AN ABSTRACT OF THE THESIS OF

Jason Anderson for the degree of Master of Science in Civil Engineering presented on August 5, 2016.

Title: An Evaluation of the Random-Parameter and Latent Class Methods for Heavy Vehicle Injury Severity and Crash Rate Analysis: An Idaho Case Study by Roadway Classification

Abstract approved:

____

Salvador Hernandez

This thesis provides a comparison of advanced econometric frameworks to account for unobserved factors in crash reported data (also referred to as unobserved heterogeneity) while identifying contributing factors by roadway classification for heavy vehicle injury severity and crash rates. The presented thesis provides two manuscripts that expand the literature regarding these advanced econometric methods using Idaho heavy vehicle crash reported data as a case study.

The first manuscript utilizes two advanced analytical techniques, namely the random-parameter multinomial logit (also referred to as the mixed logit) and latent class logit, to identify injury severity contributing factors while exploring the empirical results of the two methods. Recent efforts suggest that more studies examining the results of the two approaches be completed to facilitate the identification of a superior framework that can used for future analyses. In comparing overall model fit (log-likelihood values), marginal effects and actual severities versus predicted severities, it was found that the latent class framework for heavy vehicle injury severity analysis performed better for the Idaho crash data. Further, through a model separation test, it was found that road classifications need to be analyzed separately with 99.99% confidence.

In regard to the second manuscript, two additional advanced econometric approaches were utilized to investigate the factors that contribute to the number of crashes per million-vehicle-miles-traveled. Again, analysis was completed by road classification, as it was discovered in manuscript one that road classifications need to be analyzed separately. Due to the skewed distribution of heavy vehicle crash rates, Tobit regression was applied and compared to the empirical results of a latent class Tobit regression framework. To determine the most statistically significant method, overall model fit, partial effects and actual crash rates versus predicted crash rates were evaluated. The latent class Tobit regression framework outperformed that of the traditional Tobit regression approach for the Idaho dataset.

Through the comparison of the crash analysis framework, latent class logit and latent class Tobit regression were found to outperform their traditional counterparts. In the midst of evaluating the empirical results, this thesis has statistically determined that road classifications need to be analyzed individually. The current thesis extends the literature in regard to heavy vehicle injury severity analysis and fills the noticeable gap that exists for heavy vehicle crash rate analysis. An analytical foundation has been provided and can be used for future studies that need to model discrete outcomes or continuous response variables. Although agencies typically do not use such advanced methods, the results from this thesis can help the Idaho Department of Transportation facilitate crash countermeasures with more precision and allow them to prioritize accordingly.

An Evaluation of the Random-Parameter and Latent Class Methods for Heavy Vehicle Injury Severity and Crash Rate Analysis: An Idaho Case Study by Roadway Classification

by
Jason Anderson

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented August 5, 2016
Commencement June 2017

Master of Science thesis of Jason Anderson presented on August 5, 2016

APPROVED:

Major Professor, representing Civil Engineering

Head of the School of Civil and Construction Engineering

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Jason Anderson, Author

ACKNOWLEDGEMENTS

First, and foremost, I want to sincerely thank my advisor, Dr. Sal Hernandez, for giving me the opportunity to pursue my graduate degree. Dr. Hernandez has introduced me to an entire new path that can be taken in the engineering field, and I will forever be thankful for all that he has done for me and continues to do. Candidly, I could not see myself working with any other professor and beyond eager to begin my PhD journey with Dr. Hernandez. Dr. Hernandez, thank you.

I would also like to thank Dr. Chris Bell, Dr. Shane Brown and Dr. Tom Miller. As with Dr. Hernandez, these professors saw a potential in me and have been a crucial part of my academic journey, even if they are unware of such. I will always remember what you have allowed me to accomplish.

I want to thank Dr. Fred Mannering for his very candid feedback and help with my first paper. This has been a learning experience and his comments come with great merit.

I want to thank all the faculty at Oregon State, within the Transportation Department and outside the Transportation Department. Oregon State has been so good to me and I will make it a life's goal to represent Oregon State in the way it deserves. There's something special about the Civil Engineering Department at Oregon State and I am proud to be part of it.

Lastly, I want to thank all my loved ones. They love to give me a hard time, but when it comes to it, they have been very supportive through this process and are behind me 100%. Only a few years left now.

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (CONTINUED)

Page

TABLE OF CONTENTS (CONTINUED)

Page

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES (CONTINUED)

1.0 INTRODUCTION

1.1 Motivation

In recent years there has been an increasing interest in understanding the contributing factors to heavy vehicle crashes through the application of advanced econometric methods. For example, the ordered probit, random-parameter ordered probit, multinomial logit, random-parameter multinomial logit (also referred to as the mixed logit), and latent class logit models (see Behnood et al., 2014; Khorashadi et al., 2005; Mannering et al., 2016; Milton et al., 2008; Morgan and Mannering, 2011; Russo et al., 2014; Xiong and Mannering, 2013; Yasmin and Eluru, 2013; Yasmin et al., 2014; Ye and Lord, 2014 for a thorough review of these models). These methods have been applied to answer questions in the context of injury severity of heavy vehicle crashes holistically (i.e. the use of all data in the model), to understand the differences between urban versus rural crashes, the effect time-of-day has on injury severity and more recently the use of a binary logit model for run-of-road crashes (Khorashadi et al., 2005; Pahukula et al., 2015; Peng and Boyle, 2012). Although there have been several efforts to understand heavy vehicle crashes, the relationship between crash related factors, crash severity, and roadway classification are still not completely understood.

Similarly, while most of the recent work related to heavy vehicle crashes has focused on seeking the understanding of the contributing factors to injury severity, little attention has been paid to understanding heavy vehicle crash rates. The literature is sparse in this regard with much of the attention focused on passenger vehicle crash rate analyses. The most frequently used methods in the context of crash rate analyses have been crash frequency analyses and consist primarily of Negative Binomial and Poisson models (Abdel-Aty and Radwan, 2000; Poch and Mannering, 1996; Savolainen and Tarko, 2005; Shankar et al., 1995) and their variants the zero-inflated Poisson and zero-inflated Negative Binomial models (Carson and Mannering, 2001; Lee and Mannering, 2002; Shankar et al., 1997), random-parameter Negative Binomial models (Anastasopoulos and Mannering, 2009; Chin and Quddus, 2003; Shankar et al., 1998), Tobit and random-parameter Tobit (Anastasopoulos et al., 2012b, 2008; Islam and Hernandez, 2015), Markov switching of two different state of crash occurrence (Malyshkina et al., 2009) and Bayesian statistics on Negative Binomial models (Park et al., 2010). While literature in crash frequency and crash rates is rich for passenger cars, crash rates for heavy vehicles has not been widely studied.

With this in mind, the purpose of this thesis is to develop an advanced econometric modeling framework through the ***comparison*** of recent analytical methods in crash research to better understand heavy vehicle crashes by injury severity and crash rates as a result of roadway classification (see sections 2 and 3). These advanced econometric methods are the random-parameter multinomial logit and latent class logit modeling frameworks (applied to determine the contributing factors for injury severity as a result of roadway classification) and the random-parameter Tobit and latent class Tobit models modeling frameworks (applied to the heavy vehicle crash rate analysis).

The advantages of utilizing the *random-parameter* and *latent class* extensions of these models is that they provide the modeler with a mechanism to account for the effect of unobserved factors (also referred to as unobserved heterogeneity) that can arise from factors related to an individual (e.g. truck drivers), road and environmental, vehicle, weather, temporal effects, differences in police reporting and other factors not observed or captured in the data sets (see sections 2 and 3 for model specific advantages). By accounting for unobserved heterogeneity, these modeling extensions can correct for bias parameter estimates that can lead to incorrect inferences, which eventually could lead to ineffective implementation of safety measures.

Although both *random-parameter* and *latent class* extensions do address the issue of unobserved heterogeneity, both methods have their drawbacks. The latent class methodology accounts for possible heterogeneity by assuming that observations (e.g. large truck crashes) come from distinct *classes* based on common characteristics. However, the drawback of this assumption is that the number of classes can be small so there is a coarse approximation of the distribution of unobserved heterogeneity (Behnood et al., 2014). Unlike the latent class approach where a distribution that is not defined by the analyst and a finite number of points across a specified number of groups (or classes) is used to account for heterogeneity, in the random-parameter framework the analyst must make an assumption about the distribution of the factors that may vary across observations (unobserved heterogeneity) which may not always hold true for each observation (Pahukula et al., 2015).

Taking into account the advantages and disadvantages of the aforementioned econometric modeling frameworks, the work performed in this thesis provides a set of heavy vehicle safety models that account for unobserved factors; these models can assist the trucking industry, transportation agencies, engineers and safety planners to become more effective when it comes to safety planning. Therefore, to the best of our knowledge, these are the first attempts to compare the *random-parameter* and *latent class* modeling frameworks to determine the contributing factors for injury severity and crash rates as a result of roadway classification utilizing seven years (2007-2013) of crash reported data from the State of Idaho.

1.2 Thesis Organization

This thesis is presented in manuscript form. Chapter 2 presents the injury severity analysis and Chapter 3 presents the crash rate analysis. The thesis concludes by summarizing the key findings from both analyses, providing practical application of the findings and suggests avenues for future work.

2.0 INJURY SEVERITY ANALYSIS OF HEAVY VEHICLE CRASHES BY ROADWAY CLASSIFICATION: AN IDAHO CASE STUDY

BY: JASON C. ANDERSON
Graduate Student, School of Civil and Construction Engineering
Oregon State University, 211 Kearney Hall, Corvallis, OR 97331-3212
anderjas@oregonstate.edu

SALVADOR HERNANDEZ, PhD
Assistant Professor, School of Civil and Construction Engineering
Oregon State University, 309 Owen Hall, Corvallis, OR 97331-3212
(541)-737-4740
sal.hernandez@oregonstate.edu

ABSTRACT

To extend traditional heavy vehicle injury severity analyses, this study explores and compares two advanced econometric methods, namely, random-parameter multinomial logistic regression and latent class multinomial logistic regression, to determine contributing factors that lead to injury severity outcomes by roadway classification while accounting for unobserved heterogeneity (e.g. crash data). In the course of determining statistically significant injury severity factors, a model separation test was conducted to determine if roadway classifications should be considered separately for analysis purposes and the overall fit and marginal effect estimates of the two frameworks were assessed and compared. Empirical results show that roadway classifications should be modeled separately with a high level of confidence and that the latent class logit approach provided a better overall fit and larger marginal effects for the given dataset. The majority of significant variables are exclusive to a specific road classification. Results from this study provide a framework to better capture unobserved heterogeneity present in crash datasets and better predict injury severity outcomes which can be applied to future injury severity analyses, as well as illustrate a need to consider roadway classifications independently.

Keywords: Heavy Vehicle, Mixed Logit, Latent Class, Roadway Classification, Injury Severity

2.1 Introduction

The economic impact of heavy vehicle crashes is widely known. For instance, in 2011, heavy vehicle crashes resulted in a total of $87 billion nationwide (Federal Motor Carrier Safety Administration, 2013). Furthermore, property-damage-only crashes (no injury), crashes involving injuries and fatal crashes culminated to $16 billion, $32 billion and $39 billion, respectively (Federal Motor Carrier Safety Administration, 2013). In addition, costs due to delay and other consequences of crashes totaled $28 billion (Blincoe et al., 2015). Therefore, any decrease in heavy vehicle crashes can lead to a substantial reduction in societal costs.

More specifically, at the State level, Idaho experienced a 5.6% increase in heavy vehicle crashes from 2010 to 2013 (Idaho Office of Highway Safety, 2014). In 2014, 65.2% of all heavy vehicle crashes in Idaho resulted in no injury, 33.4% involved an injury and 1.4% were fatal (Idaho Office of Highway Safety, 2014). Heavy vehicle crashes happened most often on local roads (e.g. major collectors) accounting for 50% of all heavy vehicle crashes (Idaho Office of Highway Safety, 2014), while 50% of no injury crashes also took place on local roads. Correspondingly, 68% of all fatalities occurred on U.S. and State highways (e.g. interstates, principal arterials) and 28% of injury crashes happened on interstates (Idaho Office of Highway Safety, 2014). As seen from these statistics there is a clear need to better understand the effect of roadway classification and the associated contributing factors (observed and unobserved) on injury severity.

When fitting injury severity models, roadway classification is typically modeled as an indicator variable (e.g. a factor that may influence an injury severity outcome), yet studies that extend such an analysis to focus on injury severities by roadway classification are scarce and disaggregating heavy vehicle crashes by roadway classification can provide additional insights to assist transportation engineers, planners and agencies in mitigating these types of crashes and their associated costs. Taking this into consideration, the present study will use seven years of Idaho crash data to investigate heavy vehicle crashes by roadway classification through an injury severity analysis on four types of roadway classifications, namely, principal arterials, major collectors, interstates and other principal arterials.

Crash datasets, however, naturally do not include each and every component that contributes to the probability of a distinct injury severity (i.e. specific attributes regarding driver behavior, environmental aspects, characteristics of the roadway, etc.) and, as a result, can lead to biased estimates and inaccurate inferences—this issue is known as unobserved heterogeneity. For example, seatbelts are known to save lives, yet they are capable of causing an injury based on several unseen factors (e.g. speed at the point of impact, physiology of the driver) (Mannering et al., 2016). Likewise, crash data will often indicate if the crash occurred on a grade, but characteristics such as the percent grade and direction of travel (up/down grade) are often unknown. Variation within a known factor can also result in unobserved heterogeneity, such as gender, posted speed limits and weather conditions. Males, for instance, have variation in their physical fitness level, health status and degree of vision (e.g. 20/20 and does not require prescription glasses to drive) (Mannering et al., 2016). Posted speed limits, although regulatory, are not always adhered to and may not be describing the actual speed the driver was traveling when the crash occurred. Driver behavior in accordance with weather conditions is expected to differ by geographic location; for example, drivers living in areas that get snow each winter are going to be more experienced with driving in snowy conditions and able of mitigating the likelihood of a more severe crash. Due to such variation, to properly overcome the limitations of the data and resulting unobserved heterogeneity, this analysis is conducted through the exploration of two discrete outcome modeling techniques that identify contributing factors based on statistical significance, the random-parameter multinomial logit model and the latent class multinomial logit model.

The random-parameter multinomial logit model, often referred to as the mixed logit model, utilizes a number of linear functions to determine the probability that a crash will result in a specific injury severity. In addition, the analyst typically specifies that the random parameters are to be normally distributed when accounting for the varying effects due to heterogeneity. The mixed logit model is a common approach to injury severity analyses and has been used in recent efforts to investigate injury severity (Anastasopoulos and Mannering, 2011; Kim et al., 2013, 2010; Milton et al., 2008; Moore et al., 2011; Yasmin and Eluru, 2013). In particular, injury severity analysis regarding heavy vehicles is less documented. Islam and Hernandez (2013a) used the mixed logit approach to identify injury severity contributing factors for any crash involving a heavy vehicle in Texas without disaggregating the data into any subpopulations (i.e. roadway classification). Khorashadi et al. (2005), however, investigated heavy vehicle injury severity by rural and urban area crashes utilizing a fixed-parameter logit model and found that urban and rural areas need to be modeled separately through a transferability test. More, Pahukula et al. (2015) separated heavy vehicle crashes by time of day to fit several mixed logit models and discovered that parameter estimates are statistically different by time of day. Cerwick et al. (2014) expanded the heavy vehicle injury severity analysis by introducing a latent class logit model and comparing its estimates to that of the mixed logit model. The authors based their evaluation on overall model fit, inferences based on marginal effects and predicted crash severity outcome probabilities. In the end, Cerwick et al. (2014) suggest that studies on the comparison of these two methods be continued so that more conclusive results can lead to a superior method being provided.

