –
Age 65 & above	0.237 (0.101)[4]	-0.213 (0.121)[30]	– –	-0.471 (0.178)[42]
Restraint system use (Base: Restrained)				
Unrestrained	1.617 (0.142)[5]	– –	– –	-0.361 (0.132)[43]
Under the influence of alcohol	0.570 (0.131)[6]	– –	– –	– –
Other physical impairment	0.708 (0.140)[7]	– –	– –	– –
Vehicle Characteristics				
Vehicle Type (Base: SUV, Passenger car)				
Pickup	-0.423 (0.081)[8]	– –	– –	– –
Vans	-0.239 (0.118)[9]	– –	– –	– –
Vehicle age (Base: Vehicle age ≤ 5 years)				
Vehicle age 6-10 years	0.315 (0.069)[10]	0.218 (0.065)[31]	– –	– –
Vehicle age ≥ 11 years	0.328 (0.071)[11]	– –	– –	– –
Roadway Design and Operational Attributes				
Speed limit (Base: Speed limit less than 26 mph)				
Speed limit 26-50 mph	0.642 (0.102)[12]	– –	– –	– –
Speed limit>50mph	0.896 (0.117)[13]	– –	– –	– –
Traffic Control Device				
Other traffic control device	0.465 (0.185)[14]	– –	– –	– –
Type of intersection				
T intersection	-0.205 (0.088)[15]	– –	– –	– –
Traffic way description				
Two way-with median	0.138 (0.069)[16]	– –	– –	– –
Environmental Factor				
Time of Day (Base: 6.00 a.m. to 11.59 p.m.)				

Late night (12.00 a.m. to 5.59 a.m.)	0.281 (0.107)[17]	– –	– –	– –
Surface condition				
Snowy	-1.040 (0.153)[18]	– –	– –	– –
Crash Characteristics				
Collision object (Base: Another moving vehicle)				
Collision with large stationary object	0.379 (0.106)[19]	-0.289 (0.129)[32]	– –	– –
Collision with other object	-1.827 (0.217)[20]	-0.637 (0.349)[33]	– –	– –
Manner of collision (Base: Angular collision)				
Head-on	0.669 (0.106)[21]	– –	-0.234 (0.124)[38]	– –
Side swipe-same direction	-1.603 (0.157)[22]	– –	– –	– –
Front to rear	-1.159 (0.079)[23]	– –	– –	– –
Collision location (Base: Non-intersection)				
Driveway access	-0.440 (0.108)[24]	0.252 (0.119)[34]	0.303 (0.132)[39]	– –
Intersection	– –	0.266 (0.071)[35]	– –	0.515 (0.133)[44]
Trajectory of vehicle's motions (Base: Going straight)				
Stopped in Traffic Lane	0.324 (0.111)[25]	– –	0.529 (0.139)[40]	– –
Turning right	-0.991 (0.241)[26]	– –	– –	– –
Turning Left	-0.188 (0.094)[27]	– –	– –	– –
Negotiating a curve	– –	0.266 (0.120)[36]	– –	– –

τ_2 = Threshold between possible injury/non-incapacitating injury; τ_3= Threshold between non-incapacitating injury/incapacitating injury; τ_4= Threshold between incapacitating injury/fatal injury

ǂ Standard errors are presented in parenthesis

* Variable Numbers

Table 6.5: Estimation Results of “Driver Injury Severity” by using the Final Pooled Data Sample

Variables	Latent Propensity	τ_2	τ_3	τ_4	τ_5	τ_6	τ_7	τ_8	τ_9	τ_{10}
Constant	1.260 (0.120) ǂ	-0.493 (0.076)	0.064 (0.052)	0.941 (0.187)	-2.637 (3.692)	-2.808 (4.165)	-2.854 (4.383)	-1.19 (2.000)	-1.301 (2.414)	-0.584 (2.053)
Driver Characteristics										
Driver gender (Base: Male)										
Female	0.548 (0.061)	0.224 (0.066)	– –	–	–	–	–	–	–	–
Driver age (Base: Age 25 to 64)										
Age less than 25	-0.135 (0.065)			–	–	–	–	–	–	–
Age above 65 & above	0.255 (0.107)	-0.239 (0.122)	– –	–	–	–	–	–	–	–
Restraint system use (Base: Restrained)										
Unrestrained	1.752 (0.138)	– –	– –	–	–	–	–	–	–	–
Under the influence of alcohol	0.635 (0.160)	-0.309 (0.165)	– –	–	–	–	–	–	–	–
Other physical impairment	0.757 (0.143)	– –	– –	–	–	–	–	–	–	–
Vehicle Characteristics										
Vehicle Type (Base: SUV, Passenger car)										
Pickup	-0.379 (0.085)	– –	– –	–	–	–	–	–	–	–
Vans	-0.224 (0.114)	– –	– –	–	–	–	–	–	–	–
Vehicle age (Base: Vehicle age ≤ 5 years)										
Vehicle age 6-10 years	0.294 (0.071)	0.237 (0.065)	– –	–	–	–	–	–	–	–
Vehicle age ≥ 11 years	0.326 (0.070)	– –	– –	–	–	–	–	–	–	–
Roadway Design and Operational Attributes										
Speed limit (Base: Speed limit less than 26 mph)										
Speed limit 26-50 mph	0.641 (0.101)	– –	– –	–	–	–	–	–	–	–
Speed limit>50mph	0.965 (0.115)	– –	– –	–	–	–	–	–	–	–
Traffic Control Device										
Other traffic control device	0.569 (0.161)	– –	– –	–	–	–	–	–	–	–
Type of intersection										
T intersection	-0.203 (0.092)	– –	– –	–	–	–	–	–	–	–
Environmental Factor										
Time of Day (Base: 6.00 a.m. to 11.59 p.m.)										

Late night (12.00 a.m. to 5.59 a.m.)	0.238 (0.105)	– –	– –	–	–	–	–	–	–	–
Surface condition										
Snowy	-1.023 (0.152)	– –	– –	–	–	–	–	–	–	–
Crash Characteristics										
Collision object (Base: Another moving vehicle)										
Collision with large stationary object	0.386 (0.108)	-0.290 (0.131)	0.248 (0.091)	–	–	–	–	–	–	–
Collision with other object	-1.815 (0.211)	-0.650 (0.354)	– –	–	–	–	–	–	–	–
Manner of collision (Base: Angular collision)										
Head-on	0.693 (0.112)	– –	– –	–	–	–	–	–	–	–
Side swipe-same direction	-1.578 (0.168)	– –	– –	–	–	–	–	–	–	–
Front to rear	-1.125 (0.081)	– –	0.260 (0.106)	–	–	–	–	–	–	–
Collision location (Base: Non-intersection)										
Driveway access	-0.444 (0.102)	0.215 (0.115)	0.336 (0.145)	–	–	–	–	–	–	–
Intersection	– –	0.243 (0.072)	– –	–	–	–	–	–	–	–
Trajectory of vehicle's motions (Base: Going straight)										
Stopped in Traffic Lane	0.313 (0.115)	– –	0.410 (0.171)	–	–	–	–	–	–	–
Turning right	-0.965 (0.229)	– –	– –	–	–	–	–	–	–	–
Turning Left	-0.167 (0.090)	– –	– –	–	–	–	–	–	–	–
Negotiating a curve	– –	0.235 (0.133)	– –	–	–	–	–	–	–	–

τ_2 = Threshold between possible injury/non-incapacitating injury; τ_3= Threshold between non-incapacitating injury/incapacitating injury; τ_4= Threshold between incapacitating injury/6to30 days; τ_5= Threshold between 6to30 days/ 1 to 5 days; τ_6= Threshold between 1 to 5 days/ 7 to 24 hours; τ_7= Threshold between 7 to 24 hours/ 1 to 6 hours; τ_8 = Threshold between 1 to 6 hours/ 31 to 60 minutes; τ_9 = Threshold between 31 to 60 minutes/ 1 to 30 minutes; τ_{10} = Threshold between 1 to 30 minutes/ Died Instantly

ǂ Standard errors are presented in parenthesis

Table 6.6: Elasticity Effects

Variables	Died between 6th-30 days	Died between 2nd-5 days	Died between 7th-24 hours	Died between 2nd-6 hours	Died between 31st-60 minutes	Died between 1st-30 minutes	Died instantly
Female driver	31.816	32.347	32.812	34.120	36.099	38.344	42.184
Driver age 65 & above	32.902	33.388	33.814	35.021	36.864	38.993	42.755
Under the influence of alcohol	70.863	71.557	72.160	73.831	76.307	79.042	83.577
Vehicle age ≥ 11 years	25.424	25.756	26.045	26.854	28.067	29.426	31.699
Speed limit>50mph	77.550	78.595	79.511	82.110	86.081	90.663	98.641
Late night (12.00 a.m. to 5.59 a.m.)	19.037	19.287	19.505	20.112	21.018	22.029	23.716
Collision with large stationary object	20.272	20.523	20.743	21.358	22.284	23.328	25.108
Head-on collision	61.050	62.168	63.152	65.962	70.294	75.355	84.346

CHAPTER 7 Conclusions and Directions for Future Research

7.1 Introduction

The objective of the dissertation is to develop advanced econometric frameworks to address methodological gaps in safety literature while employing these models developed to study important empirical issues. Specifically, the primary focus of the current research is on advancing the state of the art in modeling crash injury severity for drivers. The dissertation aims to formulate and implement several methodological developments in examining the crash injury severity outcome for drivers of passenger vehicles. This road user group represents approximately 50% fatalities among all road user groups in high-income countries of the world, and hence, the proposed research endeavours to identify the various factors that affect the crash severity to assist policy makers in developing appropriate remedial measures.