The latent class approach, unlike the mixed logit method, uses a finite number of points across a specified number of groups (or classes) with a distribution that is not defined by the analyst (Hensher et al., 2015). That is, the latent class method captures heterogeneity by allowing parameters to vary across unobserved classes without the need of specifying a predefined distribution and is measured according to the mass probability of the intervals between the finite number of points. Recently, the latent class approach has become more prevalent among injury severity analyses and has been illustrated in recent literature. For example, Behnood et al. (2014) utilized the latent class multinomial logit method to determine the effects that age, gender and alcohol consumption have on injury severity. The authors fit several models based on age groups for males and females, as well as the involvement of alcohol. To statistically determine if these sub-populations should be analyzed separately, a transferability test was administered and revealed that distinct age and gender groups based on alcohol consumption should be modeled separately. Shaheed and Gkritza (2014) used a latent class approach to analyze injury severity of single-vehicle motorcycle crashes and found that the latent class method could be an advantageous tool for investigating injury severity due to estimation results and overall model fit, along with less computational effort to that of the mixed logit model. A latent segmentation based ordered logit framework was applied by Eluru et al. (2012) to identify key injury severity factors at highway-railway crossings to model highway-railway crossings by specific attributes and the authors find that two segmentations result in the best overall fit. Further, Xiong and Mannering (2013) utilized a random-parameter finite-mixture (latent class) to investigate the effects of guardian supervision on adolescent injury severity. This framework, specifically, accounts for group-specific heterogeneity (classes) as well as individual observations within classes (Xiong and Mannering, 2013).

Similarly, there have been efforts to compare the latent class approach with traditional injury severity analysis frameworks. Greene and Hensher (2003) compared the latent class and mixed logit frameworks by modeling choice of long distance travel. Although the authors believe the latent class approach is statistically a considerable improvement over the mixed logit approach, the results are data specific and a greater effort towards the comparison of the two frameworks is needed to make a more definite conclusion. The mixed logit and latent class methods were applied by Xie et al. (2012)

to investigate injury severity of single-vehicle crashes in rural areas while comparing estimation results from each approach. The authors used marginal effects to compare the two models and found that the marginal effects are consistent between the two models, yet the latent class method was superior in predicting severity outcomes. Cerwick et al. (2014) utilized the mixed logit and latent class approaches to analyze crash severity of heavy vehicles while comparing the methods based on model fit, marginal effect inferences and predicted severity outcomes. The authors discovered that overall model fit was obtained from the latent class model, yet the prediction accuracy was better for the mixed logit model.

Subsequently, the current study seeks to expand the literature on latent class and mixed logit injury severity framework analyses by identifying heavy vehicle injury severity contributing factors by roadway classification. Previous literature has used age, gender, rural, urban and alcohol consumption as subpopulations, therefore the present study aims to analyze heavy vehicle injury severity utilizing roadway classifications as subpopulations. To statistically determine if road classifications need to be considered separately, a model separation test will be conducted. Additionally, model estimations will be compared using overall fit, marginal effects and predicted injury severity outcomes. Through the identification of injury severity factors by roadway classification, transportation engineers, planners and agencies can better their heavy vehicle safety measures with more precision.

2.2 Empirical Settings

Data used for analysis consisted of police-reported crash data obtained from Idaho for years 2007 to 2013. Each year was filtered by unit type and seat type to represent drivers of heavy vehicles. The data was then combined to create a dataset that included all seven years and the four roadway classifications the largest number of heavy vehicle crashes were determined. The result was a dataset for each road classification of interest: principal arterials, major collectors, interstates and other principal arterials. To ensure that each injury severity included an adequate percentage of the population, injury severities were grouped together to create three distinct severity types: (1) no injury (property-damage-only), (2) minor injury (non-incapacitating and possible injuries) and (3) major injury (incapacitating injuries and fatalities). Grouping the injury severities was permissible, as the Independence of Irrelevant Alternatives property was not violated due to the discovery of random variables in each model. Table 1 shows the injury severity split and total observations for each road classification[1].

[1]The injury severity split on other principal arterials did not allow a model to be fit (e.g. greater than 96% of the observations were no injury crashes). For this reason, contributing factors to heavy vehicle driver injury severity could not be identified for this roadway classification.

Table 2.1: Injury Severity Split by Roadway Classification

Road Classification	Injury Severity	Observations	Percent
Principal Arterials	Property-Damage-Only (No Injuries)	1,334	85.29%
	Minor Injury (Non-Incapacitating Injuries & Possible Injuries)	197	12.60%
	Major Injury (Incapacitating Injuries & Fatalities)	33	2.11%
	Total Observations	**1,564**	
Major Collectors	Property-Damage-Only (No Injuries)	916	86.58%
	Minor Injury (Non-Incapacitating Injuries & Possible Injuries)	117	11.06%
	Major Injury (Incapacitating Injuries & Fatalities)	25	2.36%
	Total Observations	**1,058**	
Interstates	Property-Damage-Only (No Injuries)	1,402	87.13%
	Minor Injury (Non-Incapacitating Injuries & Possible Injuries)	174	10.81%
	Major Injury (Incapacitating Injuries & Fatalities)	33	2.05%
	Total Observations	**1,609**	
Other Principal Arterials	Property-Damage-Only (No Injuries)	1,501	96.22%
	Minor Injury (Non-Incapacitating Injuries & Possible Injuries)	53	3.40%
	Major Injury (Incapacitating Injuries & Fatalities)	6	0.38%
	Total Observations	**1,560**	

Several variables were found to be significant in contributing to outcome probabilities of the three injury severities. Variable descriptions and summary statistics for each road classification are shown in Table 2.2 to Table 2.4.

Table 2.2: Variable Descriptions and Summary Statistics for Principal Arterials

Variable Description	Mean	Standard Deviation
Time of Day (1 if Between 10:00PM and 5:00AM, 0 Otherwise)	0.164	0.371
Vertical Geometrics (1 if Grade, 0 Otherwise)	0.282	0.450
Weather (1 if Cloudy, 0 Otherwise)	0.260	0.439
Speed Limit (1 if Between 55MPH and 65MPH, 0 Otherwise)	0.607	0.489
Crash Location (1 if On Right Shoulder, 0 Otherwise)	0.128	0.334
Point of Impact (1 if Rear Bumper, 0 Otherwise)	0.171	0.376
Protective Device (1 if Seatbelt and Non-Activated Air Bag, 0 Otherwise	0.090	0.286
Horizontal Geometrics (1 if Straight, 0 Otherwise)	0.735	0.442
Traffic Control Device (1 if No Traffic Control Device, 0 Otherwise)	0.701	0.458
Crash Location (1 if On Roadway, 0 Otherwise)	0.786	0.410
Age (1 if Driver Between 35 and 45 Years, 0 Otherwise)	0.222	0.416
Surface Condition (1 if Wet, 0 Otherwise)	0.107	0.309
Time of Week (1 if Weekend, 0 Otherwise)	0.159	0.366
Age (1 if Driver Younger Than 35 Years, 0 Otherwise)	0.254	0.435
Protective Device (1 if Shoulder and Lap Belt, 0 Otherwise)	0.721	0.449
Horizontal Geometrics (1 if Curved, 0 Otherwise)	0.263	0.441

Table 2.3: Variable Descriptions and Summary Statistics for Major Collectors

Variable Description	Mean	Standard Deviation
Horizontal Geometrics (1 if Curve, 0 Otherwise)	0.216	0.412
Heavy Vehicle Type (1 if Tractor 2 Trailer, 0 Otherwise)	0.109	0.311
Contributing Circumstances (1 of No Contributing Circumstances, 0 Otherwise)	0.479	0.500
Point of Impact (1 if Rear Bumper, 0 Otherwise)	0.121	0.326
Time of Year (1 if Winter, 0 Otherwise)	0.256	0.437
City Limits (1 if Crash Occurred Within City Limits, 0 Otherwise)	0.146	0.353
Location of Impact (1 if Curb Line/Off Surface, 0 Otherwise)	0.266	0.442
Protective Device (1 if No Protective Device, 0 Otherwise)	0.123	0.328
Vertical Geometrics (1 if Grade, 0 Otherwise)	0.209	0.407
Time of Day (1 if Between 10:00PM and 5:00AM, 0 Otherwise)	0.086	0.280
Location (1 if Crash Occurred Within 1 Mile of An Intersection, 0 Otherwise)	0.338	0.473
Harmful Event (1 if Overturn, 0 Otherwise)	0.160	0.366
Age (1 if Driver Greater Than 50 Years, 0 Otherwise)	0.353	0.478
Speed Limit (1 if Greater Than 55MPH, 0 Otherwise)	0.173	0.378
Weather (1 if Cloudy, 0 Otherwise)	0.233	0.423

Table 2.4: Variable Descriptions and Summary Statistics for Interstates

Variable Description	Mean	Standard Deviation
Horizontal Geometrics (1 if Curve, 0 Otherwise)	0.233	0.423
Weather (1 if Snow, 0 Otherwise)	0.161	0.368
Heavy Vehicle Type (1 if Tractor 1 Trailer, 0 Otherwise)	0.691	0.462
Protective Device (1 if Shoulder and Lap Belt, 0 Otherwise)	0.759	0.427
Crash Location (1 if On Roadway, 0 Otherwise)	0.764	0.425
Contributing Circumstances (1 if Speeding Contributed to Crash, 0 Otherwise)	0.158	0.365
Location (1 if Crash Occurred Within 1 Mile of On/Off Ramp, 0 Otherwise)	0.737	0.440
Weather (1 if Cloudy, 0 Otherwise)	0.272	0.445
Protective Device (1 if Seatbelt and Non-Activated Air Bag, 0 Otherwise)	0.117	0.321
Age (1 if Driver is Less Than 30 Years, 0 Otherwise)	0.152	0.359
Harmful Event (1 if Sideswipe in Same Direction, 0 Otherwise)	0.236	0.424
Location of Impact (1 if Center Lane or Median, 0 Otherwise)	0.124	0.330
Surface Conditions (1 if Dry, 0 Otherwise)	0.516	0.500
Contributing Circumstances (1 if No Contributing Circumstances, 0 Otherwise)	0.464	0.499
Age (1 if Driver Greater Than 50 Years, 0 Otherwise)	0.375	0.484
Vertical Geometrics (1 if Level, 0 Otherwise)	0.643	0.479

2.3 Methodology

2.3.1 Mixed Logit Model

The first step was to validate the data by generating a two-thirds random sample from the complete dataset. To ensure proper proportioning, the choice variable (injury severity) was selected for strata. A mixed logit model was estimated using the two-thirds sample (the holistic model) and compared to the same model using the remaining one-third of the data. It was determined that the coefficients decreased relative to the population size, hence demonstrating a valid dataset.

The econometric modeling approach allows injury severity outcomes to be modeled as a discrete choice, therefore allowing inference regarding the outcome probabilities of each injury severity to be made. With such inference, the estimated parameters of the mixed logit model provide statistical significance of key factors that increase or decrease the probability of injury severity outcomes.

Models such as the ordered logit model and ordered probit model are capable of modeling crash severity, yet limitations of these exist, as was found by Savolainen and Mannering (2007) (Geedipally et al., 2011; Savolainen and Mannering, 2007). Due to such limitations, the mixed logit model is an improved procedure for investigating injury severities. Conjointly, to further the goodness of the parameter estimations a random-parameter technique can be applied (Gkritza and Mannering, 2008; Islam and Hernandez, 2013; Islam et al., 2014; Morgan and Mannering, 2011; Pahukula et al., 2015).

The mixed logit model begins with a linear function, where each linear function represents an injury severity of a crash and is represented as follows:

$$\mathrm{U_{in}} = \boldsymbol{\beta_i X_{in}} + \varepsilon_{\mathrm{in}} \tag{1}$$

where U_{in} is a linear function for injury severity i and crash n; i represents injury severities of no injury, minor injury and major injury; $\boldsymbol{X_{in}}$ represents the vector of explanatory variables that lead to the discrete outcome of crash n; $\boldsymbol{\beta_i}$ represents the vector of estimated parameters for injury severity i and ε_{in} is the error term that attempts to capture the unobserved factors within the model (Washington et al., 2011); but, ε_{in} is unable to capture all the unobserved factors. Police-reported crash data is often missing key variables and has variation across the available variables resulting in unobserved heterogeneity, and if disregarded, will result is biased estimates and inaccurate inferences (Mannering et al., 2016). Therefore, the mixed logit model attempts to capture this heterogeneity by allowing parameters to vary. In addition, the mixed logit model (if variables are found to be random) eliminates the independence from irrelevant alternatives (IIA) property. In other words, unobserved factors are accounted for if variables are found to be random, so grouping the injury severities into three categories is permitted (Geedipally et al., 2011). The mixed logit model is then formulated as follows (McFadden and Train, 2000; Washington et al., 2011):

$$\mathrm{P_n}(\mathrm{i}|\varphi) = \int \frac{\mathrm{e}^{(\boldsymbol{\beta}_\mathbf{i}\mathbf{X}_\mathbf{in})}}{\sum_{\forall \mathrm{I}} \mathrm{e}^{(\boldsymbol{\beta}_\mathbf{i}\mathbf{X}_\mathbf{in})}} \mathrm{k}(\boldsymbol{\beta}_\mathbf{i}|\varphi)\, \mathrm{d}\beta_\mathrm{i} \tag{2}$$

where $\mathrm{P_n}(\mathrm{i}|\varphi)$ is the weighted outcome probability of injury severity i conditional on $\mathrm{k}(\beta_\mathrm{i}|\varphi)$, where $\mathrm{k}(\beta_\mathrm{i}|\varphi)$ is the density function of $\boldsymbol{\beta}_\mathbf{i}$ and φ with distribution specified by the analyst—the density function is what allows the parameters to vary and is regularly specified to be normally distributed. All other variables have the same definition as the ordinary multinomial logit model (Washington et al., 2011).

2.3.2 Latent Class Multinomial Logit Model

As previously outlined, the latent class multinomial logit model uses a finite number of points and C number of classes to account for unobserved heterogeneity (Greene, 2012a, 2012b). The following is the formulation for outcome probabilities using the latent class multinomial logit model (Behnood et al., 2014; Cerwick et al., 2014; Eluru et al., 2012; Greene and Hensher, 2003; Shaheed and Gkritza, 2014):

$$\mathrm{P_n}(\mathrm{i}|\mathrm{C}) = \frac{e^{(\boldsymbol{\beta}_{\mathbf{iC}}\mathbf{X}_{\mathbf{in}})}}{\sum_{\forall \mathrm{I}} e^{(\boldsymbol{\beta}_{\mathbf{iC}}\mathbf{X}_{\mathbf{in}})}} \tag{3}$$

where $\mathrm{P_n}(\mathrm{i}|\mathrm{C})$ represents the outcome probability of injury severity i for heavy vehicle crash n and C is the unobserved class that accounts for heterogeneity—$\boldsymbol{\beta}$ and $\boldsymbol{X}$ are the same vectors used to formulate the mixed logit model seen in Eq. (1) and Eq. (2). Furthermore, the unconditional class outcome probability, $\mathrm{P_n}(\mathrm{C})$, is also determined by the multinomial logit formulation (Behnood et al., 2014a; Cerwick et al., 2014; Greene and Hensher, 2003):

$$\mathrm{P_n}(\mathrm{C}) = \frac{e^{(\boldsymbol{\omega}_{\mathbf{C}}\mathbf{Z}_{\mathbf{i}})}}{\sum_{\forall \mathrm{C}} e^{(\boldsymbol{\omega}_{\mathbf{C}}\mathbf{Z}_{\mathbf{i}})}} \tag{4}$$

where $\boldsymbol{Z_i}$ represents a vector of class-specific parameters used to determine the probability that heavy vehicle crash n is in class C and $\boldsymbol{\omega_C}$ is a vector of characteristics used to estimate the model parameters (e.g. explanatory/indicator variables) (Behnood et al., 2014a; Cerwick et al., 2014). Lastly, to determine the unconditional outcome probability of injury severity i for heavy vehicle crash n in class C, $\mathrm{P_n}(\mathrm{i})$, the following expression is used (Behnood et al., 2014; Cerwick et al., 2014; Greene and Hensher, 2003; William H Greene, 2012a, 2012b):

$$\mathrm{P_n}(\mathrm{i}) = \sum_{\forall \mathrm{C}} [\mathrm{P_n}(\mathrm{C})][\mathrm{P_n}(\mathrm{i}|\mathrm{C})] \tag{5}$$

2.4 Results

2.4.1 Mixed Logit Results

To determine if the mixed logit model was more significant than the fixed-parameter logit model, a log-likelihood ratio test was conducted for each model. To do this, the following equation was used (Washington et al., 2011):

$$\chi^2 = -2[\mathrm{LL}(\beta_{\mathrm{MNL}}) - \mathrm{LL}(\beta_{\mathrm{MXL}})] \tag{6}$$

where $\mathrm{LL}(\beta_{\mathrm{MNL}})$ is the log-likelihood at convergence of the model with fixed parameters, $\mathrm{LL}(\beta_{\mathrm{MXL}})$ is the log-likelihood at convergence of the model with random parameters, χ^2 is the chi-square statistic and the degrees of freedom for χ^2 being the number of random parameters in β_{MXL}. In doing so, it was determined that each mixed logit model was more significant than that of the fixed models.

With 5 degrees of freedom and a chi-square statistic of 12.97, the mixed logit model is of more significance with 98% confidence for principal arterials; results for the best fit principal arterial model are shown in Table 2.5. The major collector model, with a chi-square statistic of 9.66 and 2 degrees of freedom, should be fit using random parameters with 99.20% confidence; best fit model estimations for the major collector model are shown in Table 2.6. Lastly, a chi-square statistic of 25.65 and 6 degrees of freedom indicate that interstate injury severities should be modeled using random parameters with 99.97% confidence; Table 2.7 illustrates the results for the best fit interstate model.

Table 2.5: Best Fit Model Results & Marginal Effects for Principal Arterials

Variable	Coefficient	*t*-statistic	Marginal Effects		
			No Injury	Minor Injury	Major Injury
Property Damage Only					
Constant	3.21	3.11			
Time of Day (1 if Between 10:00PM and 5:00AM, 0 Otherwise)	-0.68	-1.45	-0.0060	0.0050	0.0010
Weather (1 if Cloudy, 0 Otherwise)	-1.04	-1.88	-0.0139	0.0118	0.0021
Crash Location (1 if Right Shoulder, 0 Otherwise)	0.46	0.82	0.0048	-0.0040	-0.0008
Point of Impact (1 if Rear Bumper, 0 Otherwise)	-1.52	-2.34	-0.0145	0.0119	0.0026
Protective Device (1 if Seatbelt and Non-Activated Air Bag, 0 Otherwise)	2.08	1.93	0.0057	-0.0039	-0.0018
Vertical Geometrics (1 if Grade, 0 Otherwise)	1.05	0.97	-0.0207	0.0160	0.0047
Standard Deviation of Parameter, Normally Distributed	*4.61*	*2.11*			
Speed (1 if Driver Traveling Between 55MPH and 65 MPH, 0 Otherwise)	0.28	0.33	-0.0448	0.0357	0.0091
Standard Deviation of Parameter, Normally Distributed	*3.23*	*1.88*			
Minor Injury					
Constant	2.24	2.32			
Horizontal Geometrics (1 if Straight, 0 Otherwise)	-1.43	-2.21	0.3683	-0.3432	-0.0251
Age (1 if Driver is Between 35 and 45 Years, 0 Otherwise)	-0.90	-1.65	0.0060	-0.0070	0.0009
Surface Condition (1 if Wet, 0 Otherwise)	-1.09	-1.65	0.0036	-0.0042	0.0006
Traffic Control Device (1 if No Control Device, 0 Otherwise)	0.21	0.38	-0.0352	0.0353	-0.0002
Standard Deviation of Parameter, Normally Distributed	*2.12*	*2.00*			
Crash Location (1 if On Roadway, 0 Otherwise)	-4.35	-2.63	0.0559	-0.0607	0.0049
Standard Deviation of Parameter, Normally Distributed	*2.57*	*1.89*			
Major Injury					
Time of Week (1 if Weekend, 0 Otherwise)	2.59	2.63	-0.0072	-0.0040	0.0112
Age (1 if Driver is Less Than 35 Years, 0 Otherwise)	-2.38	-1.95	0.0014	0.0011	-0.0026
Protective Device (1 if Lap and Shoulder Belt, 0 Otherwise)	-2.68	-3.18	0.0085	0.0061	-0.0146
Horizontal Geometrics (1 if Curved, 0 Otherwise)	1.57	1.84	-0.0040	-0.0033	0.0073
Crash Location (1 if On Roadway, 0 Otherwise)	-5.94	-2.08	0.0066	0.0034	-0.0100
Standard Deviation of Parameter, Normally Distributed	*2.96*	*1.75*			
Model Statistics					
Number of Observations	1,564				
Restricted Log-Likelihood	-1,718.23				
Log-Likelihood at Convergence	-647.78				
McFadden Pseudo R-Squared	0.62				

Table 2.6: Best Fit Model Results & Marginal Effects for Major Collectors

Variable	Coefficient	*t*-statistic	Marginal Effects		
			No Injury	Minor Injury	Major Injury
No Injury					
Constant	6.57	8.82			
Horizontal Geometrics (1 if Curve, 0 Otherwise)	-0.92	-3.10	-0.0235	0.0188	0.0047
Heavy Vehicle Type (1 if Tractor 2 Trailer, 0 Otherwise)	2.53	2.99	0.0062	-0.0052	-0.0009
Contributing Circumstances (1 if No Contributing Circumstances, 0 Otherwise)	0.90	3.00	0.0190	-0.0157	-0.0034
Point of Impact (1 if Rear Bumper, 0 Otherwise)	2.21	0.86	-0.0106	0.0086	0.0019
Standard Deviation of Parameter, Normally Distributed	*4.78*	*1.66*			
Minor Injury					
Constant	4.00	5.31			
Time of Year (1 if Winter, 0 Otherwise)	-0.78	-2.04	0.2681	-0.2419	-0.0261
City Limits (1 if Crash Occurred Within City Limits, 0 Otherwise)	-2.06	-2.53	0.0042	-0.0045	0.0003
Protective Device (1 if No Protective Device, 0 Otherwise)	1.25	3.48	-0.0135	0.0167	-0.0032
Vertical Geometrics (1 if Grade, 0 Otherwise)	0.57	1.81	-0.0105	0.0114	-0.0009
Lane of Impact (1 if Curb Line/Off Surface, 0 Otherwise)	0.72	1.38	-0.0412	0.0441	-0.0029
Standard Deviation of Parameter, Normally Distributed	*2.09*	*2.26*			
Major Injury					
Time of Day (1 if Between 10:00PM and 5:00AM, 0 Otherwise)	1.84	2.84	-0.0049	-0.0020	0.0070
Location (1 if Crash Occurred Within 1 Mile of An Intersection, 0 Otherwise)	1.00	2.03	-0.0069	-0.0027	0.0096
Harmful Event (1 if Overturn, 0 Otherwise)	2.18	4.23	-0.0158	-0.0061	0.0219
Age (1 if Driver Greater Than 50 Years, 0 Otherwise)	1.55	2.95	-0.0108	-0.0038	0.0145
Protective Device (1 if No Protective Device, 0 Otherwise)	2.40	4.69	-0.0137	-0.0062	0.0200
Speed Limit (1 if Greater Than 55MPH, 0 Otherwise)	0.93	1.70	-0.0034	-0.0012	0.0046
Weather (1 if Cloudy, 0 Otherwise)	-1.57	-1.79	0.0016	0.0006	-0.0022
Model Statistics					
Number of Observations	1,058				
Restricted Log-Likelihood	-1,162.33				
Log-Likelihood at Convergence	-371.63				
McFadden Pseudo R-Squared	0.68				

Table 2.7: Best Fit Model Results & Marginal Effects for Interstates

Variable	Coefficient	*t*-statistic	Marginal Effects		
			No Injury	Minor Injury	Major Injury
No Injury					
Constant	3.01	3.55			
Horizontal Geometrics (1 if Curve, 0 Otherwise)	-0.96	-2.80	-0.0156	0.0125	0.0030
Heavy Vehicle Type (1 if Tractor 1 Trailer, 0 Otherwise)	0.44	1.59	0.0160	-0.0135	-0.0025
Event Location (1 if On Roadway, 0 Otherwise)	2.19	5.60	0.0724	-0.0625	-0.0099
Contributing Circumstances (1 if Speeding Contributed to Crash, 0 Otherwise)	-1.14	-2.63	-0.0121	0.0101	0.0019
Weather (1 if Snow, 0 Otherwise)	3.75	2.04	-0.0030	0.0030	0.0000
Standard Deviation of Parameter, Normally Distributed	*4.05*	*2.56*			
Protective Device (1 if Shoulder and Lap Belt, 0 Otherwise)	2.01	3.05	0.0286	-0.0238	-0.0048
Standard Deviation of Parameter, Normally Distributed	*1.93*	*2.61*			
Minor Injury					
Constant	3.06	3.64			
Location (1 if Crash Occurred Within 1 Mile of On/Off Ramp, 0 Otherwise)	0.65	1.91	-0.1643	0.1495	0.0148
Protective Device (1 if Seatbelt and Non-Activated Air Bag, 0 Otherwise)	-1.82	-3.61	0.0090	-0.0090	0.0001
Harmful Event (1 if Sideswipe in Same Direction, 0 Otherwise)	-0.98	-2.26	0.0062	-0.0064	0.0003
Weather (1 if Cloudy, 0 Otherwise)	-4.88	-1.96	-0.0076	0.0076	0.0001
Standard Deviation of Parameter, Normally Distributed	*5.08*	*2.39*			
Age (1 if Driver Less Than 30 Years, 0 Otherwise)	-2.96	-1.72	-0.0032	0.0030	0.0002
Standard Deviation of Parameter, Normally Distributed	*3.71*	*2.26*			
Major Injury					
Surface Conditions (1 if Dry, 0 Otherwise)	2.18	2.52	-0.0136	-0.0048	0.0185
Contributing Circumstances (1 if No Contributing Circumstances, 0 Otherwise)	-1.87	-2.17	0.0026	0.0007	-0.0033
Protective Device (1 if Seatbelt and Non-Activated Air Bag, 0 Otherwise)	-4.81	-1.49	0.0007	0.0002	-0.0009
Vertical Geometrics (1 if Level, 0 Otherwise)	-1.39	-2.09	0.0051	0.0018	-0.0069
Lane of Impact (1 if Center Lane or Median, 0 Otherwise)	-1.75	-0.68	-0.0065	-0.0017	0.0081
Standard Deviation of Parameter, Normally Distributed	*4.46*	*2.03*			
Age (1 if Driver Greater Than 50 Years, 0 Otherwise)	-0.28	-0.18	-0.0114	-0.0030	0.0144
Standard Deviation of Parameter, Normally Distributed	*2.53*	*1.76*			
Model Statistics					
Number of Observations	1,609				
Restricted Log-Likelihood	-1767.67				
Log-Likelihood at Convergence	-605.69				
McFadden Pseudo R-Squared	0.68				

The following log-likelihood ratio test was used to validate separating the models by roadway classification (Washington et al., 2011):

$$\chi^2 = -2[\mathrm{LL}(\beta_{\mathrm{TOT}}) - \mathrm{LL}(\beta_{\mathrm{PA}}) - \mathrm{LL}(\beta_{\mathrm{MACOL}}) - \mathrm{LL}(\beta_{\mathrm{INT}})] \tag{7}$$

where $\mathrm{LL}(\beta_{\mathrm{TOT}})$ is the log-likelihood at convergence for the holistic model (model with the two-thirds dataset), $\mathrm{LL}(\beta_{\mathrm{PA}})$ is the log-likelihood at convergence of the principal arterial model, $\mathrm{LL}(\beta_{\mathrm{MACOL}})$ is the log-likelihood at convergence of the major collector model and $\mathrm{LL}(\beta_{\mathrm{INT}})$ is the log-likelihood at convergence of the interstate model. Applying Eq. (7) results in a chi-square statistic of 1,083.57 with the degrees of being the total number of estimated parameters in the three road classification models (Washington et al., 2010). Therefore, with a chi-square statistic of 1,083.57 and 49 degrees of freedom, road classifications should be modeled separately with 99.99% confidence.

The final step in confirming if road classifications are to be modeled separately was to conduct a model separation test. In regards to the model separation test, the assumption of equal variances created convergence problems within the constants due to large amounts of heterogeneity and generated highly erroneous constant estimates and log-likelihood values (Greene, 2012b). To avoid such bias, the heteroscedastic extreme value multinomial logit model was used to conduct the separation test (Greene, 2012b). Using the log-likelihood values obtained from the heteroscedastic extreme value multinomial logit model, the following log-likelihood ratio test was applied (Washington et al., 2011):

$$\chi^2 = -2\left[LL\left(\beta_{X_{1_{X_2}}}\right) - LL\left(\beta_{X_1}\right)\right] \tag{8}$$

where $LL(\beta_{X_1})$ is the log-likelihood at convergence of model X_1 and $LL\left(\beta_{X_{1_{X_2}}}\right)$ is the log-likelihood at convergence of model X_1 using the data from model X_2. For example, the best fit mixed logit model for principal arterials provides beta and constant estimates (model X_1). The beta and constant values are then fixed and the same principal arterial model is ran using the data from major collectors (model X_2)—this output is $LL\left(\beta_{X_{1_{X_2}}}\right)$. The chi-square statistics and degrees of freedom[2] for this log-likelihood ratio test are shown Table 2.8. Each chi-square statistic and its corresponding degrees of freedom further illustrate, with 99.99% confidence, that road classifications should be modeled independently.

Table 2.8: Chi-Square Statistics & Degrees of Freedom for Road Classification Separation Test

X_1	X_2		
	Principal Arterial	Major Collector	Interstate
Principal Arterial	0	359.67 (12)	70.31 (12)
Major Collector	647.02 (10)	0	172.50 (10)
Interstate	250.39 (12)	358.96 (10)	0

2.4.2 Latent Class Results

The best fit mixed logit models were fit as latent class multinomial logit models to compare overall model fit, marginal effects and the number of correctly predicted

[2] The degrees of freedom for this log-likelihood ratio test is equal to the number of variables in model X_1 using the data from model X_2

injury severities. Although several previous studies (Collins et al., 1993; Hagenaars and McCutcheon, 2002; Magidson and Vermunt, 2004; Nylund et al., 2007) state that the best fit number of latent classes should be selected using the Bayesian Information Criterion (BIC), others suggest that that the number of classes should be selected based on the Akaike Information Criterion (AIC) (Louviere et al., 2000; Shen, 2009). For this study, the best fit number of latent classes are based on the AIC—for these larger sample sizes, the AIC was smallest for each number of latent classes investigated. Table 2.9, Table 2.10 and Table 2.11 display the latent class model specifications and Table 2.12, Table 2.13 and Table 2.14 show the corresponding marginal effects.