The current dissertation contributes substantially towards methodological gaps in the state of the art for driver injury severity analysis along six directions: (1) appropriate model framework, (2) underreporting issue in severity analysis, (3) exogenous factor homogeneity assumption (4) multiple dependent variables in severity analysis, (5) continuum of fatal crashes and (6) data pooling from multiple data sources. In addition to making the aforementioned methodological contributions, the dissertation also makes a substantial empirical contribution to the existing safety literature. In this chapter major conclusions from the model frameworks presented earlier are summarized. The rest of the chapter is organized as follows. Sections 7.2 through 7.6 discuss the findings of each chapter briefly while also present the methodological and empirical contributions of the dissertation. Section 7.7 concludes the dissertation by presenting the directions for future research.

7.2 Evaluating Alternate Discrete Outcome Frameworks for Modeling Crash Injury Severity

In Chapter 2, a comprehensive empirical comparison of the ordered and unordered outcome models was presented to identify the more relevant framework to model crash injury severity. The alternative modeling approaches considered for the exercise included: for the ordered outcome framework - ordered logit (OL), generalized ordered logit (GOL), mixed generalized ordered logit (MGOL) and for the unordered outcome framework - multinomial logit (MNL), nested logit (NL), ordered generalized extreme value logit (OGEV) and mixed multinomial logit (MMNL) model. The empirical analysis was based on the 2010 General Estimates System (GES) database. The performance of the alternative frameworks were examined in the context of model estimation and validation (at the aggregate and disaggregate level). Further, the performance of the model frameworks in the presence of underreporting was explored – with and without corrections to the estimates. Important findings and the implications from the research effort of Chapter 2 include:

- The comparison exercise clearly highlighted the superiority of the MGOL in terms of data fit compared to MMNL model.
- In the context of underreporting, both MMNL and MGOL performed almost equivalently and the performance for both of these structures were improved with the correction measure.
- The host of validation statistics confirmed that neither the ordered nor the unordered frameworks exclusively outperforms each other both at the aggregate and the disaggregate levels. The results of validation also indicated that MGOL and MMNL offer very similar prediction for the various sub-samples at the aggregate and disaggregate level.
- Overall, the results of the empirical comparison provided credence to the belief that an ordered system that allows for exogenous variable effects to vary across alternatives and accommodates unobserved heterogeneity offer almost equivalent results to that of the corresponding unordered systems in the context of driver injury severity.
- Implications: The results presented in Chapter 2 have significant implications for safety research. There is growing recognition in the safety community that modeling injury severity as exogenous to seat belt use, alcohol consumption, or collision type is not realistic. For instance, the common unobserved factors that influence seat belt usage might also influence injury severity (see Eluru and Bhat, 2007). Incorporating such interactions in a joint framework increases the complexity of the models involved. However, by allowing for injury severity to follow an ordered response structure we can reduce the complexity of the joint model because of the single error term of this structure. The unordered model would lead to a more cumbersome modeling approach because of the multiple error terms involved (Eluru, 2013). Recent research has demonstrated the advantages of such joint frameworks (see for example Castro et al., 2013; Narayanmoorthy et al., 2012).

7.3 A Latent Segmentation based Generalized Ordered Logit Model – Heterogeneity in Driver Injury Severity Modeling

Chapter 3 proposed an econometric model for examining driver injury severity that accommodates systematic heterogeneity based on crash characteristics and relaxes the constant threshold assumption of traditional ordered logit model. The data for this research effort was drawn from the Victoria crash database from Australia for the years 2006 through 2010. The empirical analysis involved the estimation of models using six different statistical frameworks: 1) OL, 2) GOL, 3) LSOL with two segments, 4) LSOL with three segments, 5) LSGOL with two segments and 6) LSGOL with three segments. The comparison exercise, based on information criterion metrics, highlighted the superiority of the LSGOL model with two segments on the estimation sample in terms of data fit compared to the other ordered outcome models. Other important results from Chapter 3 includes:

- In the LSGOL approach, drivers were assigned probabilistically to two segments – high risk segment and low risk segment - based on a host of crash characteristics.
- The crash characteristics that affected the allocation of drivers into segments included: collision object, trajectory of vehicle's motion and manner of collision.
- The elasticity effects estimate indicated that the most significant variables in terms of increase in serious/fatal injury (from both models) for drivers were driver age above 65, driver ejection, not wearing seat belts, and collision in high speed zone. In terms of serious/fatal injury reduction, the important factors were presence of pedestrian control, presence of roundabout, driving a panel van, unpaved road condition and presence of passengers.
- The predictive performance evaluation of the estimated models on a validation sample revealed that the LSGOL model represented superior performance compared to other estimated models.
- In summary, the comparison exercise supported the hypothesis that LSGOL is a promising ordered response framework for accommodating population heterogeneity and for relaxing the fixed threshold assumption in the context of driver injury severity.

7.4 Examining Driver Injury Severity in Two Vehicle Crashes – A Copula Based Approach

The focus of Chapter 4 was to jointly model the collision type and injury severity outcome of drivers involved in a two vehicle collisions using a copula-based joint multinomial logit-ordered logit modeling framework. Our study also accommodated the potential heterogeneity (across drivers) in the dependency effect of collision type and injury severity outcome within a closed form copula framework. It examined collision type as a vehicle level variable using a combination of collision type and the initial point of contact. The proposed model was estimated using driver injury severity data for two vehicle crashes from the state of Victoria, Australia for the year 2006 through 2010. The empirical analysis involved estimation of models by using six different copula structures: 1) Gaussian, 2) FGM, 3) Clayton, 4) Gumbel, 5) Frank and 6) Joe. Results of the estimated copula models revealed a significant level of dependency between collision type and injury severity outcome and thus confirmed the importance of accommodating dependence between collision type and injury severity outcome in the analysis of driver injury severity. Other important findings and implications from the empirical analysis include:

- The best model fit was obtained for a combination model of Frank-Clayton copulas (Frank copula structure for rear-ender and head-on collision and Clayton dependency structure with the remaining collision type).
- In model estimation, except for far-angular collision type, all other copula dependencies were characterized by at least one additional exogenous variable. This provided support to our hypothesis that the dependency structures were not constant across the entire database.
- The results suggested that the impact of exogenous variables varied (for some variables) in magnitude as well as in sign across collision types. The impacts of these variables were also substantially different from the estimates of independent MNL-OL model.
- The elasticity effect estimates for independent variables clearly highlighted that each collision type has a fundamentally distinct injury severity profile underscoring the importance of examining the effect of various exogenous variables on driver injury severity outcome by different collision types.
- The validation experiments also revealed the enhanced performance of copula based model compared to independent model.

- Implications: The findings of Chapter 4 provided a more complete picture of injury severity profile associated with different collision type in the context of driver injury severity, thus target based countermeasures could be devised from such an approach to address the entire profile of collision mechanism.

7.5 Analyzing the Continuum of Fatal Crashes: A Generalized Ordered Approach

The focus of Chapter 5 of the dissertation was to identify the associated risk factors of driver fatalities while recognizing that fatality is not a single state but rather is made up of a timeline between dying instantly to dying within thirty days by using Fatality Analysis Reporting System (FARS) database for the year 2010. This research attempt introduced in this dissertation is a first attempt to analyze the fatal injury from a new perspective and examined fatality as a continuous spectrum based on survival time ranging from dying within thirty days of crash to dying instantly. Moreover, the correction for endogeneity bias was pinned down in the current study context by employing a two-stage residual inclusion (2SRI) approach. In the research effort, we estimated six different models: 1) ordered logit (OL), 2) generalized ordered logit (GOL), 3) mixed generalized ordered logit (MGOL), 4) OL with the 2SRI treatment, 5) GOL with the 2SRI treatment and 6) MGOL model with the 2SRI treatment while employing a comprehensive set of exogenous variables (driver characteristics, vehicle characteristics, roadway design and operational attributes, environmental factors, crash characteristics and situational variables). The comparison exercise highlighted the superiority of the MGOL model with the 2SRI treatment on the sample in terms of data fit compared to the other ordered outcome models in the current study context.

- The factors that contributed to an increase in the likelihood of early death include: alcohol impairment, previous record of other harmful motor vehicle convictions, medium and higher speed limit, presence of stop sign, presence of other traffic control device, cloudy weather, head-on crashes, collision at non-intersection locations, driver ejection, presence of more passengers and longer EMS response time.
- The factors that contributed to a decrease in the likelihood of early death include: young driver, previous record of license suspension and revocation, crashes during off-peak and evening peak periods and front-to-rear crashes.