Table 2.9: Latent Class Multinomial Logit Model Results for Principal Arterials

Variable	Latent Class 1						Latent Class 2					
	No Injury		Minor Injury		Major Injury		No Injury		Minor Injury		Major Injury	
	Beta	*t*-statistic	Beta	*t*-statistic	Beta	*t*-statistic	Beta	*t*-statistic	Beta	*t*-statistic	Beta	*t*-statistic
Constant												
No Injury	1.23	1.82	-	-	-	-	97.88	0.00	-	-	-	-
Minor Injury	-	-	-	-	-	-	-	-	-	-	-	-
Crash Characteristics	-	-	-	-	-	-	-	-	-	-	-	-
Event Location (1 if On Right Shoulder, 0 Otherwise)	-	-	-	-	-	-	-	-	-	-	-	-
Point of Impact (1 if Rear Bumper, 0 Otherwise)	-0.61	-1.31	-	-	-	-	-1.60	-2.25	-	-	-	-
Event Location (1 if On Roadway, 0 Otherwise)	-	-	-	-	-	-	-	-	-	-	-	-
Roadway & Environmental Characteristics	-	-	-	-	-	-	-	-	-	-	-	-
Vertical Geometrics (1 if Grade, 0 Otherwise)	0.16	0.35	-	-	-	-	-1.37	-2.00	-	-	-	-
Weather (1 if Cloudy, 0 Otherwise)	-0.08	-0.19	-	-	-	-	-1.31	-2.32	-	-	-	-
Speed Limit (1 if Greater Than 55MPH and Less Than or Equal to 65 MPH, 0 Otherwise)	-0.05	-0.12	-	-	-	-	-1.31	-2.30	-	-	-	-
Horizontal Geometrics (1 if Straight, 0 Otherwise)	-	-	-0.41	0.33	-	-	-	-	-1.97	-2.33	-	-
Traffic Control Device (1 if No Control Device, 0 Otherwise)	-	-	-	-	-	-	-	-	-	-	-	-
Surface Condition (1 if Wet, 0 Otherwise)	-	-	0.31	0.53	-	-	-	-	-1.75	-2.52	-	-
Horizontal Geometrics (1 if Curved, 0 Otherwise)	-	-	-	-	1.11	1.73	-	-	-	-	1.98	2.03
Driver Characteristics	-	-	-	-	-	-	-	-	-	-	-	-
Protective Device (1 if Seatbelt and Non-Activated Air Bag, 0 Otherwise)	-	-	-	-	-	-	-	-	-	-	-	-
Age (1 if Driver is Between 35 and 45 Years, 0 Otherwise)	-	-	-0.36	-0.75	-	-	-	-	-0.71	-1.61	-	-
Age (1 if Driver is Less Than 35 Years, 0 Otherwise)	-	-	-	-	-1.94	-2.03	-	-	-	-	-1.67	-1.55
Protective Device (1 if Lap and Shoulder Belt, 0 Otherwise)	-	-	-	-	-2.76	-4.34	-	-	-	-	-0.92	-1.53
Temporal & Spatial Characteristics	-	-	-	-	-	-	-	-	-	-	-	-
Time of Day (1 if Between 10:00 p.m. and 5:00 a.m., 0 Otherwise)	-0.53	-1.2	-	-	-	-	-1.15	-1.72	-	-	-	-
Time of Week (1 if Weekend, 0 Otherwise)	-	-	-	-	1.62	2.28	-	-	-	-	1.66	2.74
Class Probability (*t*-statistic)	0.795 (22.61)						0.205 (5.84)					
Number of Observations	1,564											
Log-Likelihood at Zero	-1718.23											
Log-Likelihood at Convergence	-638.98											
Adjusted R-Squared	0.62											
Akaike Information Criterion	1356.00											

Table 2.10: Latent Class Multinomial Logit Model Results for Major Collectors

Variable	Latent Class 1						Latent Class 2					
	No Injury		Minor Injury		Major Injury		No Injury		Minor Injury		Major Injury	
	Beta	*t*-statistic	Beta	*t*-statistic	Beta	*t*-statistic	Beta	*t*-statistic	Beta	*t*-statistic	Beta	*t*-statistic
Constant												
No Injury	66.75	0.00	-	-	-	-	6.30	4.05	-	-	-	-
Minor Injury	-	-	63.39	0.00	-	-	-	-	5.08	3.37	-	-
Crash Characteristics												
Contributing Circumstances (1 if No Contributing Circumstances, 0 Otherwise)	-	-	-	-	-	-	-	-	-	-	-	-
Point of Impact (1 if Rear Bumper, 0 Otherwise)	0.51	0.55	-	-	-	-	-4.02	-2.87	-	-	-	-
Lane of Impact (1 if Curb Line/Off Surface, 0 Otherwise)	-	-	2.27	3.09	-	-	-	-	-0.61	-0.47	-	-
Harmful Event (1 if Overturn, 0 Otherwise)	-	-	-	-	-	-	-	-	-	-	-	-
Roadway & Environmental Characteristics												
Horizontal Geometrics (1 if Curved, 0 Otherwise)	-0.99	-2.23	-	-	-	-	-0.39	-0.38	-	-	-	-
Vertical Geometrics (1 if Grade, 0 Otherwise)	-	-	0.06	0.12	-	-	-	-	1.87	1.67	-	-
Speed Limit (1 if Greater Than or Equal to 55 MPH, 0 Otherwise)	-	-	-	-	-	-	-	-	-	-	-	-
Weather (1 if Cloudy, 0 Otherwise)	-	-	-	-	-	-	-	-	-	-	-	-
Driver Characteristics												
Protective Device (1 if No Protective Device, 0 Otherwise)	-	-	0.00	1.00	-	-	-	-	3.35	2.76	2.69	1.87
Age (1 if Driver Greater Than 50 Years, 0 Otherwise)	-	-	-	-	1.70	1.41	-	-	-	-	2.72	2.68
Vehicle Characteristics												
Heavy Vehicle Type (1 if Tractor 2-Trailor, 0 Otherwise)	-	-	-	-	-	-	-	-	-	-	-	-
Temporal & Spatial Characteristics												
Time of Year (1 if Winter, 0 Otherwise)	-	-	-0.01	-0.01	-	-	-	-	-2.94	-1.90	-	-
City Limits (1 if Inside City Limits, 0 Otherwise)	-	-	-	-	-	-	-	-	-	-	-	-
Time of Day (1 if Between 10:00PM and 5:00AM, 0 Otherwise)	-	-	-	-	1.24	0.62	-	-	-	-	2.42	2.06
Location (1 if Crash Occurred Within 1 Mile of An Intersection, 0 Otherwise)	-	-	-	-	-1.07	-0.97	-	-	-	-	3.12	2.81
Class Probability (*t*-statistic)	0.789 (12.79)						0.211 (3.43)					
Number of Observations	1,058											
Log-Likelihood at Zero	-1162.33											
Log-Likelihood at Convergence	-359.18											
Adjusted R-Squared	0.69											
Akaike Information Criterion	792.40											

Table 2.11: Latent Class Multinomial Logit Model Results for Interstates

Variable	Latent Class 1						Latent Class 2					
	No Injury		Minor Injury		Major Injury		No Injury		Minor Injury		Major Injury	
	Beta	*t*-statistic	Beta	*t*-statistic	Beta	*t*-statistic	Beta	*t*-statistic	Beta	*t*-statistic	Beta	*t*-statistic
Constant												
Property Damage Only Crash	31.10	0.00	-	-	-	-	3.30	3.12	-	-	-	-
Minor Injury Crash	-	-	30.58	0.00	-	-	-	-	1.99	2.74	-	-
Crash Characteristics												
Event Location (1 if On Roadway, 0 Otherwise)	5.62	4.65	-	-	-	-	-1.87	-2.02	-	-	-	-
Contributing Circumstances (1 if Speeding Contributed to Crash, 0 Otherwise)	0.85	1.19	-	-	-	-	-1.66	-2.80	-	-	-	-
Harmful Event (1 if Sideswipe in Same Direction, 0 Otherwise)	-	-	2.67	2.34	-	-	-	-	-	-	-	-
Lane of Impact (1 if Center Lane or Median, 0 Otherwise)	-	-	-	-	0.71	1.12	-	-	-	-	1.76	2.37
Contributing Circumstances (1 if No Contributing Circumstances, 0 Otherwise)	-	-	-	-	-0.16	-0.17	-	-	-	-	-1.86	-2.01
Roadway & Environmental Characteristics												
Horizontal Geometrics (1 if Curved, 0 Otherwise)	-0.57	-1.51	-	-	-	-	-0.93	-2.07	-	-	-	-
Weather (1 if Snow, 0 Otherwise)	3.70	2.53	-	-	-	-	-0.43	-0.80	-	-	-	-
Weather (1 if Cloudy, 0 Otherwise)	-	-	-1.27	-3.10	-	-	-	-	-0.11	-0.27	-	-
Surface Condition (1 if Dry, 0 Otherwise)	-	-	-	-	-	-	-	-	-	-	-	-
Vertical Geometrics (1 if Level, 0 Otherwise)	-	-	-	-	-0.65	-1.04	-	-	-	-	-1.20	-1.91
Vehicle Characteristics												
Heavy Vehicle Type (1 if Tractor 1-Trailer, 0 Otherwise)	-0.36	-0.95	-	-	-	-	0.98	2.48	-	-	-	-
Driver Characteristics												
Protective Device (1 if Shoulder and Lap Belt, 0 Otherwise)	0.69	1.56	-	-	-	-	1.51	2.59	-	-	-	-
Protective Device (1 if Seatbelt and Non-Activated Air Bag, 0 Otherwise)	-	-	-1.08	-1.66	-	-	-	-	-2.28	-2.81	-2.71	-2.05
Age (1 if Driver Younger Than 30 Years, 0 Otherwise)	-	-	-	-	-	-	-	-	-	-	-	-
Age (1 if Driver Older Than 50 Years, 0 Otherwise)	-	-	-	-	0.86	1.42	-	-	-	-	1.25	2.07
Temporal & Spatial Characteristics												
Location (1 if Crash Occurred Within 1 Mile of On/Off Ramp, 0 Otherwise)	-	-	-	-	-	-	-	-	-	-	-	-
Class Probability (*t*-statistic)			0.718 (9.12)						0.282 (3.58)			
Number of Observations	1,609											
Log-Likelihood at Zero	-1,767.67											
Log-Likelihood at Convergence	-581.75											
Adjusted	0.67											
Akaike Information Criterion	1241.50											

Table 2.12: Estimated Marginal Effects of Latent Class Principal Arterial Model

Variable	Severity Function	Marginal Effects on Outcome Probabilities of Injury Severity		
		No Injury	Minor Injury	Major Injury
Time of Day (1 if Between 10:00PM and 5:00AM, 0 Otherwise)	No Injury	-0.0094	0.0068	0.0005
Weather (1 if Cloudy, 0 Otherwise)	No Injury	-0.0110	0.0068	0.0008
Crash Location (1 if Right Shoulder, 0 Otherwise)	No Injury	0.0118	-0.0104	-0.0004
Point of Impact (1 if Rear Bumper, 0 Otherwise)	No Injury	-0.0112	0.0075	0.0011
Protective Device (1 if Seatbelt and Non-Activated Air Bag, 0 Otherwise)	No Injury	0.0006	-0.0002	-0.0001
Vertical Geometrics (1 if Grade, 0 Otherwise)	No Injury	-0.0097	0.0064	0.0004
Speed (1 if Driver Traveling Between 55MPH and 65 MPH, 0 Otherwise)	No Injury	-0.0284	0.0170	0.0016
Horizontal Geometrics (1 if Straight, 0 Otherwise)	Minor	-1.6490	1.5505	-0.0986
Age (1 if Driver is Between 35 and 45 Years, 0 Otherwise)	Minor	0.0057	-0.0047	0.0006
Surface Condition (1 if Wet, 0 Otherwise)	Minor	0.0027	-0.0019	0.0003
Traffic Control Device (1 if No Control Device, 0 Otherwise)	Minor	-0.0167	0.0150	-0.0018
Crash Location (1 if On Roadway, 0 Otherwise)	Minor	0.0050	-0.0034	0.0007
Time of Week (1 if Weekend, 0 Otherwise)	Major	-0.0050	-0.0034	0.0054
Age (1 if Driver is Less Than 35 Years, 0 Otherwise)	Major	0.0016	0.0012	-0.0018
Protective Device (1 if Lap and Shoulder Belt, 0 Otherwise)	Major	0.0077	0.0044	-0.0058
Horizontal Geometrics (1 if Curved, 0 Otherwise)	Major	-0.0043	-0.0046	0.0066
Crash Location (1 if On Roadway, 0 Otherwise)	Major	-0.3495	-0.3172	0.4771

Table 2.13: Estimated Marginal Effects of Latent Class Major Collector Model

Variable	Severity Function	Marginal Effects on Outcome Probabilities of Injury Severity		
		No Injury	Minor Injury	Major Injury
Horizontal Geometrics (1 if Curve, 0 Otherwise)	No Injury	-0.0240	0.0220	0.0016
Heavy Vehicle Type (1 if Tractor 2 Trailer, 0 Otherwise)	No Injury	0.0036	-0.0016	-0.0002
Contributing Circumstances (1 if No Contributing Circumstances, 0 Otherwise)	No Injury	0.0200	-0.0162	-0.0012
Point of Impact (1 if Rear Bumper, 0 Otherwise)	No Injury	-0.0098	0.0060	0.0013
Time of Year (1 if Winter, 0 Otherwise)	Minor	0.0329	-0.0307	-0.0022
City Limits (1 if Crash Occurred Within City Limits, 0 Otherwise)	Minor	0.0025	-0.0025	0.0000
Protective Device (1 if No Protective Device, 0 Otherwise)	Minor	-0.0082	0.0121	-0.0039
Vertical Geometrics (1 if Grade, 0 Otherwise)	Minor	-0.0108	0.0121	-0.0009
Lane of Impact (1 if Curb Line/Off Surface, 0 Otherwise)	Minor	-0.0691	0.0707	-0.0017
Time of Day (1 if Between 10:00PM and 5:00AM, 0 Otherwise)	Major	-0.0036	-0.0016	0.0029
Location (1 if Crash Occurred Within 1 Mile of An Intersection, 0 Otherwise)	Major	-0.0092	-0.0060	0.0059
Harmful Event (1 if Overturn, 0 Otherwise)	Major	-0.1077	-0.0352	0.1411
Age (1 if Driver Greater Than 50 Years, 0 Otherwise)	Major	-0.0097	-0.0041	0.0071
Protective Device (1 if No Protective Device, 0 Otherwise)	Major	-0.1058	-0.0367	0.1407
Speed Limit (1 if Greater Than 55MPH, 0 Otherwise)	Major	-0.0012	-0.0005	0.0012
Weather (1 if Cloudy, 0 Otherwise)	Major	-0.0001	-0.0001	0.0001

Table 2.14: Estimated Marginal Effects of Latent Class Interstate Model

Variable	Severity Function	Marginal Effects on Outcome Probabilities of Injury Severity		
		No Injury	Minor Injury	Major Injury
Horizontal Geometrics (1 if Curve, 0 Otherwise)	No Injury	-0.0142	0.0106	0.0021
Heavy Vehicle Type (1 if Tractor 1 Trailer, 0 Otherwise)	No Injury	0.0145	-0.0132	-0.0010
Event Location (1 if On Roadway, 0 Otherwise)	No Injury	-0.0317	0.0422	0.0045
Contributing Circumstances (1 if Speeding Contributed to Crash, 0 Otherwise)	No Injury	-0.0122	0.0097	0.0012
Weather (1 if Snow, 0 Otherwise)	No Injury	0.0001	0.0004	0.0001
Protective Device (1 if Shoulder and Lap Belt, 0 Otherwise)	No Injury	0.0584	-0.0472	-0.0071
Location (1 if Crash Occurred Within 1 Mile of On/Off Ramp, 0 Otherwise)	Minor	0.0140	-0.0053	-0.0087
Protective Device (1 if Seatbelt and Non-Activated Air Bag, 0 Otherwise)	Minor	0.0099	-0.0100	0.0002
Harmful Event (1 if Sideswipe in Same Direction, 0 Otherwise)	Minor	-0.0124	0.0029	-0.0002
Weather (1 if Cloudy, 0 Otherwise)	Minor	0.0088	-0.0084	0.0005
Age (1 if Driver Less Than 30 Years, 0 Otherwise)	Minor	0.0048	-0.0049	0.0004
Surface Conditions (1 if Dry, 0 Otherwise)	Major	-0.1539	-0.0906	0.2452
Contributing Circumstances (1 if No Contributing Circumstances, 0 Otherwise)	Major	0.0019	0.0004	-0.0014
Protective Device (1 if Seatbelt and Non-Activated Air Bag, 0 Otherwise)	Major	0.0010	0.0004	-0.0006
Vertical Geometrics (1 if Level, 0 Otherwise)	Major	0.0046	0.0025	-0.0059
Lane of Impact (1 if Center Lane or Median, 0 Otherwise)	Major	-0.0042	-0.0027	0.0050
Age (1 if Driver Greater Than 50 Years, 0 Otherwise)	Major	-0.0064	-0.0032	0.0078

2.5 Discussion

2.5.1 Mixed Logit

A total of 36 variables were found to be significant throughout the three road classification models, with just two variables being significant in each model. The two variables found to be significant for all three road classification models were cloudy weather conditions and horizontal curves. Cloudy weather conditions decrease the probability of an injury severity for each road classification. For instance, cloudy weather decreases the likelihood of no injury on principal arterials, reduces the probability of a major injury on major collectors and decreases the outcome probability of a minor injury on interstates. In addition, cloudy weather conditions was found to be random and normally distributed with a mean of -4.88 and standard deviation of 5.08 on interstates. This indicates that for 16.7% of heavy vehicles the estimated parameter mean is greater than zero, while 83.3% of heavy vehicles have an estimated parameter mean less than zero. That is to say, cloudy conditions increase the probability of sustaining a minor injury on interstates for 16.7% of heavy vehicles, yet decrease minor injury probability 83.3%.

Cloudy conditions have previously been found to influence injury severity. For instance, Kim et al. (2013) found cloudy conditions to increase the probability of fatal crashes in California, however, Mohamed et al. (2013) found that cloudy conditions reduce the probability of fatal crashes in New York and Montreal. A possible explanation for injury severity reduction in New York, Montreal, and now Idaho, could be attributed to driver experience in cloudy conditions. As for the second variable found to be significant for all classifications, horizontal curves reduce no injury probability for major collectors and interstates, but increase the likelihood of a major injury on principal arterials.

Seven variables were found to be significant between at least two of the road classifications. Major collectors and interstates have one shared significant variable, principal arterials and major collectors have three shared significant variables, and principal arterials and interstates have three shared significant variables. Of these seven variables, four were found to be random. Crashes happening on the roadway (e.g. not on the shoulder or median) were found to be random and normally distributed for minor injury and major injury crashes on principal arterials. Particularly, for minor injury crashes, a mean of -4.35 and standard deviation of 2.57 suggests that for 4.5% of heavy vehicles the estimated parameter mean is greater than zero and less than zero for 95.5%. Further, major injury crashes have a mean of -5.94 and standard deviation of 2.96, therefore increasing the probability of a major injury for 2.2% of heavy vehicles and decreasing the probability for 97.8%. A decrease in major injuries for crashes occurring on the roadway was also found by Xie et al. (2012).

Crashes that happened on a grade were found to be random and normally distributed on principal arterials for no injury crashes. With a mean of 1.05 and standard deviation of 4.61, 41% of heavy vehicles have a decrease in no injury probability and 59% of heavy vehicles have an increase in no injury probability. The percent grade is not given in the data, or if the crash occurred while traveling up/down grade, therefore the randomness in this variable may be accounting for such characteristics.

Crashes with the rear bumper being the point of impact were found to be random and normally distributed on major collectors for no injury crashes. Rear bumper impact, with a mean of 2.21 and standard deviation of 4.78, indicates that 32.2% of heavy vehicles have a decrease in no injury probability and 67.8% of heavy vehicles have an increase in no injury probability. In regards to rear bumper impact, posted speed limits were provided in the data but speed at the time of the crash was not. With this in mind, rear bumper impact may be random in an attempt to capture the speed at which the impact took place—lower speeds are apt to result in less severe injuries while higher speeds can lead to more severe injuries.

Drivers wearing a seatbelt was found to be random and normally distributed for no injury crashes on interstates. A mean of 2.01 and standard deviation of 1.93 suggest that for 14.9% of heavy the probability of a no injury crash decreases, but increases the probability of no injury crashes for 85.1% of heavy vehicles. Although seatbelts are designed to save lives, they can also cause injury and this is demonstrated by the randomness in this variable. Recent studies (see Kashani and Mohaymany, 2011; Obeng, 2011; Russo et al., 2014; Ye and Lord, 2014) have also found seatbelt usage to impact injury severity outcomes.

Drivers under the age of 30 were found to be random and normally distributed on interstates. A mean of -1.72 and standard deviation of 2.29 propose that for 22.6% of heavy vehicles the probability of suffering a minor injury increases, but decreases the probability of a minor injury for 77.4% of heavy vehicles—this is also found in previous injury severity studies (Islam and Hernandez, 2013; Pahukula et al., 2015; Yasmin et al., 2014). Snowy weather conditions on interstates was found to be random and normally distributed for no injury crashes. Snowy weather, with a mean of 3.97 and standard deviation of 4.08, decreases the outcome probability of no injury for 16.5% of heavy vehicles, but increases the probability for 83.5% of heavy vehicles. This differs from some previous findings (i.e. Yasmin et al., 2014) in which the winter months (when snowy weather is present) increase the probability of major injury crashes. The difference in findings could be credited to Idaho drivers being more experienced with driving in inclement weather conditions.

Crashes occurring on weekends was significant exclusively for principal arterials and increases the probability of a major injury crash. This finding could be a result of heavy vehicle drivers maneuvering with more risk due to the lack of typical commuter traffic seen on principal arterials during the week. Driving a tractor 2-trailer increases the outcome probability of no injury crashes and is exclusive to major collectors. Islam et al. (2014) also found multiple unit heavy vehicles to effect injury severity, however it was found to increase minor injury probability. Interstates, exclusively, see an increase in outcome probability of no injury crashes when a tractor 1-trailer is involved. If tractor 2-trailers and tractor 1-trailers are involved in crashes that occur during congested conditions, it is probable that these crashes happen at very low speeds, thus despite the size of the vehicle less severe injuries may be expected.

Marginal effect values for principal arterials indicate that straight horizontal geometrics have the greatest effect on injury severity. Marginal effects show that principal arterials with straight horizontal geometrics have a 0.368 higher probability of no injury, and a 0.343 and 0.025 lower probability of a minor injury and major injury, respectively. On major collectors, crashes happening in the winter have the greatest impact on injury severity. Marginal effects reveal that winter crashes have a 0.268 higher probability of no injury, yet a 0.242 and 0.026 lower probability of minor and major injuries. Interstates see the location of the crash having the greatest effect on injury severity. For crashes that happened within 1 mile of an on/off ramp, marginal effects indicate a 0.164 lower probability of no injury while showing a 0.150 and 0.015 higher probability for minor injury and major injury crashes.

2.5.2 Latent Class

Upon fitting a best fit mixed logit model, the same variables were used to fit a latent class multinomial logit model. For the latent class model, two classes were found to be the best fit number of classes for each road classification using the AIC (Louviere et al., 2000; Shen, 2009). At first glance, the latent class approach appears to account for more of the heterogeneity within the dataset. For example, several variables in the principal arterial model are significant in one class and not the other, have the same sign in Class 1 and Class 2, or are not significant in either class. These findings are in line with Shaheed and Gkritza (2014) and indicate that a large amount of heterogeneity has been accounted for between Class 1 and Class 2. Among the variables significant in one class and not the other are cloudy weather conditions, crashes where the point of impact was the rear bumper and drivers younger than 35 years. Variables with the same sign include a posted speed limit between 55 miles per hour and 65 miles per hour, curved horizontal geometrics and crashes that happened between the hours of 10:00 p.m. and 5:00 a.m. Finally, variables insignificant in Class 1 and Class 2 include crashes that took place on the right shoulder and no traffic control device present. Although the variables are different, the same observations are seen in the major collector model and the interstate model.

Marginal effects for the latent class models show that other factors have the largest effect on injury severity outcome, increase the effect of a variable found to have the largest impact in the mixed logit models, or have the opposite effect on injury severity. For example, the principal arterial mixed logit marginal effects indicated that straight horizontal geometrics have the greatest effect on injury severity. This was also found to be true in the latent class model, but marginal effects show that principal arterials with straight horizontal geometrics have a 1.65 lower probability of no injury, 1.55 higher probability of a minor injury and a 0.099 lower probability of a major injury compared to the mixed logit approach. In addition, the latent class marginal effects found that crashes occurring on the roadway have a 0.477 higher probability of a major injury, yet have a 0.350 and 0.317 lower probability of no injury and minor injury.

On major collectors, the latent class marginal effects show that crashes in which the heavy vehicle overturned and crashes in which no protective device was used impact injury severity the greatest. Specifically, overturned heavy vehicles have a 0.141 higher probability of a major injury while no injury and minor injury have a 0.108 and 0.035 lower probability, respectively. For crashes that no protective device was worn, the latent class marginal effects show a lower probability for no injury and minor injury tantamount to the effect of an overturned vehicle—the same is true for major injury. The lane of impact also impacts injury severity significantly on major collectors based on the latent class marginal effects. Marginal effects find that crashes happening on the curb line, or off the surface, have a 0.070 lower probability of no injury, 0.071 higher probability of a minor injury and a 0.002 lower probability of a major injury.

As for the greatest effect on interstates, the latent class marginal effects revealed that dry surface conditions have the utmost impact on injury severity. For instance, dry surfaces have a 0.154 lower probability for no injury and minor injury, although dry surface conditions have a 0.245 higher probability of a major injury.

The probability split of Class 1 and Class 2 for principal arterials is 0.795 to 0.205 and was highly significant at 99%. Similarly, the probability split of Class 1 and Class 2 for major collectors is 0.789 to 0.211 and highly significant at 99%, as is the probability split of Class 1 and Class 2 for interstates at 0.718 to 0.282. The significance of the class splits indicate a large amount of group-specific (classes) heterogeneity in which variables may have differing effects on injury severity outcomes.

2.5.3 Model Performance

The overall fit of the latent class model, based on log-likelihood values, is better than the mixed logit model for each road classification. The better fit belonging to the latent class model is also seen in the findings of Xie et al. (2012) in which their results indicate that the latent class framework provides a better statistical fit for their data.

In addition, variables effecting injury severity the greatest were substantially different according to the latent class method (straight horizontal geometrics was the only shared variable). In each classification, the latent class provided two or more variables that had a significant increase or decrease in injury severity probability while the mixed logit method had one per road classification. This may indicate that the latent class approach is accounting for more heterogeneity by informing the analyst of several high impact variables, as the marginal effect values became larger as overall model fit increased (e.g. marginal effect values increased from the fixed-parameter logit model to the mixed logit model and increased again from the mixed logit model to the latent class model). Further, variables with the greatest effect provided by the latent class model are variables that intuitively make sense (e.g. heavy vehicles that overturn are to increase major injury probability), while increases in no injury crashes were predominantly seen in the mixed logit models.

To further assess model performance, actual choices (injury severities) versus predicted injury severities were examined. As seen in Table 2.15, both approaches under-predicted each severity, but the latent class framework did predict more actual injury severities for each roadway classification. For principal arterials, the mixed logit approach predicted approximately 77% of the actual severities and the latent class approach predicted roughly 89% of the actual severities. The mixed logit model predicted about 80% of the actual severities on major collectors while the latent class model predicted nearly 87% of the actual severities. Finally, the mixed logit framework predicted approximately 80% of the actual severities on interstates whereas the latent class framework predicted almost 89% of the actual severities.

Table 2.15: Actual Injury Severity versus Predicted Injury Severity

Road Classification	Injury Severity	Actual	Mixed Logit Predicted	Latent Class Predicted
Principal Arterials	No Injury	1,334	1,161	1,286
	Minor Injury	197	40	105
	Major Injury	33	3	7
	Total	**1,564**	**1,204**	**1,398**
Major Collectors	No Injury	916	817	858
	Minor Injury	117	27	49
	Major Injury	25	5	12
	Total	**1,058**	**849**	**919**
Interstates	No Injury	1,402	1,246	1,355
	Minor Injury	174	34	69
	Major Injury	33	3	5
	Total	**1,609**	**1,283**	**1,429**

2.6 Insights & Future Work

The present study provides a heavy vehicle injury severity analysis while assessing two econometric frameworks used to capture the unobserved heterogeneity present in crash reported data. The mixed logit method attempts to capture the heterogeneity by assuming a given parameter varies based on a normal distribution while the latent class approach permits parameters to vary across a specified number of classes without needing to define a distribution. Using overall model fit, marginal effects and actual severities versus predicted severities the estimates for each framework were examined.

Using Idaho as a case study to investigate heavy vehicle injury severities by roadway classification, it was determined that roadway classifications should be modeled separately. In addition to the model separation test, the large number of statistically significant variables found to be exclusive to each model indicate that injury severity factors differ by roadway classification. With the highest percentage of heavy vehicle crashes in Idaho occurring on road classifications presented in this study, Idaho officials can use these findings as guidance to take action to reduce heavy vehicle crashes by focusing on specific factors that are unique to a road classification. For example, Idaho agencies can focus on mitigating weekend crashes that lead to major injuries on principal arterials by disaggregating the data by weekend crashes and performing an analysis to determine contributing factors. To highlight the difference in roadway classifications, 8 of the 16 significant variables for principal arterials were exclusive to principal arterials, 9 of the 15 significant variables for major collectors were exclusive to major collectors and 10 of the 16 significant variables for interstates were exclusive to interstates.

Furthermore, the latent class framework performed better based on overall model fit and is line with Chu (2014) and other past studies. As for the rate of prediction, the latent class framework predicted more correct injury severities by category and a higher percentage of overall severities. Marginal effects also show a difference between the two frameworks, where the latent class method provided several variables with large effects on injury severity while the mixed logit approach did not. The fit measures provided by the latent class method suggest a need to consider the latent class framework to account for heterogeneity when investigating the contributing factors that lead to injury severity outcomes. Findings in the current study are also in line with Cerwick et al. (2014), in which they suggested that a smaller dataset may find that the latent class approach is more favorable for injury severity analyses.

In summary, this study presents two precise frameworks to model injury severity outcomes by roadway classification while accounting for heterogeneity. Through this Idaho case study it was determined that road classifications need to be investigated separately, while also establishing that for the given dataset the latent class framework is preferred based on overall model fit, predicted injury severities and marginal effects inferences. Results suggest that future studies should consist of modeling other roadway classifications based on the need of the study area to assist transportation agencies, planners and engineers in developing the safest transportation infrastructure possible. That is, rather than generalizing the contributing factors that lead to injury severity, factors can be identified by geographic region and road classification within that area. The presented methodology, latent class multinomial logistic regression, should also be extended to other studies that focus on injury severity outcomes to provide more results regarding the superior framework for injury severity analyses.

REFERENCES

Anastasopoulos, P.C., Mannering, F., 2011. An Empirical Assessment of Fixed and Random Parameter Logit Models Using Crash- and Non-Crash-Specific Injury Data. Accident Analysis and Prevention 43, 1140–1147.

Behnood, A., Roshandeh, A.M., Mannering, F., 2014. Latent Class Analysis of the Effects of Age, Gender, and Alcohol Consumption on Driver-Injury Severities. Analytic Methods in Accident Research 3–4, 56–91.

Blincoe, L., Miller, T.R., Zaloshnja, E., Lawrence, B.A., 2015. The Economical and Societal Impact of Motor Vehicle Crashes, 2010 (Revised May 2015).

Cerwick, D.M., Gkritza, K., Shaheed, M.S., Hans, Z., 2014. A Comparison of the Mixed Logit and Latent Class Methods for Crash Severity Analysis. Analytic Methods in Accident Research 3–4, 11–27.

Chu, H.C., 2014. Assessing Factors Causing Severe Injuries in Crashes of High-Deck Buses in Long-Distance Driving on Freeways. Accident Analysis and Prevention 62, 130–136.

Collins, L.M., Fidler, P.L., Wugalter, S.E., Long, J.D., 1993. Goodness-of-Fit Testing for Latent Class Models. Multivariate Behavioral Research 28, 375–389.

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

Federal Motor Carrier Safety Administration, 2013. Commercial Motor Vehicle Facts. URL https://www.fmcsa.dot.gov/safety/data-and-statistics/commercial-motor-vehicle-facts-—-march-2013 (accessed 2.11.15).

Geedipally, S., Turner, P., Patil, S., 2011. Analysis of Motorcycle Crashes in Texas with Multinomial Logit Model. Transportation Research Record 2265, 62–69.

Gkritza, K., Mannering, F., 2008. Mixed Logit Analysis of Safety-Belt Use in Single- and Multi-Occupant Vehicles. Accident Analysis and Prevention 40, 443–451.

Greene, W.H., 2012a. LIMDEP Version 10 Reference Guide, 10th ed. Econometric Software, Inc., New York.

Greene, W.H., 2012b. NLOGIT 5 Reference Guide. Econometric Software, Inc., New York.

Greene, W.H., Hensher, D.A., 2003. A Latent Class Model for Discrete Choice Analysis: Contrasts with Mixed Logit. Transportation Research Part B: Methodological 37, 681–698.