- We found that after controlling for endogeneity, the coefficient on the logarithm of EMS response time was intuitive and statistically significant indicating that EMS response time is correlated with unobserved determinants generating endogeneity in the outcome model of the time to death of drivers.
- The elasticity effects indicated that there were considerable differences in the elasticity effects across different fatal crash categories, suggesting that fatality is not a single state but rather is made up of multiple discrete states from dying instantly to dying within thirty days of crash.
- The most significant variables in terms of lower survival probability for drivers were crashes on high speed limit roads, crashes on medium speed limit roads and head-on crashes. In terms of longer survival probability, the important factors were old driver, front-to-rear crash and crashes during off-peak period.
- The elasticity analysis assisted in providing a clear picture of attribute impact on driver time-to-death variables.
- Implications: The variable effects identified in Chapter 5 have important implications in terms of enforcement, engineering and educational strategies. In terms of engineering measures, a forgiving road environment should be designed for a high and medium speed limit road location to allow the drivers more space to recover from a driving error. Moreover, policies concerning enforcement for reducing traffic violations have the potential to reduce head-on crashes.

7.6 Pooling Data from Fatality Analysis Reporting System (FARS) and Generalized Estimates System (GES) to Explore the Continuum of Injury Severity Spectrum

The focus of Chapter 6 was to develop a framework for pooling of data from Fatality Analysis Reporting System (FARS) and Generalized Estimates System (GES) data for the year 2010. In this chapter, we developed and tested a simple approach to combine information from FARS and GES databases toward a pooled database that brings together the strengths of individual databases. Further, the pooled data set was employed to undertake injury severity analysis with a very refined characterization of fatality along with other injury severity levels by employing a Generalized Ordered Logit model. Important findings and implications from the empirical analysis include:

- The validation exercise confirmed that the severity parameter estimates obtained using the pooled data closely resembled the severity parameter estimates obtained using the un-pooled GES data.
- The empirical analysis indicated that inclusion of multiple discrete states of fatal injury category improves the model fit and provides more information in examining the injury severity outcome.
- These results also suggest that considering a fine resolution categorization of fatal crashes in examining the crash injury severity outcome offers the potential to provide useful information for policy makers in developing the EMS system and trauma triage.
- Implications: The empirical effects identified in Chapter 6 have important implications in terms of enforcement, engineering and educational strategies. In terms of engineering measures, a forgiving road environment should be designed for a high speed limit road location to allow the drivers more space to recover from a driving error. Head-on collisions are often caused by drivers violating traffic rules, driving across the centerline, driving too fast for the roadway conditions and thus by losing control of their vehicles (Zhang and Ivan, 2005). Therefore, policies concerning the enforcement in reducing the traffic violation have the potential to reduce this type of collision. With respect to enforcement and education, our results endorse a continuous education program and stricter enforcement to prevent drunk-driving. Mortality rate for the alcohol impaired drivers is observed to be twice compared to the sober drivers (Stübig et al., 2012) in the event of road traffic crash. Besides, research suggests that alcohol-impaired drivers not only put themselves at a risk of high injury severity but also risk others involved in the crash to a high injury severity (Rana et al., 2010). Therefore, public health effort and education campaigns against alcohol-intoxicated driving are needed for this group of drivers.

7.7 Directions for Future Research

The summary of findings and the contributions of the dissertation in examining driver injury severity are discussed in the preceding sections of this chapter. In the following sections, the limitations of the research efforts are discussed while also present potential research extensions for future.

7.7.1 Limitations and Potential Extensions

Limitations: The dissertation is not without limitations. The models estimated in our research efforts are based on data from particular region or countries. Hence, it is not possible to generalize the findings for other region or countries based on the findings of our studies. The wide variety of the data inventory would allow us to identify important determinants of safety for developed countries. Another aspect related to data is that we have used cross-sectional crash databases comprising of either one or multiple years of data. However, in our analyses, we have not considered differences across the different years. It will be an interesting exercise to model the impact of temporal effects on driver injury severity models.

In our research effort, we categorized the spectrum of fatal crashes in seven refined categories of fatalities ranging from fatality after thirty days to instant death, specifically in Chapters 5 and 6. However, some of the earlier studies (Trunkey, 1983) argued that the distribution of survival times after traffic crash is "trimodal". There are also studies (Clark et al., 2012) that contradict the trimodal distribution of survival time after crash. Thus, it will be an interesting exercise to identify the discrete categories or distribution of fatal crashes in examining the impact of exogenous variable on fatality spectrum.

For examining the spectrum of fatal crashes we have considered FARS database of the US since it reports the exact timeline of the fatal occurrence within thirty days from the time to crash. However, to our knowledge, no other police reported crash database includes such information of fatal crashes. Thus, it would not be possible to examine the continuum of fatal crashes by using other available police reported crash databases. In this regards, it might be interesting to explore other source of crash databases, for instance: hospital reported crash data, to examine driver fatalities at a disaggregate level of fatal crashes.

Potential Extensions: The major focus of current dissertation is to contribute to the state of the art for modeling driver injury severity analysis along six directions as discussed earlier. However, there are many other methodological issues (for instance spatial and temporal correlations, accommodating soft psychometric measures) in safety literature that requires further investigations as highlighted recently in the article by Mannering and Bhat (2014). While literature in severity analysis is vast and growing, future research efforts should also continue to address the methodological gaps in analyzing crash injury severity outcome for different road user groups.

In term of methodology, safety research would benefit from incorporating temporal effects within the econometric frameworks developed in this dissertation. Latent segmentation based model incorporates effects of both observed and unobserved heterogeneity while assuming that these impacts remain constant over time. But the effect of observed exogenous variable may vary based on a latent time-dependent state variable. More recently, safety researchers have employed Markov switching technique in capturing unobserved heterogeneity across time period in the context of crash injury severity analysis (Malyshkina and Mannering, 2009; Xiong et al., 2014). Such an approach allows for two time-varying unobserved states of roadway safety. Within this framework, the impacts of control variables are not only allowed to vary but also allowed to interact and change from one state to another over time. Thus, developing a Markov switching latent segmentation based model would add more flexibility towards capturing potential heterogeneity in the pursuits of estimating more generalized crash injury severity models.

From an empirical standpoint, the road safety literature would benefit from exploring a wide spectra of driver attributes, specifically driver behaviour and driving record, in examining crash injury severity outcomes. In terms of driver behaviour, inappropriate driving actions are extensively identified in the literature as a major contributor of traffic collisions. For instance, using the Driver Behaviour Questionnaire, several studies (Elliott et al., 2007; Parker et al., 1995) found that violations and errors were significantly related to crashes. In fact, driver actions are identified to be a contributing factor in more than 90% of road crashes (Rumar, 1985). However, the conventional police/hospital reported crash databases may not include relevant behavioural, physiological and psychological characteristics of individual involved in traffic crashes. If data on insurance claims could potentially be used to augment the traditional police reported crash database the modeling exercise can be substantially enhanced. The driving history (for example: previous crash records, previous speeding or impairment records, previous license convictions, previous traffic law violations) as reported in insurance reports can be used as a "proxy" for driver behaviour. Thus, it would be interesting to examine crash injury severity outcomes by using matched police reported crash databases and insurance reports of the drivers.

Future research efforts should also consider and investigate the opportunities to examine crash-frequency and crash-severity models jointly. In road safety literature, crash-frequency and crash-severity models are mostly examined separately. However, few studies (see Savolainen et al., 2011 for a detailed discussion) have examined frequency of different crash severity levels simultaneously mostly by using non-crash-specific data. A two-stage model has also been proposed recently for examining the crash frequency-severity process jointly (Wang et al., 2011). Another possible alternate method to tie these two different approaches is to integrate the outcome of crash-frequency model, specifically macroscopic model, in examining the crash severity level outcomes. The outcome of a macroscopic crash-frequency model can potentially be used for network screening and for ranking different planning zones. Further, it would be possible to explore the relationship between the total number of crashes and the subsequent severity outcomes of a particular zone by integrating the rank variable as an independent variable in the crash-severity model. By using the specifications of such an integrated crash frequency-severity model, it would be possible to forecast the change in crash frequency and severity levels from implementing a specific countermeasure at planning level. The derived information will also benefit the emergency medical service resources.

REFERENCES

Abay, K. A., R. Paleti, and C. R. Bhat. The Joint Analysis of Injury Severity of Drivers in Two-Vehicle Crashes Accommodating Seat Belt Use Endogeneity. Transportation Research Part B: Methodological, Vol. 50, 2013, pp. 74-89.

Abdel-Aty, M. A., and H. T. Abdelwahab. Configuration Analysis of Two-Vehicle Rear-End Crashes. Transportation Research Record, No. 1840, 2003, pp. 140-147.

Abdel-Aty, M. Analysis of Driver Injury Severity Levels at Multiple Locations Using Ordered Probit Models, Journal of Safety Research, Vol. 34, No. 5, 2003, pp. 597-603.

Alexander, J., P. Barham, and I. Black. Factors Influencing the Probability of an Incident at a Junction: Results from an Interactive Driving Simulator. Accident Analysis and Prevention, Vol. 34, No. 6, 2002, pp. 779-92.

Al-Ghamdi, A. Using Logistic Regression to Estimate the Influence of Accident Factors on Accident Severity. Accident Analysis and Prevention, Vol. 34, No. 6, 2002, pp. 729-741.

Arditi, D., D. E. Lee, and G. Polat. Fatal Accidents in Nighttime Vs. Daytime Highway Construction Work Zones. Journal of Safety Research, Vol. 38, No. 4, 2007, pp. 399-405.