Hagenaars, J.A., McCutcheon, A.L. (Eds.), 2002. Applied Latent Class Analysis. Cambridge University Press, New York.

Hensher, D.A., Rose, J.M., Greene, W.H., 2015. Applied Choice Analysis, 2nd ed. Cambridge University Press, United Kingdom.

Idaho Office of Highway Safety, 2014. Idaho Traffic Crashes.

Islam, M., Hernandez, S., 2013. Modeling Injury Outcomes of Crashes Involving Heavy Vehicles on Texas Highways. Transportation Research Record 2388, 28–36.

Islam, S., Jones, S.L., Dye, D., 2014. Comprehensive Analysis of Single and Multi-Vehicle Large Truck At-Fault Crashes on Rural and Urban Roadways In Alabama. Accident Analysis and Prevention 67, 148–158.

Kashani, A.T., Mohaymany, A.S., 2011. Analysis of the Traffic Injury Severity on Two-Lane, Two-Way Rural Roads Based on Classification Tree Models. Safety Science 49, 1314–1320.

Khorashadi, A., Niemeier, D., Shankar, V., Mannering, F., 2005. Differences in Rural and Urban Driver-Injury Severities in Accidents Involving Large-Trucks: An Exploratory Analysis. Accident Analysis and Prevention 37, 910–921. doi:10.1016/j.aap.2005.04.009

Kim, J.K., Ulfarssom, G.F., Shankar, V.N., Mannering, F., 2010. A Note on Modeling Pedestrian-Injury Severity in Motor-Vehicle Crashes with the Mixed Logit Model. Accident Analysis and Prevention 42, 1751–1758.

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

Louviere, J.J., Hensher, D.A., Swait, J.D., 2000. Stated Choice Methods: Analysis and Applications. Cambridge University Press, New York.

Magidson, J., Vermunt, J.K., 2004. Latent Class Models 11.

Mannering, F., Shankar, V., Bhat, C.R., 2016. Unobserved Heterogeneity and the Statistical Analysis of Highway Accident Data. Analytic Methods in Accident Research 11, 1–16.

McFadden, D., Train, K., 2000. Mixed MNL Models for Discrete Response. Journal of Applied Econometrics 15, 447–470.

Milton, J.C., Shankar, V., Mannering, F., 2008. Highway Accident Severities and the Mixed Logit Model: An Exploratory Empirical Analysis. Accident Analysis and Prevention 40, 260–266.

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

Moore, D.N., Schneider IV, W.H., Savolainen, P.T., Farzaneh, M., 2011. Mixed Logit Analysis of Bicyclist Injury Severity Resulting From Motor Vehicle Crashes at Intersetion and Non-Intersection Locations. Accident Analysis and Prevention 43, 621–630.

Morgan, A., Mannering, F., 2011. The Effects of Road-Surface Conditions, Age, and Gender on Driver-Injury Severities. Accident Analysis and Prevention 43, 1852–1863.

Nylund, K.L., Asparouhov, T., Muthén, B.O., 2007. Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. Structural Equation Modeling 14, 535–569.

Obeng, K., 2011. Gender Differences in Injury Severity Risks in Crashes at Signalized Intersections. Accident Analysis and Prevention 43, 1521–1531.

Pahukula, J., Hernandez, S., Unnikrishnan, A., 2015. A Time of Day Analysis of Crashes Involving Large Trucks in Urban Areas. Accident Analysis and Prevention 75, 155–163.

Russo, B.J., Savolainen, P., Schneider IV, W.H., Anastasopoulos, P.C., 2014. Comparison of Factors Affecting Injury Severity in Angle Collisions by Fault Status Using a Random Parameters Bivariate Ordered Probit Model. Analytic Methods in Accident Research 2, 21–29.

Savolainen, P., Mannering, F., 2007. Probabilistic Models of Motorcyclists' Injury Severities in Single - and Multi - Vehicle Crashes. Accident Analysis and Prevention 39, 955–963.

Shaheed, M.S., Gkritza, K., 2014. A Latent Class Analysis of Single-Vehicle Motorcycle Crash Severity Outcomes. Analytic Methods in Accident Research 2, 30–38.

Shen, J., 2009. Latent Class Model or Mixed Logit Model? A Comparison by Transport Mode Choice Data. Applied Economics 41, 2915–2924.

Washington, S.P., Karlaftis, M.G., Mannering, F., 2011. Statistical and Econometric Methods for Transportation Data Analysis, 2nd ed, Chapman & Hall/CRC. Chapman & Hall/CRC.

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

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

Yasmin, S., Eluru, N., 2013. Evaluating Alternate Discrete Outcome Frameworks for Modeling Crash Injury Severity. Accident Analysis and Prevention 59, 506–521.

Yasmin, S., Eluru, N., Bhat, C.R., Tay, R., 2014. A Latent Segmentation Based Generalized Ordered Logit Model to Examine Factors Influencing Driver Injury Severity. Analytic Methods in Accident Research 1, 23–38.

Ye, F., Lord, D., 2014. Comparing Three Commonly Used Crash Severity Models on Sample Size Requirements: Multinomial Logit, Ordered Probit and Mixed Logit Models. Analytic Methods in Accident Research 1, 72–85.

3.0 HEAVY VEHICLE CRASH RATE ANALYSIS: A COMPARISON OF HETEROGENEITY METHODS USING IDAHO CRASH DATA

BY: JASON C. ANDERSON
Graduate Student, School of Civil and Construction Engineering
Oregon State University, 211 Kearney Hall, Corvallis, OR 97331-3212
anderjas@oregonstate.edu

SALVADOR HERNANDEZ, PhD
Assistant Professor, School of Civil and Construction Engineering
Oregon State University, 309 Owen Hall, Corvallis, OR 97331-3212
(541)-737-4740
sal.hernandez@oregonstate.edu

ABSTRACT

Studies investigating crash rates by roadway classification are few and far between and even more so if extended to focus on heavy vehicles. This study explores and compares two advanced econometric methods, namely random-parameter Tobit regression and latent class Tobit regression, to determine contributing factors for heavy vehicle crashes per million-vehicle-miles-traveled while accounting for the unobserved heterogeneity present in crash data. The increasing crash rates in Idaho, crash proportion by roadway classification and available data make for an ideal case study. Empirical results show that although the random-parameter Tobit regression model provides better insight into heavy vehicle crash rates than the fixed-parameter approach, the latent class Tobit regression model was found to be the preferred methodology for the given dataset. Traffic volumes, roadway characteristics and traffic control devices were among the variables found to be statistically significant. Results from this study provide a framework to better account for heterogeneity, while identifying key factors by roadway classification that influence heavy vehicle crash rates. The illustrated framework and analysis by roadway classification can provide guidance to transportation agencies and policy makers, and prompt future studies that include latent class analysis and/or analysis by road classification.

Keywords: Heavy Vehicle, Random-Parameter Tobit, Latent Class, Crash Rate, Road Classification

3.1 Introduction

Heavy vehicle crashes have a substantial economic impact on commerce and society [trucks with a gross vehicle weight rating (GVWR) greater than 4,535.9 kg (10,000 pounds)]. For example, in the US, heavy vehicle crashes where about $87 billion in 2011 and costs due to delay and other consequences where roughly $28 billion (Blincoe et al., 2015; Federal Motor Carrier Safety Administration, 2013). These values will continue to increase as the economy continues to grow, and with this growth, the volume of heavy vehicles on the nation's freight infrastructure will also increase as seen from 2010 to 2013 where a 2.30% increase of heavy vehicles was experienced (about 6,000,000 vehicles) (Bureau of Transportation Statistics, 2015). This number is expected to continue to grow in the coming years and crashes associated with heavy vehicles will remain a concern for safety planners and other safety related agencies. Although heavy vehicle crashes have decreased over the past two decades, the number of fatal crash involvements per 100-million vehicle-miles-traveled compared to passenger cars is higher (1.34 versus 1.08 for the year 2014)(Federal Motor Carrier Safety Administration, 2016; NHTSA, n.d.). In Idaho, the state experienced a 5.6% increase in heavy vehicle crashes and a 4.4% increase in heavy vehicle crashes per million vehicle-miles-traveled (MVMT) from 2010 to 2013 (Idaho Office of Highway Safety, 2014). Moreover, 50% of these heavy vehicle crashes occurred on local roads, 28% of injury crashes happened on interstates and 68.2% of fatalities took place on U.S. and State Highways (Idaho Office of Highway Safety, 2014). These statistics illustrate the need for continued research in understanding the relationship between heavy vehicle crash rates and road classification (i.e., by functional class of road).

There have been a number of studies that have addressed crash frequency through the application of count and spatial based models (Aguero-Valverde and Jovanis, 2008; Bhat et al., 2014; Castro et al., 2012; Chen and Tarko, 2012; Park and Lord, 2007; Shea et al., 2015; Ukkusuri et al., 2011; Xie et al., 2014; Xie and Zhang, 2008; Xu and Huang, 2015; Ye et al., 2009; Zhang et al., 2012). However, most of these studies have primarily focused on data related to pedestrians, passenger car or all traffic mixes in a single modeling framework and do not model heavy vehicle crash rates explicitly. Although there have been several recent efforts to understand heavy vehicle injury severity factors (Islam et al., 2014a; Khorashadi et al., 2005; Milton et al., 2008; Moore et al., 2011; Pahukula et al., 2015; Zhu and Srinivasan, 2011), heavy vehicle crash rate analyses are sparse. This is especially true for heavy vehicle crash rates by functional class of road. A possible reason for this deficiency in the literature may stem from the availability of sufficient data to capture the complex interactions of multiple crash rate factors under a single framework by functional class.

Recent studies have addressed the issue of insufficient data through the application of statistical and econometric methods that account for unobserved factors (also referred to as unobserved heterogeneity) which are factors unknown to the analyst and that vary across the sample population, see Mannering and Bhat for a complete review of these methods (Mannering et al., 2016). For instance, weather conditions which continually changes over time as well as the truck driver response to the changing weather condition (not present in the data). These models allow the analyst to account for these variations and make more informed inferences regarding the effects of the contributing factors (Mannering et al., 2016).

With this in mind, the present study seeks to identify contributing factors that impact heavy vehicle crashes per MVMT by road classification (principle arterials, major collectors and intestates) through the application and performance based comparison of two “heterogeneity” models, namely random parameters- and latent class- Tobit regression. The Tobit modeling framework is selected due to the nature of crash rate data. Similar to frequency models, a crash rate analysis is likely to have several observations in which no crash has occurred, therefore a censoring method is recommended to account for the skewed nature of the response variable (crash rate). It has be shown that the Tobit regression modeling framework can account for the skewed nature of crash rate data without omitting observations by censoring the analysis at a given value (Washington et al., 2010). These models have been successfully applied to related transportation safety data, for example, Anastasopoulos et al. use the random-parameter Tobit model to determine factors that influence crash rates per 100-million VMT on highways (Anastasopoulos et al., 2012a). Islam and Hernandez investigated fatalities per million truck-miles-traveled and fatalities per ton-mile of freight for heavy vehicles through the application of a random-parameter Tobit regression model and Chen et al. utilized a random-effects Tobit model to analyze crash rates with refined-scale data (Chen et al., 2014; Islam and Hernandez, 2015). Due to the repetitive nature of crash rates, the random-effects approach is able to account for the censoring effects and serial correlation[3], as well as attempt to capture the unobserved heterogeneity (Chen et al., 2014). From a latent class Tobit regression application, there are no known applications to transportation safety data, however the method has been applied to social science studies (see (Brown et al., 2015; Jedidi et al., 1993)).

Therefore, the present study will use the random-parameter Tobit method to identify significant contributing factors to crash rates by roadway functional class while accounting for heterogeneity. However, variables not found to be random in the random-parameter method may in fact have differing effects on heavy vehicle crash rates. Hence, to determine if such variables effect crash rates differently, the current study will be extended by investigating the results of the Tobit latent class approach by disaggregating the Tobit model into unobserved groups (or classes; see Methodology section). To accomplish this, an extensive crash database collected and maintained by the Idaho Department of Transportation (IDT) is used. The findings of this study can provide insight that can aid safety planners and other safety related agencies in identifying appropriate countermeasures to help reduce and mitigate heavy vehicle crashes. To the best of the authors' knowledge, these are first attempts at developing these types of models for heavy vehicle crash rate analysis.

---

[3] In this case, serial correlation is referring to the roadway segments. The same segment appears over and over, therefore it is possible to experience spatial correlation.

3.2 Source of Data

The current study uses 7 years of police-recorded crash data obtained from the state of Idaho from 2007 to 2013. Each year was filtered to represent heavy vehicle crashes, then combined with traffic data from the Idaho Department of Transportation (IDT) (e.g., AADT and VMT) utilizing segment codes and milepost markers that were present in both datasets. The segment codes and milepost numbers of the location of the crash were used to determine the intermediate segments within the milepost intervals in the traffic data—these segments are used for the modeling process being that each segment had distinct traffic data. Using the complete dataset consisting of exposure variables (i.e., roadway geometrics, traffic control devices, number of lanes, etc.) and traffic volumes, several indicator variables were created to identify specific exposure conditions and traffic volumes that impact crash rates by road classification in Idaho. Table 3.1 to Table 3.3 display the response variable and indicators found to be significant throughout the modeling process[4].

[4] No model was fit for Other Principal Arterials due to the overdispersed nature of the response variable, crashes per million-vehicle-miles traveled.

Table 3.1: Descriptive Statistics for Significant Variables on Principal Arterials

Variable	Mean	Standard Deviation	Minimum	Maximum
Crashes per Million-Vehicle Miles-Traveled (Response Variable)	0.127	0.208	0.005	3.488
Speed Limit (1 if 65MPH, 0 Otherwise)	0.443	0.497	-	-
Traffic Control Device (1 if No Device, 0 Otherwise)	0.702	0.458	-	-
Road Configuration (1 if 2-Way & 2-Way Left-Turn Lane	0.126	0.333	-	-
Heavy Vehicle AADT (1 if Less Than or Equal to 300, 0 Otherwise)	0.158	0.365	-	-
Passenger Vehicle AADT (1 if Greater Than 10,500, 0 Otherwise)	0.108	0.310	-	-
Total AADT (1 if Between 5,000 and 7,000, 0 Otherwise)	0.167	0.373	-	-

Note: AADT = Average Annual Daily Traffic; Total of 862 Segments

Table 3.2: Descriptive Statistics for Significant Variables on Major Collectors

Variable	Mean	Standard Deviation	Minimum	Maximum
Crashes per Million-Vehicle Miles-Traveled (Response Variable)	0.620	1.204	0.017	13.105
Speed Limit (1 if Less Than or Equal to 40 MPH, 0 Otherwise)	0.241	0.428	-	-
Traffic Control Device (1 if Stop Sign, 0 Otherwise)	0.159	0.366	-	-
Horizontal Geometrics (1 if Straight, 0 Otherwise)	0.789	0.408	-	-
Road Configuration (2-Way & Double-Yellow Painted Divider, 0 Otherwise)	0.129	0.335	-	-
Total AADT (1 if Less Than 500, 0 Otherwise)	0.228	0.420	-	-
Passenger Vehicle AADT (1 if Greater Than 2,500, 0 Otherwise)	0.227	0.419	-	-

Note: AADT = Average Annual Daily Traffic; Total of 768 Segments

Table 3.3: Descriptive Statistics for Significant Variables on Interstates

Variable	Mean	Standard Deviation	Minimum	Maximum
Crashes per Million-Vehicle Miles-Traveled (Response Variable)	0.034	0.047	0.003	0.726
Speed Limit (1 if 75 MPH, 0 Otherwise)	0.675	0.469	-	-
Total AADT (1 When Less Than 6,500, 0 Otherwise)	0.164	0.370	-	-
Passenger Vehicle AADT (1 if Greater Than 15,000, 0 Otherwise)	0.116	0.321	-	-
Heavy Vehicle AADT (1 if Between 2,000 and 3,000, 0 Otherwise)	0.256	0.437		
Horizontal Geometrics (1 if Curved, 0 Otherwise)	0.230	0.421	-	-
Road Configuration (1 if 2-Way and Raised/Depressed Divider)	0.929	0.258		
Surface Defects (1 if No Surface Defects, 0 Otherwise)	0.960	0.195	-	-

Note: AADT = Average Annual Daily Traffic; Total of 379 Segments

3.4 Methodology

This section will outline the process of modeling crashes per MVMT by roadway classification.

3.4.1 Dependent Variable

To model heavy vehicle crash rates, a rate for each segment is calculated using the traffic data provided by Idaho. The following equation was used to calculate MVMT for each segment (Golembiewski and Chandler, 2011):

$$\mathrm{R_s} = \frac{\sum_{\mathrm{y}=1}^{\mathrm{n}} \mathrm{N_{ys}}}{\left[\sum_{\mathrm{y}=1}^{\mathrm{n}} \mathrm{AADT_{ys}} \times \mathrm{L_s} \times 365\right] \big/ 1{,}000{,}000} \tag{1}$$

where R_s is the number of crashes per MVMT on segment s, y is the year (2007 to 2013), N_{ys} is the number of heavy vehicle crashes in year y on segment s, $AADT_{ys}$ is the average annual daily traffic for year y on segment s and L_s is the length of segment s in miles.

Segments in the traffic dataset that had zero crashes were omitted due to uncertainty by not being represented in the crash data. That is, each crash was associated with a segment code and milepost marker with corresponding exposure variables, therefore inference in regard to exposure variables between milepost markers that did not involve a crash would be inaccurate. This resulted in many segments for each road classification, each of which had at least one crash. Although at least one crash was observed in each segment, the crash rate (response variable) was still skewed and needed to be addressed during analysis. To illustrate the distribution of crash rates for each road classification, Figure 3.1 to Figure 3.4 are presented below.

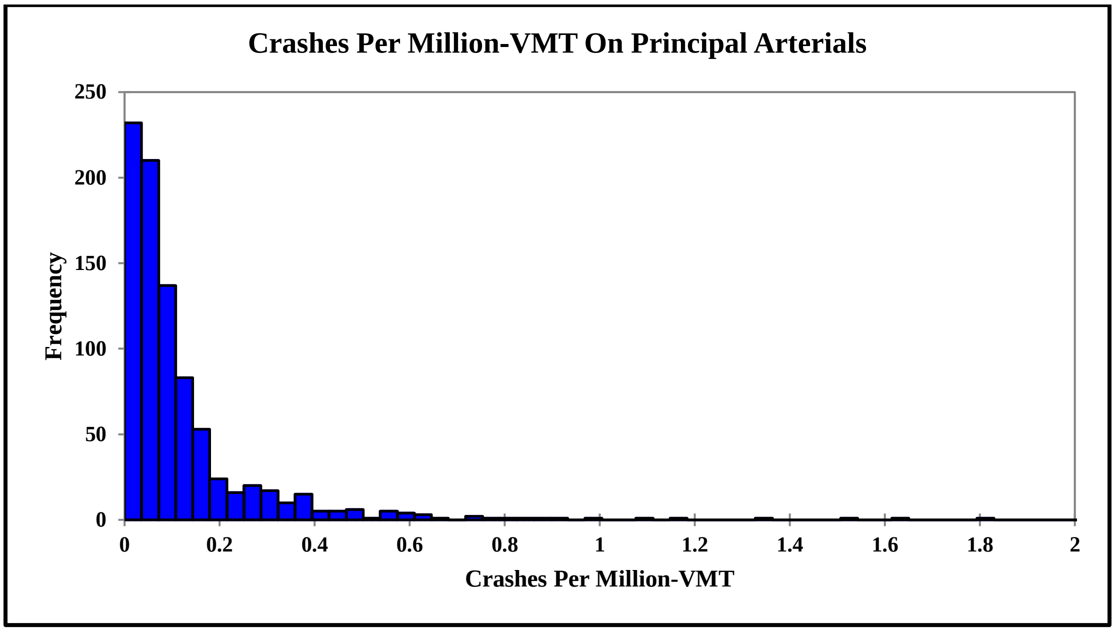

Figure 3.1: Distribution of Crashes per Million-VMT on Principal Arterials

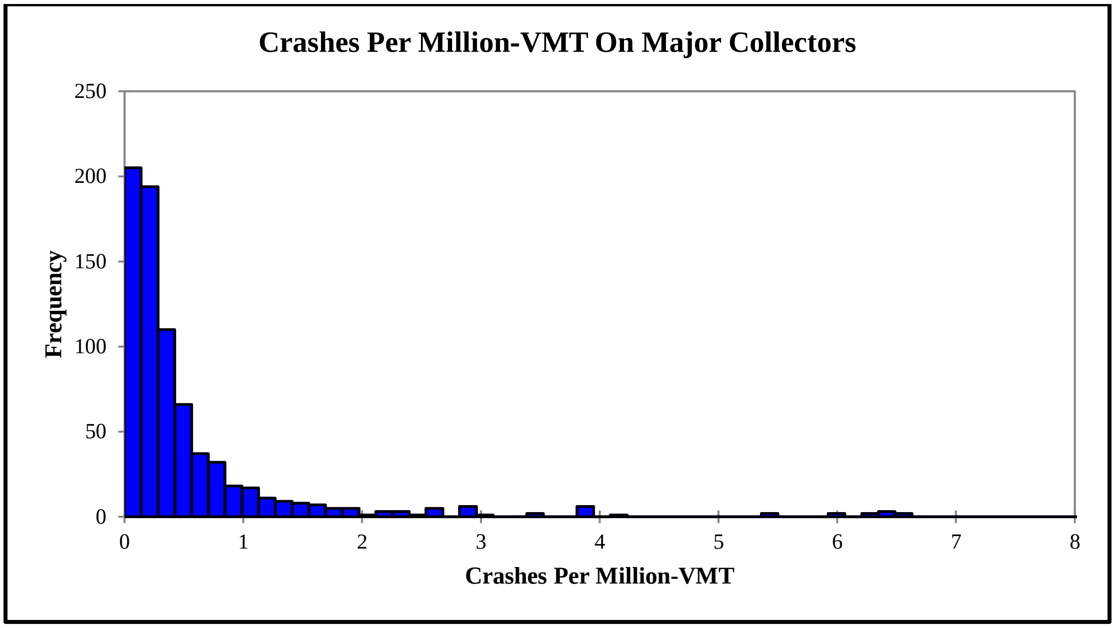

Figure 3.2: Distribution of Crashes per Million-VMT on Major Collectors

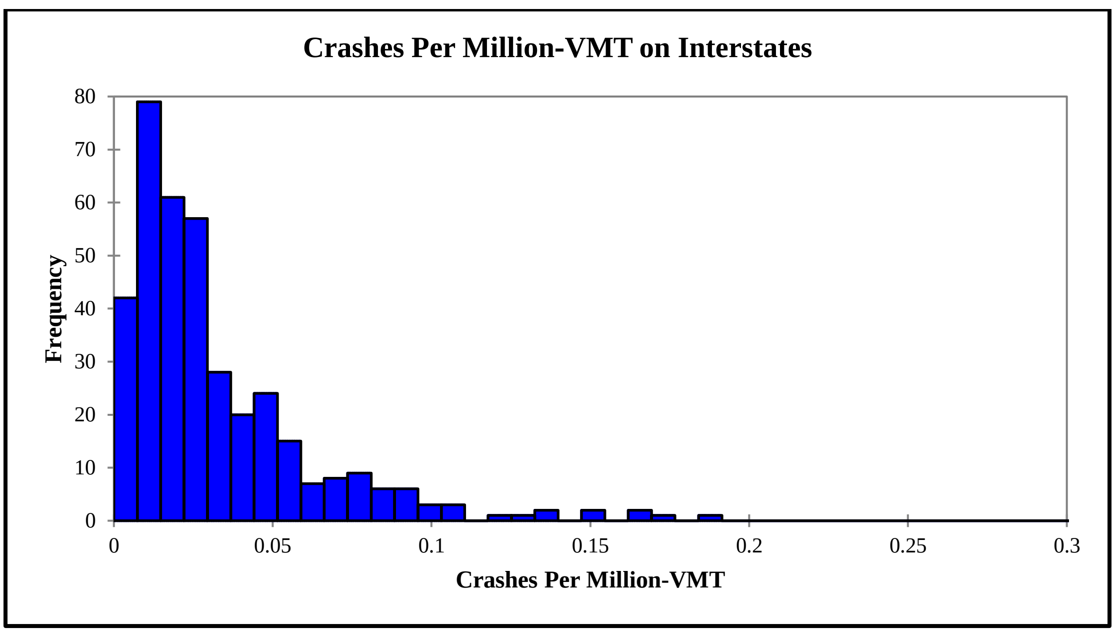

Figure 3.3: Distribution of Crashes per Million-VMT on Interstates

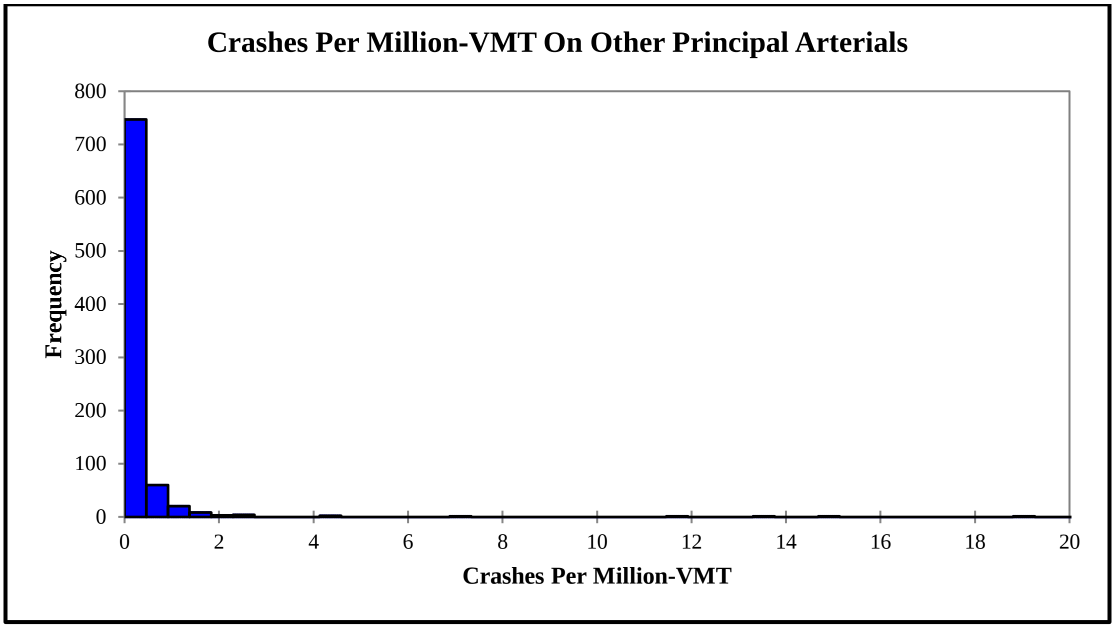

Figure 3.4: Distribution of Crashes per Million-VMT on Other Principal Arterials

3.4.1 Random-Parameter Tobit Model

The distribution of crash rates presented in Figure 1 illustrate the need to utilize a method that can account for the large lower bound cluster of observations[5] while maintaining the linear assumptions required for regression of a continuous dependent variable (in this case heavy vehicle crash rates by roadway classification). With this premise in mind, we seek to develop a statistical model that can be used to determine the contributing factors on heavy vehicle crash rates by roadway classification. Namely, this study will apply the Tobit regression modeling framework first introduced

---

[5] The data for Other Principal Arterials, as seen in Figure 4, is too centered near zero. Even through censoring the model, the population size for modeling was too small and produced erroneous estimates. For this reason, the model for Other Principal Arterials has been omitted from this study.

by James Tobin (Tobin, 1958). However, key variables not available within many crash datasets as previously mentioned and variation across the available variables is likely to result in unobserved heterogeneity, and if neglected, will lead to biased estimates and inaccurate inferences (see (Mannering et al., 2016) for further discussion). To account for unobserved heterogeneity (Anastasopoulos and Mannering, 2009; Anastasopoulos et al., 2012a; Carlsson and Martinsson, 2007; Islam and Hernandez, 2015; Yu et al., 2015), the current study will apply the random-parameter approach to the traditional Tobit regression framework (William H. Greene, 2012). As mentioned earlier, Anastasopoulos et al., Islam and Hernandez and Chen et al. all apply successful applications of the random-parameter Tobit regression model to transportation safety data (Anastasopoulos et al., 2012a; Chen et al., 2014; Islam and Hernandez, 2015). Hence, for this work, the standard Tobit model is expressed for heavy vehicle crashes as:

$$\begin{aligned} &\mathrm{Y}_s^* = \boldsymbol{\beta}'\mathbf{X_s} + \varepsilon_s \quad \text{with} \quad \varepsilon_s \sim \mathrm{N}[0, \sigma^2] \text{ and } s = 1, 2, \dots \mathrm{N} \\ &\mathrm{Y}_s = \mathrm{Y}_s^* \ \text{ if } \ \mathrm{Y}_s^* > \mathrm{L} \\ &\mathrm{Y}_s = 0 \ \text{ if } \ \mathrm{Y}_s^* \leq \mathrm{L} \end{aligned} \tag{2}$$

where N is the number of observations, Y_s is the number of crashes per MVMT (the response variable), L is the number that the model is left-censored at, $\boldsymbol{X_s}$ is the vector of explanatory variables (AADT, roadway geometrics, etc.), $\boldsymbol{\beta'}$ is the vector of estimated parameters and ε_s is the normally and independently distributed error term with a mean of zero and a constant variance, σ^2. To determine the likelihood, the following function applies (Brown et al., 2015):

$$\mathrm{LL} = \prod_{0}\left[1 - \Psi\left(\frac{\boldsymbol{\beta}'\mathbf{X}_{\mathbf{s}}}{\sigma}\right)\right]\prod_{1}\left(\frac{1}{\sigma}\right)\psi\left(\frac{\mathrm{Y}_{\mathrm{s}} - \boldsymbol{\beta}'\mathbf{X}_{\mathbf{s}}}{\sigma}\right) \tag{3}$$

where $\Psi\left(\frac{\boldsymbol{\beta}'\mathbf{X}_{\mathbf{s}}}{\sigma}\right)$ is the standard normal distribution function and $\psi\left(\frac{\mathrm{Y}_{\mathrm{s}}-\boldsymbol{\beta}'\mathbf{X}_{\mathbf{s}}}{\sigma}\right)$ is the standard normal density function.

Further, in an attempt to capture the unobserved heterogeneity, the random-parameter approach is now applied to the Tobit framework and estimable parameters can be written as follows (William H. Greene, 2012):

$$\boldsymbol{\beta}_{\mathrm{s}} = \boldsymbol{\beta} + \phi_{\mathrm{s}} \tag{4}$$

where the equivalent log-likelihood function is represented as (Brown et al., 2015):

$$\mathrm{LL} = \sum_{\forall s} \ln \int_{\phi_s} g(\phi_s) P(Y_s^* | \phi_s) d\phi_s \tag{6}$$

where $g(\phi_s)$ is the probability density function of ϕ_s and $P(Y_s^*|\phi_s)$ is the probability of the Tobit model. As stated in previous studies (Anastasopoulos et al., 2012a; Islam and Hernandez, 2015), the maximum likelihood estimations encounter computing issues due to its complexity. To address this issue, a common approach developed by Halton is used to simulate the maximum likelihood by utilizing Halton draws to solve the complex integral seen in Eq. (6) and has been shown to be preferable over merely random draws (Bhat, 2003; Halton, 1960; Train, 1999). The present study uses the Limdep software NLOGIT5 to generate the Halton draws and estimate the maximum likelihood.