Arnedt, J., G. Wilde, P. Munt, and A. MacLean. How do Prolonged Wakefulness and Alcohol Compare in the Decrements They Produce on A Simulated Driving Task?. Accident Analysis and Prevention, Vol. 33, No. 3, 2001, pp. 337-344.

Awadzi, K. D., S. Classen, A. Hall, R. P. Duncan, and C. W. Garvan. Predictors of Injury among Younger and Older Adults in Fatal Motor Vehicle Crashes. Accident Analysis and Prevention, Vol. 40, No.6, 2008, pp.1804-1810.

Aziz, H. M. A., V. S. Ukkusuri, and S. Hasan. Exploring the Determinants of Pedestrian–Vehicle Crash Severity in New York City. Accident Analysis and Prevention, Vol. 50, 2013, pp. 1298-1309.

Baker, T. K., T. Falb, R. Voas, and J. Lacey. Older Women Drivers: Fatal Crashes in Good Conditions. Journal of Safety Research, Vol. 34, No. 4, 2003, pp. 399-405.

Bass, F. M., and D. R. Wittink, Pooling Issues and Methods in Regression Analysis with Examples in Marketing Research. Journal of Marketing Research, Vol. 12, No. 4, 1975, pp. 414-425.

Bédard, M., G. H. Guyatt, M. J. Stones, and J. P. Hirdes. The Independent Contribution of Driver, Crash, and Vehicle Characteristics to Driver Fatalities. Accident Analysis and Prevention, Vol. 34, No. 6, 2002, pp. 717-727.

Ben-Akiva, M. E., and R. S. Lerman. Discrete Choice Analysis: Theory and Application to Travel Demand. The MIT Press, Cambridge, 1985.

Bhat, C. R. Endogenous Segmentation Mode Choice Model with an Application to Intercity Travel. Transportation Science, Vol. 31, No. 1, 1997, pp. 34-48.

Bhat, C. R. Quasi-random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model, Transportation Research Part B: Methodological, Vol. 35, No. 7, 2001, pp. 677-693.

Bhat, C. R. The Maximum Approximate Composite Marginal Likelihood (Macml) Estimation of Multinomial Probit-Based Unordered Response Choice Models. Transportation Research Part B: Methodological, Vol. 45, No. 7, 2011, pp. 923-39.

Bhat, C. R., and N. Eluru. A Copula-Based Approach to Accommodate Residential Self-Selection Effects in Travel Behavior Modeling. Transportation Research Part B: Methodological, Vol. 43, No. 7, 2009, pp. 749-65.

Bhat, C.R., and V. Pulugurta. A Comparison of Two Alternative Behavioral Choice Mechanisms for Household Auto Ownership Decisions. Transportation Research Part B, Vol. 32, No. 1, 1998, pp. 61-75.

Blincoe, L., A. Seay, E. Zaloshnja, T. Miller, E. Romano, S. Lutcher, and R. Spicer. The Economic Impact of Motor Vehicle Crashes 2000, NHTSA Technical Report, 2002.

Braver, E. R. Race, Hispanic Origin, and Socioeconomic Status in Relation to Motor Vehicle Occupant Death Rates and Risk Factors Among Adults. Accident Analysis and Prevention, Vol. 35, No. 3, 2003, pp. 295– 309.

Brodsky, H. Delay in Ambulance Dispatch to Road Accidents. American Journal of Public Health, Vol. 82, No. 6, 1992, pp. 873-875.

Brown, L. H., A. Khanna, and R. C. Hunt. Rural Vs Urban Motor Vehicle Crash Death Rates: 20 Years of Fars Data. Prehospital Emergency Care, Vol. 4, No. 1, 2000, pp. 7-13.

Campos-Outcalt, D., C. Bay, A. Dellapena, and M. K. Cota. Motor Vehicle Crash Fatalities by Race/Ethnicity in Arizona, 1990– 96. Injury Prevention, Vol. 9, No. 3, 2003, pp. 251– 256.

Castro, M., R. Paleti, and C. R. Bhat. A Spatial Generalized Ordered Response Model to Examine Highway Crash Injury Severity. Accident Analysis and Prevention, Vol. 52, 2013, pp. 188-203.

Chalya, P. L., J. B. Mabula, R. M. Dass, N. Mbelenge, I. H. Ngayomela, A. B. Chandika, and J. M. Gilyoma. Injury Characteristics and Outcome of Road Traffic Crash Victims at Bugando Medical Centre in Northwestern Tanzania. Journal of Trauma Management and Outcomes, Vol. 6, No.1, 2012, pp. 1-8.

Chamberlain, G. Analysis of Covariance with Qualitative Data. The Review of Economic Studies, Vol. 47, No. 1, 1980, pp. 225-238.

Chang, L.-Y., and F. Mannering. Analysis of Injury Severity and Vehicle Occupancy in Truck- and Non-Truck-Involved Accidents. Accident Analysis and Prevention, Vol. 31, No. 5, 1999, pp. 579-592.

Chen, F., and S. Chen. Injury Severities of Truck Drivers in Single- and Multi-Vehicle Accidents on Rural Highways. Accident Analysis and Prevention, Vol. 43, No. 5, 2011, pp. 1677-1688.

Chen, H., L. Cao, and D. B. Logan. Analysis of Risk Factors Affecting the Severity of Intersection Crashes by Logistic Regression. Traffic Injury Prevention, Vol. 13, No. 3, 2012, pp. 300-307.

Chen, L., S. P. Baker, E. R. Braver, and G. Li. Carrying Passengers as a Risk Factor for Crashes Fatal to 16- and 17-Year-Old Drivers. Journal of American Medical Association, Vol. 283, No. 12, 2000, pp. 1578-82.

Chiou, Y. C., C. C. Hwang, C. C. Chang, and C. Fu. Modeling Two-Vehicle Crash Severity by a Bivariate Generalized Ordered Probit Approach. Accident Analysis and Prevention. Vol. 51, 2013, pp. 175-84.

Chipman, M. L. Side Impact Crashes - Factors Affecting Incidence and Severity: Review of the Literature. Traffic Injury Prevention. Vol. 5, No. 1, 2004, pp. 67-75.

Clark D. E., B. M. Cushing. Predicted Effect of Automatic Crash Notification on Traffic Mortality. Accident Analysis and Prevention, Vol. 34, No. 4, 2002, pp. 507-13.

Clark, D. E., J. Qian, K. C. Sihler, L. D. Hallagan, and R. A. Betensky. The Distribution of Survival Times after Injury. World journal of surgery, Vol. 36, No. 7, 2012, pp. 1562-1570.

Clark, D. E., R. J. Winchell, and R. A. Betensky. Estimating the Effect of Emergency Care on Early Survival after Traffic Crashes. Accident Analysis and Prevention, Vol. 60, No. 0, 2013, pp. 141-147.

Collins, L. M., P. L. Fidler, S. E. Wugalter, and J. D. Long. Goodness-of-Fit Testing for Latent Class Models. Multivariate Behavioral Research, Vol. 28, No. 3, 1993, pp. 375-89.

Conroy, C., G. T. Tominaga, S. Erwin, S. Pacyna, T. Velky, F. Kennedy, M. Sise, and R. Coimbra. The Influence of Vehicle Damage on Injury Severity of Drivers in Head-on Motor Vehicle Crashes. Accident Analysis and Prevention, Vol. 40, No. 4, 2008, pp. 1589-1594.

Cowley, R. A., F. Hudson, E. Scanlan, W. Gill, R. J. Lally, W. Long, and A. O. Kuhn. An Economical and Proved Helicopter Program for Transporting the Emergency Critically Ill and Injured Patient in Maryland. Journal of Trauma - Injury, Infection, and Critical Care, Vol. 13, No. 12, 1973, pp. 1029-38.

Das, A., and M. A. Abdel-Aty. A Combined Frequency-Severity Approach for the Analysis of Rear-End Crashes on Urban Arterials. Safety Science, Vol. 49, No. 8-9, 2011, pp. 1156-63.

de Lapparent, M. Willingness to Use Safety Belt and Levels of Injury in Car Accidents. Accident Analysis and Prevention, Vol. 40, No. 3, 2008, pp. 1023-1032.

Depaire, B., G. Wets, and K. Vanhoof. Traffic Accident Segmentation by Means of Latent Class Clustering. Accident Analysis and Prevention, Vol. 40, No. 4, 2008, pp. 1257-1266.

Dissanayake, S., and J. J. Lu. Factors Influential in Making an Injury Severity Difference to Older Drivers Involved in Fixed Object–Passenger Car Crashes. Accident Analysis and Prevention, Vol. 34, No. 5, 2002, pp. 609-618.

Edwards, J. B. The Relationship between Road Accident Severity and Recorded Weather. Journal of Safety Research, Vol. 29, No, 4, 1998, pp. 249-262.

Elliott, M.A., C.J. Armitage, and C.J. Baughan. Using the Theory of Planned Behaviour to Predict Observed Driving Behaviour. British Journal of Social Psychology, Vol. 46, No. 1, 2007, pp. 69-90.

Eluru, N. Evaluating Alternate Discrete Choice Frameworks for Modeling Ordinal Discrete Variables. Accident Analysis and Prevention, Vol. 55, 2013, pp. 1-11.

Eluru, N., and C. R. Bhat. A Joint Econometric Analysis of Seat Belt Use and Crash-Related Injury Severity. Accident Analysis and Prevention, Vol. 39, No. 5, 2007, pp.1037-1049.

Eluru, N., C. R. Bhat, and D. A. Hensher. A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes. Accident Analysis and Prevention, Vol. 40, No. 3, 2008, pp. 1033-1054.

Eluru, N., M. Bagheri, L. F. Miranda-Moreno, and L. Fu. A Latent Class Modeling Approach for Identifying Vehicle Driver Injury Severity Factors at Highway-Railway Crossings. Accident Analysis and Prevention, Vol. 47, 2012, pp. 119-127.

Eluru, N., R. Paleti, R. Pendyala, and C. Bhat. Modeling Injury Severity of Multiple Occupants of Vehicles. Transportation Research Record, No. 2165, 2010, pp. 1-11.

Elvik, R. To What Extent Can Theory Account for the Findings of Road Safety Evaluation Studies?. Accident Analysis and Prevention, Vol. 36, No. 5, pp. 2004, pp. 841-49.

Elvik, R., and A. B. Mysen. Incomplete Accident Reporting: Meta-analysis of Studies Made in 13 Countries. Transportation Research Record, No. 1665, 1999, pp. 133-140.

Evans, L Traffic Safety. Bloomfield, Michigan: Science Serving Society, 2004.

Evans, L., and M. C. Frick. Seating Position in Cars and Fatality Risk. American Journal of Public Health, Vol. 78, No. 11, 1988, pp. 1456-1458.

Fabbri, A., G. Marchesini, A. M. Morselli-Labate, F. Rossi, A. Cicognani, M. Dente, T. Iervese, et al. Positive Blood Alcohol Concentration and Road Accidents. A Prospective Study in an Italian Emergency Department. Emergency Medicine Journal, Vol. 19, No. 3, 2002, pp. 210-14.

Fatality Analysis Reporting System (FARS). Passenger Vehicle Driver Deaths by Vehicle Type, 1975-2010. National Highway Traffic Safety Administration (NHTSA), Washington, DC, 2010.

Feero, S., J. R. Hedges, E. Simmons, and L. Irwin. Does out-of-Hospital Ems Time Affect Trauma Survival?. The American Journal of Emergency Medicine, Vol. 13, No. 2, 1995, pp. 133-35.

Fredette, M., L. S. Mambu, A. Chouinard, and F. Bellavance. Safety Impacts Due to the Incompatibility of Suvs, Minivans, and Pickup Trucks in Two-Vehicle Collisions, Accident Analysis and Prevention, Vol. 40, No. 6, 2008, pp. 1987-95.

Gårder, P. Segment Characteristics and Severity of Head-On Crashes on Two-Lane Rural Highways in Maine. Accident Analysis and Prevention, Vol. 38, No, 4, 2006, pp. 652-661.

Gates, J., S. Dubois, N. Mullen, B. Weaver, and M. Bédard. The Influence of Stimulants on Truck Driver Crash Responsibility in Fatal Crashes. Forensic Science International, Vol. 228, No. 1-3, 2013, pp. 15-20.

Golias, J. C., and H. S. Tzivelou. Aspects of Road-Accident Death Analyses. Journal of Transportation Engineering, Vol. 118, No. 2, 1992, pp. 299-311.

Gonzalez, R. P., G. Cummings, M. Mulekar, and C. B. Rodning. Increased Mortality in Rural Vehicular Trauma: Identifying Contributing Factors through Data Linkage. Journal of Trauma - Injury, Infection and Critical Care, Vol. 61, No. 2, 2006, pp. 404-09.

Gonzalez, R. P., G. R. Cummings, H. A. Phelan, M. S. Mulekar, and C. B. Rodning. Does Increased Emergency Medical Services Prehospital Time Affect Patient Mortality in Rural Motor Vehicle Crashes? A Statewide Analysis. The American Journal of Surgery, Vol. 197, No. 1, 2009, pp. 30-34.

Gray, R., and D. Regan. Glare Susceptibility Test Results Correlate with Temporal Safety Margin When Executing Turns across Approaching Vehicles in Simulated Low-Sun Conditions. Ophthalmic and Physiological Optics, Vol. 27, No. 5, 2007, pp. 440-450.

Greene, W. H., and D. A. Henshe. A Latent Class Model for Discrete Choice Analysis: Contrasts with Mixed Logit. Transportation Research Part B, Vol. 37, No. 8, 2003, pp. 681-698.

Hanna, C. L., L. Laflamme, and C. R. Bingham. Fatal Crash Involvement of Unlicensed Young Drivers: County Level Differences According to Material Deprivation and Urbanicity in the United States. Accident Analysis and Prevention, Vol. 45, No. 0, 2012, pp. 291-95.

Harper, J. S., W. M. Marine, C. J. Garret, D. Lezotte, and S. R. Lowenstein. Motor Vehicle Crash Fatalities: A Comparison of Hispanic and Non-Hispanic Motorists in Colorado. Annals of Emergency Medicine, Vol. 36, No. 6, 2000, pp. 589– 596.

Helai, H., H. Chin, and M. Haque. Severity of Driver Injury and Vehicle Damage in Traffic Crashes at Intersections: A Bayesian Hierarchical Analysis. Accident Analysis and Prevention, Vol. 40, No. 1, 2008, pp. 45-54.

Holdridge, J. M., V. N. Shankar, and G. F. Ulfarsson. The Crash Severity Impacts of Fixed Roadside Objects. Journal of Safety Research, Vol. 36, No. 2, 2005, pp. 139-47.

Hu, W., and E. Donnell. Median Barrier Crash Severity: Some New Insights. Accident Analysis and Prevention, Vol. 42, No. 6, 2010, pp. 1697-1704.

Huang, H., H. C. Chin, and M. M. Haque. Severity of Driver Injury and Vehicle Damage in Traffic Crashes at Intersections: A Bayesian Hierarchical Analysis. Accident Analysis and Prevention, Vol. 40, No. 1, 2008, pp. 45-54.

International Traffic Safety Data and Analysis Group (IRTAD). Road Safety Annual Report 2013. International Transport Forum, Paris, France, 2013.

Isenberg, D., D. C. Cone, and F. E. Vaca. Motor Vehicle Intrusion Alone does not Predict Trauma Center Admission or Use of Trauma Center Resources. Prehospital Emergency Care, Vol. 15, No. 2, 2011, pp. 203-207.

Islam, S., and F. Mannering. Driver Aging and Its Effect on Male and Female Single-Vehicle Accident Injuries: Some Additional Evidence. Journal of Safety Research, Vol. 37, No. 3, 2006, pp. 267-276.

Janssen, W. Seat-belt Wearing and Driving Behavior: An Instrumented-Vehicle Study. Accident Analysis and Prevention, Vol. 26, No, 2, 1994, pp. 249-261.

Jin, Y., X. Wang, and X. Chen. Right-Angle Crash Injury Severity Analysis Using Ordered Probability Models. 2010 International Conference on Intelligent Computation Technology and Automation, ICICTA 2010, 2010, pp. 206-209.

Ju, Y. H., and S. Y. Sohn. Time to Death Analysis of Road Traffic Accidents in Relation to Delta V, Drunk Driving, and Restraint Systems. Traffic Injury Prevention, Vol. 15, No. 8, 2014, pp. 771-777.

Jung, S., X. Qin, and D. A. Noyce. Injury Severity of Multivehicle Crash in Rainy Weather. Journal of Transportation Engineering, Vol. 138, No. 1, 2011, pp. 50-59.

Jurado-Piña, R., J. M. Pardillo-Mayora, and R. Jiménez. Methodology to Analyze Sun Glare Related Safety Problems at Highway Tunnel Exits. Journal of Transportation Engineering, Vol. 136, No. 6, 2010, pp. 545-53.

Khattak, A. J. Injury Severity in Multivehicle Rear-End Crashes. Transportation Research Record, No. 1746, 2001, pp. 59-68.

Khattak, A. J., and M. Rocha. Are SUVs "Supremely Unsafe Vehicles"?: Analysis of Rollovers and Injuries with Sport Utility Vehicles. Transportation Research Record, No. 1840, 2003, pp. 167-177.

Khattak, A. J., M. D. Pawlovich, R. R. Souleyrette, and S. L. Hallmark. Factors Related to More Severe Older Driver Traffic Crash Injuries. Journal of Transportation Engineering, Vol. 128, No. 3, 2002, pp. 243-249.

Khattak, A., and K. Knapp. Interstate Highway Crash Injuries during Winter Snow and Nonsnow Events. Transportation Research Record, No. 1746, 2001, pp. 30-36.

Khattak, A., P. Kantor, and F. Council. Role of Adverse Weather in Key Crash Types on Limited-Access Roadways: Implications for Advanced Weather Systems. Transportation Research, No. 1621, 1998, pp. 10-19.

Khorashadi, A., D. Niemeier, V. Shankar, and F. Mannering. Differences in Rural and Urban Driver-Injury Severities in Accidents Involving Large-Trucks: An Exploratory Analysis. Accident Analysis and Prevention, Vol. 37, No. 5, 2005, pp. 910-921.

Kim, J.-K., G. F., Ulfarsson, Kim, S., and V. N. Shankar. Driver-Injury Severity in Single-Vehicle Crashes in California: A Mixed Logit Analysis of Heterogeneity Due to Age and Gender. Accident Analysis and Prevention, Vol. 50, 2013, pp. 1073-1081.

Kim, J.-K., S. Kim, G. Ulfarsson, and L. Porrello. Bicyclist Injury Severities in Bicycle-motor Vehicle Accidents. Accident Analysis and Prevention, Vol. 39, No. 2, 2007, pp. 238-251.

Kockelman, K. M., and Y. J. Kweon. Driver Injury Severity: An Application of Ordered Probit Models. Accident Analysis and Prevention, Vol. 34, No. 3, 2002, pp. 313-321.

Konduri, K., S. Astroza, B. Sana, R. Pendyala, and S. Jara-Díaz. Joint Analysis of Time Use and Consumer Expenditure Data. Transportation Research Record, No. 2231, 2011, pp. 53-60.

Krafft, M., A. Kullgren, C. Tingvall, O. Boström, and R. Fredriksson. How Crash Severity in Rear Impacts Influences Short- and Long-term Consequences to the Neck. Accident Analysis and Prevention, Vol. 32, No. 2, 2000, pp. 187-195.

Krull, K., A. J. Khattak, and F. Council. Injury Effects of Rollovers and Events Sequence in Single-Vehicle Crashes. Transportation Research Record, No. 1717, 2000, pp. 46-54.

Kweon, Y. J., and K. M. Kockelman. Overall Injury Risk to Different Drivers: Combining Exposure, Frequency, and Severity Models. Accident Analysis and Prevention, Vol. 35, No. 4, 2003, pp. 441-450.

Lambert-Bélanger, A., S. Dubois, B., Weaver, N., Mullen, and M., Bédard. Aggressive Driving Behaviour in Young Drivers (Aged 16 through 25) Involved in Fatal Crashes. Journal of Safety Research, Vol. 43, No. 5–6, 2012, pp. 333-38.

Lécuyer, J. F., and A. Chouinard. Study on the Effect of Vehicle Age and the Importation of Vehicles 15 Years and Older on the Number of Fatalities, Serious Injuries and Collisions in Canada. Proceedings of the Canadian Multidisciplinary Road Safety Conference XVI. Winnipeg: The Canadian Association of Road Safety Professionals (CARSP), 2006.

Lee, L.-F. Generalized Econometric Models with Selectivity. Econometrica, Vol. 51, No. 2, 1983, pp. 507-512.

Loo, B. P. Y., and K. L. Tsui. Factors Affecting the Likelihood of Reporting Road Crashes Resulting in Medical Treatment to the Police. Injury Prevention, Vol. 13, No. 3, 2007, pp. 186-89.

Lord, D., and F. L. Mannering. The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives. Transportation Research Part A: Policy and Practice. Vol. 44, No. 5, 2010, pp. 291-305.

Louviere, J., R. Meyer, D. Bunch, R. Carson, B. Dellaert, W. M. Hanemann, D. Hensher, and J. Irwin. Combining Sources of Preference Data for Modeling Complex Decision Processes. Marketing Letters, Vol. 10, No. 3, 1999, pp. 205-217.

Lyman, S., S. A. Ferguson, E. R. Braver, and A. F. Williams. Older Driver Involvements in Police Reported Crashes and Fatal Crashes: Trends and Projections. Injury Prevention, Vol. 8, No. 2, 2002, pp. 116-20.

Mackay, G., J. Hill, S. Parkin, and J. Munns. Restrained Occupants on the Nonstruck Side in Lateral Collisions. Accident Analysis and Prevention, Vol. 25, No. 2, 1993, pp. 147-152.

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

Malyshkina, N. V., and F. L. Mannering. Markov Switching Multinomial Logit Model: An Application to Accident-Injury Severities. Accident Analysis and Prevention, Vol. 41, No. 4, 2009, pp. 829-838.

Mannering, F. L., and C. R. Bhat. Analytic Methods in Accident Research: Methodological Frontier and Future Directions. Analytic Methods in Accident Research, Vol. 1, 2014, pp. 1-22.

Mao, Y., J. Zhang, G. Robbins, K. Clarke, M. Lam, and W. Pickett. Factors Affecting the Severity of Motor Vehicle Traffic Crashes Involving Young Drivers in Ontario. Injury Prevention: Journal of the International Society for Child and Adolescent Injury Prevention, Vol. 3, No. 3, 1997, pp. 183-189.

Marson, C. A., and J. C. Thomson. The Influence of Prehospital Trauma Care on Motor Vehicle Crash Mortality. Journal of Trauma - Injury, Infection and Critical Care, Vol. 50, No. 5, 2001, pp. 917-20.

Mayrose, J., D.V.K. Jehle. Vehicle Weight and Fatality Risk for Sport Utility Vehicle-Versus-Passenger Car Crashes, Journal of Trauma, Vol. 53, 2002, pp. 751-753.

McFadden, D. Econometric Models of Probabilistic Choice in C. F. Manski and D. McFadden (ed.). Structural Analysis of Discrete Data with Econometric Applications. Cambridge, Massachusetts: M.I.T. Press, 1981, pp. 198-272.

McLellan, B., Rizoli, S., Brenneman, F., Boulanger, B., Sharkey, P., Szalai, J., 1996. Injury Pattern and Severity in Lateral Motor Vehicle Collisions: A Canadian Experience. Journal of Trauma - Injury, Infection and Critical Care 41 (4), 708-713.

Meng, Q., and J. Weng. Uncertainty Analysis of Accident Notification Time and Emergency Medical Service Response Time in Work Zone Traffic Accidents. Traffic Injury Prevention, Vol. 14, No. 2, 2013, pp. 150-58.

Ministry of Infrastructure and Transport. Road deaths in Australia 1925–2008, Department of Infrastructure, Transport, Regional Development and Local Government. Australia Government, 2010.

Mohamed, M. G., N. Saunier, L. F. Miranda-Moreno, and S. V. Ukkusuri. A Clustering Regression Approach: A Comprehensive Injury Severity Analysis of Pedestrian-Vehicle Crashes in New York, US and Montreal, Canada. Safety Science, Vol. 54, 2013, pp. 27-37.

Morgan, A., and F. L. Mannering. The Effects of Road-Surface Conditions, Age, and Gender on Driver-Injury Severities, Accident Analysis and Prevention, Vol. 43, No. 5, 2011, pp. 1852-1863.

Narayanamoorthy, S., R. Paleti, and C.R. Bhat, A Spatial Model for Examining Bicycle and Pedestrian Injuries, Technical paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, July 2012.

National Highway Traffic Safety Administration (NHTSA), Traffic Safety Facts 2012, A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System, NHTSA, Washington, D.C., 2010. Accessed online at http://www-nrd.nhtsa.dot.gov/Pubs/812032.pdf on September 9th 2014.

National Highway Traffic Safety Administration (NHTSA), Traffic Safety Facts: 2010 Motor Vehicle Crashes Overview, NHTSA, Washington, D.C., 2012.

Nordhoff, L. S. Motor Vehicle Collision Injuries: Biomechanics, Diagnosis, and Management. Jones and Bartlett Learning, 2005.

Nylund, K. L., T. Asparouhov, and B. O. Muthén. Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. Structural Equation Modeling, Vol. 14, No. 4, 2007, pp. 535-569.

O'Neill, B., S.Y. Kyrychenko. Crash Incompatibilities between Cars and Light Trucks: Issues and Potential Countermeasures, Society of Automotive Engineers, 2004, pp. 39-47.

Obeng, K. Injury Severity, Vehicle Safety Features, and Intersection Crashes. Traffic Injury Prevention, Vol. 9, No. 3, 2008, pp. 268-76.

O'Donnell, C. J., and D. H. Connor. Predicting the Severity of Motor Vehicle Accident Injuries Using Models of Ordered Multiple Choice. Accident Analysis and Prevention, Vol. 28, No. 6, 1996, pp. 739-753.

Offaly History. Famous Offaly People – a series of short biographies, Mary Ward 1827-1869. Offaly Historical & Archaeological Society, Bury Quay, Tullamore, Co. Offaly, 2007.

O'Neill, B. Preventing Passenger Vehicle Occupant Injuries by Vehicle Design- a Historical Perspective from Iihs. Traffic Injury Prevention, Vol. 10, No. 2, 2009, pp. 113-26.

Palanca, S., D. M. Taylor, M. Bailey, and P. A. Cameron. Mechanisms of Motor Vehicle Accidents that Predict Major Injury. Emergency Medicine, Vol. 15, No. 5-6, 2003, pp. 423-428.

Paleti, R., N. Eluru, and C. R. Bhat. Examining the Influence of Aggressive Driving Behavior on Driver Injury Severity in Traffic Crashes. Accident Analysis and Prevention, Vol. 42, No. 6, 2010, pp. 1839-1854.

Parker, D., J. Reason, A. Manstead, and S. Stradling. Driving Errors, Driving Violations and Accident Involvement. Ergonomics, Vol. 38, No. 5, 1995, pp. 1036-1048.

Persaud, B. N., R. A. Retting, P. E. Garder, and D. Lord. Safety Effect of Roundabout Conversions in the United States: Empirical Bayes Observational before-after Study. Transportation Research Record, No. 1751, 2001, pp. 1-8.

Plainis, S., I. Murray, and I. Pallikaris. Road Traffic Casualties: Understanding the Night-time Death Toll. Injury Prevention, Vol. 12, No. 2, 2006, pp. 125-128.

Portoghese, A., E. Spissu, C. R. Bhat, N. Eluru, and I. Meloni. A Copula-Based Joint Model of Commute Mode Choice and Number of Non-Work Stops During the Commute. International Journal of Transport Economics, Vol. 38, No. 3, 2011, pp. 337-62.

Prentkovskis, O., E. Sokolovskij, and V. Bartulis. Investigating Traffic Accidents: A Collision of Two Motor Vehicles. Transport, Vol. 25, No. 2, 2010, pp. 105-115.

Preusser, D. F., A. F. Williams, S. A. Ferguson, R. G. Ulmer, and H. B. Weinstein. Fatal Crash Risk for Older Drivers at Intersections. Accident Analysis and Prevention, Vol. 30, No. 2, 1998a, pp. 151-59.

Preusser, D. F., S. A. Ferguson, and A. F. Williams. The Effect of Teenage Passengers on the Fatal Crash Risk of Teenage Drivers. Accident Analysis and Prevention, Vol. 30, No. 2, 1998b, pp. 217-22.

Preusser, D., A. Williams, and A. Lund. Characteristics of Belted and Unbelted Drivers. Accident Analysis and Prevention, Vol. 23, No. 66, 1991, pp. 475-482.

Quinn, C. The health-economic applications of copulas: Methods in applied econometric research. Health, Econometrics and Data Group (HEDG) Working Paper 07/22, Department of Economics, University of York, 2007.

Rana, T., S. Sikder, and A. Pinjari. Copula-Based Method for Addressing Endogeneity in Models of Severity of Traffic Crash Injuries. Transportation Research Record, No. 2147, 2010, pp. 75-87.

Renski, H., A. J. Khattak, and F. M. Council. Effect of Speed Limit Increases on Crash Injury Severity: Analysis of Single-Vehicle Crashes on North Carolina Interstate Highways. Transportation Research Record, No. 1665, 1999, pp. 100-108.

Retting, R. A., B. N. Persaud, P. E. Garder, and D. Lord. Crash and Injury Reduction Following Installation of Roundabouts in the United States. American Journal of Public Health, Vol. 91, No. 4, 2001, pp. 628-631.

Retting, R. A., H. B. Weinstein, and M. G. Solomon. Analysis of Motor-Vehicle Crashes at Stop Signs in Four U.S. Cities. Journal of Safety Research, Vol. 34, No. 5, 2003, pp. 485-489.

Rifaat, S., and R. Tay. Effects of Street Patterns on Injury Risks in Two-Vehicle Crashes. Transportation Research Record, No. 2012, 2009, pp. 61-67.

Romano, E., R. Voas, and S. Tippetts. Stop Sign Violations: The Role of Race and Ethnicity on Fatal Crashes. Journal of Safety Research, Vol. 37, No. 1, 2006, pp. 1-7.

Rumar, K. The Role of Perceptual and Cognitive Filters on Observed Behavior. In Evans, L., and R. Schwing, Human Behavior and Traffic Safety. New York: Plenum Press, 1985, pp. 151-165.

Ryb, G. E., P. C. Dischinger, G. McGwin, and R. L. Griffin. Crash-related mortality and model year: Are newer vehicles safer? Paper presented at The Annals of Advances in Automotive Medicine, Vol. 55, 2011, pp. 113-121.

Sánchez-Mangas, R., A. García-Ferrrer, A. De Juan, and A. M. Arroyo. The Probability of Death in Road Traffic Accidents. How Important Is a Quick Medical Response?. Accident Analysis and Prevention, Vol. 42, No. 4, 2010, pp. 1048-56.

Sasser, S. M., R. C. Hunt, M. Faul, D. Sugerman, W. S. Pearson, T. Dulski, M. M. Wald, G. J. Jurkovich, C. D. Newgard, E. B. Lerner, A. Cooper, S. C. Wang, M. C. Henry, J. P. Salomone, R. L. Galli. Guidelines for Field Triage of Injured Patients Recommendations of the National Expert Panel on Field Triage, 2011. Morbidity and Mortality Weekly Report, No. 61(RR-1), 2012, pp. 1-23.

Sauaia, A., F. A. Moore, E. E. Moore, K. S. Moser, R. Brennan, R. A. Read, and P. T. Pons Epidemiology of Trauma Deaths: A Reassessment. Journal of Trauma-Injury, Infection, and Critical Care, Vol. 38, No. 2, 1995, pp. 185-193.

Savolainen, P. T., and F. L. Mannering. Probabilistic Models of Motorcyclists' Injury Severities in Single- and Multi-Vehicle Crashes. Accident Analysis and Prevention, Vol. 39, No. 5, 2007, pp. 955-963.

Savolainen, P. T., F. L. Mannering, D. Lord, and M. A. Quddus. The Statistical Analysis of Highway Crash-Injury Severities: A Review and Assessment of Methodological Alternatives. Accident Analysis and Prevention, Vol. 43, No. 5, 2011, pp. 1666-1676.

Schiff, M. A., and P. Cummings. Comparison of Reporting of Seat Belt Use by Police and Crash Investigators: Variation in Agreement by Injury Severity. Accident Analysis and Prevention, Vol. 36, No. 6, 2004, pp. 961-65.

Schneider, W., P. Savolainen, and K. Zimmerman. Driver Injury Severity Resulting from Single-Vehicle Crashes Along Horizontal Curves on Rural Two-Lane Highways. Transportation Research Record, No. 2102, 2009, pp. 85-92.

Shankar, V., and F. Mannering. An Exploratory Multinomial Logit Analysis of Single-Vehicle Motorcycle Accident Severity. Journal of Safety Research, Vol. 27, No, 3, 1996, pp. 183-194.

Shibata, A., and K. Fukuda. Risk Factors of Fatality in Motor Vehicle Traffic Accidents. Accident Analysis and Prevention, Vol. 26, No. 3, 1994, pp. 391-397.

Siskind, V., D. Steinhardt, M. Sheehan, T. O'Connor, and H. Hanks. Risk Factors for Fatal Crashes in Rural Australia. Accident Analysis and Prevention, Vol. 43, No. 3, 2011, pp. 1082-1088.

Sivak, M., B. Schoettle, and J. Rupp. Survival in Fatal Road Crashes: Body Mass Index, Gender, and Safety Belt Use. Traffic Injury Prevention, Vol. 11, No. 1, 2010, pp. 66-68.

Sivakumar, A. and Polak, J.W. (2013). Exploration of Data-Pooling Techniques: Modeling Activity Participation and Household Technology Holdings. Presented at the 92nd Annual Meeting of the Transportation Research Board, 2013, Washington D.C., USA.

Sklar, A. Random Variables, Joint Distribution Functions, and Copulas. Kybernetika, Vol. 9, 1973, pp. 449-460.

Small, K. A. A Discrete Choice Model for Ordered Alternatives. Econometrica, Vol. 55, No. 2, 1987, pp. 409-424.

Sobhani A., N. Eluru, and A. Faghih-Imani. A Latent Segmentation based Multiple Discrete Continuous Extreme Value Model. Transportation Research Part B, Vol. 58, 2013, pp. 154-169.

Sobhani, A., W. Young, D. Logan, and S. Bahrololoom. A Kinetic Energy Model of Two-Vehicle Crash Injury Severity. Accident Analysis and Prevention, Vol. 43, No. 3, 2011, pp. 741-754.

Soderstrom, C. A., M. F. Ballesteros, P. C. Dischinger, T. J. Kerns, R. D. Flint, and G. S. Smith. Alcohol/Drug Abuse, Driving Convictions, and Risk-Taking Dispositions among Trauma Center Patients. Accident Analysis and Prevention, Vol. 33, No. 6, 2001, pp. 771-82.

Srinivasan, K. K. Injury Severity Analysis with Variable and Correlated Thresholds: Ordered Mixed Logit Formulation. Transportation Research Record, No. 1784, 2002, pp. 132-142.

Stewart, R. D. Prehospital care of trauma. In: Management of Blunt Trauma. Baltimore, MD: Williams and Wilkins, 1990, pp. 23–29.

Stübig, T., M. Petri, C. Zeckey, S. Brand, C. Müller, D. Otte, C. Krettek, and C. Haasper. Alcohol Intoxication in Road Traffic Accidents Leads to Higher Impact Speed Difference, Higher Iss and Mais, and Higher Preclinical Mortality. Alcohol, Vol. 46, No. 7, 2012, pp. 681-86.

Subramanian, R. Motor Vehicle Traffic Crashes as a Leading Cause of Death in the United States, 2003. Traffic Safety Facts, NHTSA Research Note, 2006.

Tay, R., and S. M. Rifaat. Factors Contributing to the Severity of Intersection Crashes. Journal of Advanced Transportation, Vol. 41, No. 3, 2007, pp. 245-65.

Tay, R., J. Choi, L. Kattan, A. Khan. A Multinomial Logit Model of Pedestrian-Vehicle Crash Severity. International Journal of Sustainable Transportation, Vol. 5, No. 4, 2011, pp. 233-249.

Terza, J. V., A. Basu, and P. J. Rathou. Two-stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling. Journal of Health Economics, Vol. 27, No. 3, 2008, pp. 531-543.

Terza, J.V. Ordinal Probit: A Generalization. Communications in Statistics: Theory and Methods, Vol. 14, No. 1, 1985, pp. 1-11.

Thompson, J. P., M. R. J. Baldock, J. L. Mathias, and L. N. Wundersitz. An Examination of the Environmental, Driver and Vehicle Factors Associated with the Serious and Fatal Crashes of Older Rural Drivers. Accident Analysis and Prevention, Vol. 50, 2013, pp. 768-75.

Tohira, H., I. Jacobs, D. Mountain, N. Gibson, and A. Yeo. Differences in risk factors between early and late trauma death after road traffic accidents. Paper read at 2012 IRCOBI Conference Proceedings - International Research Council on the Biomechanics of Injury. 2012.

Toy, E. L., and J. K. Hammitt. Safety Impacts of SUVs, Vans, and Pickup Trucks in Two-vehicle Crashes. Risk Analysis, Vol. 23, No. 4, 2003, pp. 641-650.

Travis, L. L., D. E. Clark, A. E. Haskins, and J. A. Kilch. Mortality in Rural Locations after Severe Injuries from Motor Vehicle Crashes. Journal of Safety Research, Vol. 43, No. 5-6, 2012, pp. 375-80.

Trivedi, P.K., D.M., Zimmer. Copula modeling: an introduction for practitioners. Foundations and Trends in Econometrics, Vol. 1, No. 1, 2007, pp. 1-110.

Trunkey DD. Trauma. Sci Am, Vol. 249, 1983, pp. 28–35.

Tsui, K. L., F. L. So, N. N. Sze, S. C. Wong, and T. F. Leung. Misclassification of Injury Severity among Road Casualties in Police Reports. Accident Analysis and Prevention, Vol. 41, No. 1, 2009, pp. 84-89.

Tulloh, B. R. Positive Correlation between Blood Alcohol Level and Iss in Road Trauma. Injury, Vol. 25, No. 8, 1994, pp. 539-43.

Ulfarsson, G. F., and F. L. Mannering. Differences in Male and Female Injury Severities in Sport-Utility Vehicle, Minivan, Pickup and Passenger Car Accidents, Accident Analysis and Prevention, Vol. 36, No. 2, 2004, pp. 135-47.

Valent, F., F. Schiava, C. Savonitto, T. Gallo, S. Brusaferro, and F. Barbone. Risk Factors for Fatal Road Traffic Accidents in Udine, Italy. Accident Analysis and Prevention, Vol. 34, No. 1, 2002, pp. 71-84.

Verma, V., F. Gagliardi, and C. Ferretti. On Pooling of Data and Measures. Working Paper no° 84/2009, DMQ, Università di Siena, 2009.

Viano, D. C., and C. S. Parenteau. Severe-to-fatal Injury Risks in Crashes with Two Front-Seat Occupants by Seat Belt Use. Traffic Injury Prevention, Vol. 11, No. 3, 2010, pp. 294-299.

Wang, C., M. A. Quddus, and S. G. Ison. Predicting Accident Frequency at Their Severity Levels and Its Application in Site Ranking Using a Two-Stage Mixed Multivariate Model. Accident Analysis and Prevention, Vol. 43, No. 6, 2011, pp. 1979-90.

Wang, X., and K. M. Kockelman. Use of Heteroscedastic Ordered Logit Model to Study Severity of Occupant Injury Distinguishing Effects of Vehicle Weight and Type. Transportation Research Record, No. 1908, 2005, pp. 195-204.

Washington, S., M. G. Karlaftis, F. L. Mannering. Statistical and Econometric Methods for Transportation Data Analysis. Boca Raton: Chapman & Hall/CRC, 2003.

Wen, C.-H., and F. S. Koppelman. The Generalized Nested Logit Model. Transportation Research Part B: Methodological, Vol. 35, No. 7, 2001, pp. 627-641.

Weninger, P., and H. Hertz. Factors Influencing the Injury Pattern and Injury Severity after High Speed Motor Vehicle Accident-a Retrospective Study. Resuscitation, Vol. 75, No. 1, 2007, pp. 35-41.

Williams, A. F. Nighttime Driving and Fatal Crash Involvement of Teenagers. Accident Analysis and Prevention, Vol. 17, No. 1, 1985, pp. 1-5.

Williams, A. F., and V. I. Shabanova. Responsibility of Drivers, by Age and Gender, for Motor-Vehicle Crash Deaths. Journal of Safety Research. Vol. 34, No. 5, 2003, pp. 527-31.

Windmeijer, F. A. G. Goodness-of-fit Measures in Binary Choice Models. Econometric Reviews, Vol. 14, No. 1, 1995, pp. 101-116.

World Health Organization (WHO). Global Health Estimates Summary Tables: DALYs by Cause, Age and Sex. World Health Organization, Geneva, Switzerland, 2013a.

World Health Organization (WHO). Global Status Report on Road Safety 2013: Supporting a Decade of Action. World Health Organization, Geneva, Switzerland, 2013b.

Xie, Y., K. Zhao, and N. Huynh. Analysis of Driver Injury Severity in Rural Single-Vehicle Crashes. Accident Analysis and Prevention, Vol. 47, 2012, pp. 36-44.

Xie, Y., Y. Zhang, and F. Liang. Crash Injury Severity Analysis Using Bayesian Ordered Probit Models, Journal of Transportation Engineering, Vol. 135, No. 1, 2009, pp. 18-25.

Xiong, Y., and F. L. Mannering. The Heterogeneous Effects of Guardian Supervision on Adolescent Driver-Injury Severities: A Finite-Mixture Random-Parameters Approach. Transportation Research Part B, Vol. 49, 2013, pp. 39-54.

Xiong, Y., J. L. Tobias, and F. L. Mannering. The Analysis of Vehicle Crash Injury-Severity Data: A Markov Switching Approach with Road-Segment Heterogeneity. Transportation Research Part B: Methodological, Vol. 67, 2014, pp. 109-128.

Yamamoto, T., and V. N. Shankar. Bivariate Ordered-Response Probit Model of Driver's and Passenger's Injury Severities in Collisions with Fixed Objects. Accident Analysis and Prevention, Vol. 36, No. 5, 2004, pp. 869-876.

Yamamoto, T., J. Hashiji, and V. N. Shankar. Underreporting in Traffic Accident Data, Bias in Parameters and the Structure of Injury Severity Models. Accident Analysis and Prevention, Vol. 40, No. 4, 2008, pp. 1320-1329.

Yan, X., E. Radwan, and M. Abdel-Aty. Characteristics of Rear-End Accidents at Signalized Intersections Using Multiple Logistic Regression Model. Accident Analysis and Prevention. Vol. 37, No. 6, 2005, pp. 983-95.

Yasmin, S., S. Anowar, and R. Tay. Effects of Drivers' Actions on Severity of Emergency Vehicle Collisions. Transportation Research Record, No. 2318, 2012, pp. 90-97.

Yasmin. S., N. Eluru, and S. V. Ukkusuri. Alternative Ordered Response Frameworks for Examining Pedestrian Injury Severity in New York City. Journal of Transportation Safety and Security, Vol. 6, No. 4, 2014, pp. 275-300.

Yau, K. K. W. Risk Factors Affecting the Severity of Single Vehicle Traffic Accidents in Hong Kong. Accident Analysis and Prevention, Vol. 36, No. 3, 2004, pp. 333-40.

Ye, F., and D. Lord. Investigation of effects of underreporting crash data on three commonly used traffic crash severity models. Transportation Research Record, No. 2241, 2011, pp. 51-58.

Ye, X., R. M. Pendyala, F. S. Al-Rukaibi, and K. Konduri. A Joint Model of Accident Type and Severity for Two-Vehicle Crashes. Presented at 87th Annual Meeting of the Transportation Research Board, Washington, D.C., 2008.

Young, R. K., and J. Liesman. Estimating the Relationship between Measured Wind Speed and Overturning Truck Crashes Using a Binary Logit Model. Accident Analysis and Prevention, Vol. 39, No. 3, 2007, pp. 574-580.

Zador, P. L., S. A. Krawchuk, and R. B. Voas. Alcohol-Related Relative Risk of Driver Fatalities and Driver Involvement in Fatal Crashes in Relation to Driver Age and Gender: An Update Using 1996 Data. Journal of Studies on Alcohol, Vol. 61, No. 3, 2000, pp. 387-95.

Zeckey, C., S. Dannecker, F. Hildebrand, P. Mommsen, R. Scherer, C. Probst, C. Krettek, and M. Frink. Alcohol and Multiple Trauma-Is There an Influence on the Outcome?. Alcohol, Vol. 45, No. 3, 2011, pp. 245-51.

Zhang, C., and J. N. Ivan. Effects of Geometric Characteristics on Head-on Crash Incidence on Two-Lane Roads in Connecticut. Transportation Research Record, No. 1908, 2005, pp. 159-164.

Zhang, X., H. Yao, G. Hu, M. Cui, Y. Gu, and H. Xiang. Basic Characteristics of Road Traffic Deaths in China. Iranian Journal of Public Health, Vol. 42, No. 1, 2013, pp. 7-15.

Zuxuan, D., J. N. Ivan, and P. Gårder. Analysis of Factors Affecting the Severity of Head-on Crashes Two-Lane Rural Highways in Connecticut. Transportation Research Record, No. 1953, 2006, pp. 137-146.