3.4.2 Latent Class Tobit Model

Although the random-parameter method accounts for unobserved heterogeneity, there is a possible disadvantage due to the assumption that the parameters vary in a predefined distribution and that parameters vary only across singular observations (see (Mannering et al., 2016) for further discussion). With this in mind, the latent class approach attempts to capture unobserved heterogeneity by allowing estimable parameters to vary with a discrete distribution across unobserved groups of observations (or classes). The heterogeneity is accounted for by defining a finite number of points and measuring the mass probability of the intervals between points. Applying this to the Tobit regression structure results in the following:

$$\begin{aligned} &Y_s^*|(\text{Class} = C) = \boldsymbol{\beta'}_{\boldsymbol{C}}\mathbf{X_s} + \varepsilon_{s|C} \quad \text{with} \quad \varepsilon_{s|C} \sim N[0, \sigma^2{}_C] \text{ and } s = 1, 2, \ldots N \\ &Y_s = Y_s^* \text{ if } Y_s^* > L \\ &Y_s = 0 \text{ if } Y_s^* \leq L \end{aligned} \tag{6}$$

where $\boldsymbol{\beta'}_{\boldsymbol{C}}$ is a vector of estimated parameters belonging to class C and $Y_s^*|(\text{Class} = C)$ is the number of crashes per MVMT of segment s in class C. The corresponding log-likelihood function can now be written as (Brown et al., 2015):

$$LL = \sum_{i=1}^{N} \log\left[\sum_{C=1}^{C} P_{sC}(\delta_C, \omega_s)\left[f\left(Y_s|\text{Class} = C, \mathbf{X_s}, \boldsymbol{\beta'}_{\mathbf{C}}, \sigma_C\right)\right]\right] \tag{7}$$

where $P_{sC}(\delta_C, \omega_s)$ is the logit probability of being in class C and represented by the multinomial logit form (Brown et al., 2015; William H Greene, 2012c):

$$P_{sC}(\delta_C, \omega_s) = \frac{e^{(\omega_s \delta_C)}}{\sum_{C=1}^{C} e^{(\omega_s \delta_C)}} \quad \text{with } C = 1, 2, \ldots C \text{ and } \delta_C = 0 \text{ for normalization} \tag{8}$$

Lastly, after the parameters have been estimated a second estimation is conducted to determine the posterior probabilities of crash rate Y_s belonging to class C (William H Greene, 2012c). The posterior probability is represented as follows:

$$P(\text{Class} = C|\text{Crash Rate } Y_s) = \frac{f(\text{Crash Rate } Y_s|\text{Class} = C)P(\text{Class } C)}{\sum_{C=1}^{C} f(\text{Crash Rate } Y_s|\text{Class} = C)\, P(\text{Class } C)} \tag{9}$$

3.5 Model Estimation Results

3.5.1 Random-Parameter Tobit Model

As described previously, the maximum likelihood method was used to estimate the parameters for the fixed-parameter and random-parameter models. Several distributions were considered for the distribution of the random parameters—normal, uniform, triangular and lognormal—but only the normal distribution was found to have a statistically significant mean and standard deviation. No variables were found to be random for principal arterials or interstates, yet Table 3.4 and Table 3.6 show that the fixed-parameter models have a better log-likelihood at convergence than the log-likelihood at zero (constant only). To determine the statistical significance the following log-likelihood ratio test was used (Washington et al., 2010):

$$\chi^2 = -2[\mathrm{LL}(0) - \mathrm{LL}(\beta)] \tag{10}$$

where $\mathrm{LL}(0)$ is the log-likelihood at zero (constant only) and $\mathrm{LL}(\beta)$ is the log-likelihood at convergence for the fixed-parameter model. This ratio test provides a chi-square statistic with degrees of freedom being the number of estimated parameters excluding the constant. However, variables were found to be random for major collectors, shown in Table 3.5, and to statistically asses the more significant model the ensuing log-likelihood ratio test was conducted between the fixed-parameter and random-parameter models:

$$\chi^2 = -2[\mathrm{LL}(\beta_{\mathrm{FP}}) - \mathrm{LL}(\beta_{\mathrm{RP}})] \tag{11}$$

where $\mathrm{LL}(\beta_{\mathrm{FP}})$ is the log-likelihood at convergence of the fixed-parameter model, $\mathrm{LL}(\beta_{\mathrm{RP}})$ is the log-likelihood at convergence of the random-parameter model and degrees of freedom for χ^2 being the number of random parameters. One more goodness of fit measure was applied, the Maddala Pseudo R^2 value (Veall and Zimmerman, 1996):

$$\mathrm{R}^2 = 1 - \mathrm{e}^{\left(\frac{-2[\mathrm{LL}(\beta) - \mathrm{LL}(0)]}{\mathrm{N}}\right)} \tag{12}$$

where $\mathrm{LL}(\beta)$ is the log-likelihood at convergence for the best fit model (random or fixed), $\mathrm{LL}(0)$ is the log-likelihood at zero and N is the number of observations.

With regard to the principal arterial model, the chi-square statistic of 83.11 and 6 degrees of freedom indicated with 99.99% confidence that the fixed-parameter model is preferred over the model with simply the constant. For interstates, a chi-square statistic of 61.25 and 6 degrees of freedom showed with 99.99% that the fixed-parameter model is superior to the model with no estimated parameters. In the case of major collectors, in which variables were found to be random, the chi-square statistic of 80.81 and 2 degrees of freedom demonstrated with 99.99% confidence that the random-parameter model is statistically preferred.

Table 3.4: Best Fit Tobit Regression Estimates for Principal Arterials

Variable	Coefficient	*t*-statistic	Partial Effect
Constant	0.12	4.48	
Speed Limit (1 if 65MPH, 0 Otherwise)	-0.11	-3.95	-0.041
Traffic Control Device (1 if No Device, 0 Otherwise)	-0.12	-4.33	-0.047
Road Configuration (1 if 2-Way & 2-Way Left-Turn Lane	0.10	2.81	0.041
Heavy Vehicle AADT (1 if Less Than or Equal to 300, 0 Otherwise)	0.10	2.88	0.038
Passenger Vehicle AADT (1 if Greater Than 10,500, 0 Otherwise)	-0.22	-4.70	-0.084
Total AADT (1 if Between 5,000 and 7,000, 0 Otherwise)	-0.06	-1.71	-0.023
Sigma, σ	0.31	26.85	
Number of Observations		862	
Log-Likelihood at Zero		-396.68	
Log-Likelihood at Convergence		-355.12	
χ^2		83.11	
Maddala Pseudo R^2		0.092	

Table 3.5: Best Fit Tobit Regression Estimates for Major Collectors

Variable	Fixed-Parameter Tobit			Random-Parameter Tobit		
	Coefficient	*t*-statistic	Partial Effect	Coefficient	*t*-statistic	Partial Effect
Constant	-1.31	-5.94		-1.21	-5.40	
Speed Limit (1 if Less Than or Equal to 40 MPH, 0 Otherwise)	0.52	2.87	0.205	0.42	2.30	0.120
Standard Deviation of Parameter, Normally Distributed	-	-	-	*0.73*	*6.67*	-
Traffic Control Device (1 if Stop Sign, 0 Otherwise)	0.45	2.22	0.178	0.40	2.01	0.116
Horizontal Geometrics (1 if Straight, 0 Otherwise)	0.91	4.38	0.356	0.73	3.67	0.210
Road Configuration (2-Way & Double-Yellow Painted Divider, 0 Otherwise)	0.70	3.10	0.275	0.36	1.53	0.104
Standard Deviation of Parameter, Normally Distributed	-	-	-	*0.99*	*5.45*	-
Total AADT (1 if Less Than 500, 0 Otherwise)	1.03	5.61	0.406	0.97	4.88	0.279
Passenger Vehicle AADT (1 if Greater Than 2,500, 0 Otherwise)	-0.60	-2.90	-0.238	-0.50	-2.22	-0.143
Sigma, σ	1.82	25.91		1.59	56.33	
Number of Observations	768			768		
Log-Likelihood at Zero	-1003.14			-1003.14		
Log-Likelihood at Convergence	-967.07			-926.67		
χ^2	72.13			80.81		
Maddala Pseudo R^2	0.090			0.181		

Table 3.6: Best Fit Tobit Regression Estimates for Interstates

Variable	Coefficient	*t*-statistic	Partial Effect
Constant	0.09	4.54	
Speed Limit (1 if 75 MPH, 0 Otherwise)	-0.02	-2.20	-0.008
Total AADT (1 When Less Than 6,500, 0 Otherwise)	0.03	3.88	0.016
Passenger Vehicle AADT (1 if Greater Than 15,000, 0 Otherwise)	-0.03	-2.28	-0.013
Heavy Vehicle AADT (1 if Between 2,000 and 3,000, 0 Otherwise)	0.02	2.06	0.008
Horizontal Geometrics (1 if Curved, 0 Otherwise)	0.02	2.36	0.010
Road Configuration (1 if 2-Way and Raised/Depressed Divider)	-0.03	-2.20	-0.014
Surface Defects (1 if No Surface Defects, 0 Otherwise)	-0.05	-2.81	-0.023
Sigma, σ	0.06	20.74	
Number of Observations		379	
Log-Likelihood at Zero		211.91	
Log-Likelihood at Convergence		242.54	
χ^2		61.25	
Maddala Pseudo R^2		0.149	

3.5.2 Latent Class Tobit Model

Latent class regression models for each road classification are shown in Table 3.7 to Table 3.9. In line with previous studies, the number of latent classes for each model were selected using the Bayesian Information Criterion (BIC)—the number of latent classes that produced the smallest BIC were used (Collins et al., 1993; Hagenaars and McCutcheon, 2002; Magidson and Vermunt, 2004; Nylund et al., 2007). However, Louviere et al. (2000) and Shen (2009) suggest that the smallest Akaike Information Criterion (AIC) be used to determine the best fit number of classes and was considered for the major collector latent class model. The AIC for the major collector model was much less than the BIC, while the BIC was much less for the principal arterial and interstate models. The best fit major collector model has four latent classes, a lower AIC and population size of 768. Similar findings were presented by Yang (2004) in which a population size of 700 resulted in four classes being the best fit number of latent classes due to a lower AIC.

The class-split for each classification is highly significant and the best fit number of classes are different for each model. Principal arterials have a best fit model with three latent classes, major collectors a best fit with four latent classes and interstates a best fit with two latent classes.

Table 3.7: Best Fit Latent Class Tobit Regression Estimates for Principal Arterials

Variable	Latent Class 1		Latent Class 2		Latent Class 3		Partial Effect
	Coefficient	*t*-statistic	Coefficient	*t*-statistic	Coefficient	*t*-statistic	
Constant	1.51	2.59	0.26	7.90	0.04	11.83	
Speed Limit (1 if 65MPH, 0 Otherwise)	0.05	0.06	-0.12	-3.25	0.00	-1.07	-0.004
Traffic Control Device (1 if No Device, 0 Otherwise)	-1.06	-2.06	-0.11	-3.15	-0.01	-1.93	-0.008
Road Configuration (1 if 2-Way & 2-Way Left-Turn Lane)	-0.94	-0.74	0.09	2.31	0.00	-0.12	0.000
Heavy Vehicle AADT (1 if Less Than or Equal to 300, 0 Otherwise)	0.93	1.86	0.05	1.34	0.00	0.69	0.005
Passenger Vehicle AADT (1 if Greater Than 10,500, 0 Otherwise)	-0.53	-0.10	-0.12	-1.58	-0.01	-1.56	-0.007
Total AADT (1 if Between 5,000 and 7,000, 0 Otherwise)	1.15	0.65	0.00	-0.05	0.00	-0.09	0.003
Sigma, σ	0.61	2.31	0.15	10.52	0.02	15.46	
Class Probability (*t*-statistic)	0.022 (2.77)		0.262 (9.06)		0.716 (25.24)		

Number of Observations	862
Log-Likelihood at Zero	166.81
Log-Likelihood at Convergence	195.60
Akaike Information Criterion	-339.20
Bayesian Information Criterion	**-215.50**

Table 3.8: Best Fit Latent Class Tobit Regression Estimates for Major Collectors

Variable	Latent Class 1		Latent Class 2		Latent Class 3		Latent Class 4		Partial Effect
	Coefficient	*t*-statistic	Coefficient	*t*-statistic	Coefficient	*t*-statistic	Coefficient	*t*-statistic	
Constant	2.11	0.71	0.08	3.83	0.07	0.93	0.43	1.69	
Speed Limit (1 if Less Than or Equal to 40 MPH, 0 Otherwise)	2.66	1.40	0.02	1.16	0.05	0.88	0.43	2.67	0.099
Traffic Control Device (1 if Stop Sign, 0 Otherwise)	1.94	1.05	0.01	0.39	0.06	1.19	0.55	3.08	0.012
Horizontal Geometrics (1 if Straight, 0 Otherwise)	-1.28	-0.61	0.02	1.00	0.13	2.17	0.02	0.07	-0.052
Road Configuration (2-Way & Double-Yellow Painted Divider, 0 Otherwise)	2.46	1.07	0.02	0.81	0.13	1.85	-0.09	-0.44	0.066
Total AADT (1 if Less Than 500, 0 Otherwise)	2.28	1.17	0.01	0.33	0.29	5.52	0.74	4.26	0.070
Passenger Vehicle AADT (1 if Greater Than 2,500, 0 Otherwise)	-3.71	-1.04	-0.03	-1.39	-0.06	-0.90	0.35	1.87	-0.059
Sigma, σ	3.10	3.83	0.07	9.64	0.16	5.44	0.44	5.38	
Class Probability (*t*-statistic)	0.076 (3.93)		0.577 (13.22)		0.227 (4.97)		0.120 (4.98)		

Number of Observations	768
Log-Likelihood at Zero	-437.00
Log-Likelihood at Convergence	-414.63
Akaike Information Criterion	**899.30**
Bayesian Information Criterion	1061.80

Table 3.9: Best Fit Latent Class Tobit Regression Estimates for Interstates

Variable	Latent Class 1		Latent Class 2		Partial Effect
	Coefficient	*t*-statistic	Coefficient	*t*-statistic	
Constant	0.72	5.05	-0.01	-0.20	
Speed Limit (1 if 75 MPH, 0 Otherwise)	0.00	0.02	-0.01	-1.49	-0.004
Total AADT (1 When Less Than 6,500, 0 Otherwise)	0.00	-0.10	0.03	2.23	0.007
Passenger Vehicle AADT (1 if Greater Than 15,000, 0 Otherwise)	-0.01	-0.29	-0.02	-0.61	-0.008
Heavy Vehicle AADT (1 if Between 2,000 and 3,000, 0 Otherwise)	0.00	-0.12	0.02	2.32	0.006
Horizontal Geometrics (1 if Curved, 0 Otherwise)	-0.01	-0.55	0.02	1.63	0.002
Road Configuration (1 if 2-Way and Raised/Depressed Divider)	-0.01	-0.24	0.01	0.48	0.002
Surface Defects (1 if No Surface Defects, 0 Otherwise)	-0.70	-4.81	0.01	0.36	-0.108
Sigma, σ	0.03	4.21	0.04	11.84	
Class Probability (*t*-statistic)	0.361 (2.30)		0.639 (4.06)		

Number of Observations	379
Log-Likelihood at Zero	363.56
Log-Likelihood at Convergence	385.48
Akaike Information Criterion	-733.00
Bayesian Information Criterion	**-658.10**

3.6 Discussion

3.6.1 Tobit Model

High passenger vehicle AADT (PAADT) decreases crash rates for each road classification and has a significant impact on crash rates based on the partial effects. Partial effects show that PAADT greater than 10,500 on principal arterials decreases the number of heavy vehicle crashes per MVMT by 0.084. Similarly, PAADT greater than 2,500 on major collectors reduces the number of heavy vehicle crashes per MVMT by 0.143 and PAADT greater than 15,000 on interstates reduces heavy vehicle crashes by 0.013 per MVMT. Conversely, low total AADT (passenger vehicles and heavy vehicles) increases crash rates. For instance, partial effects indicate that AADT less than 500 on major collectors increases the number of heavy vehicle crashes by 0.279

per MVMT. Likewise, AADT less than 6,500 on interstates results in a 0.016 crashes per MVMT increase. These findings are in line with previous work (Anastasopoulos and Mannering, 2009; Anastasopoulos et al., 2012a, 2012b, 2008; Dickerson et al., 2000; Qi et al., 2007; Zhou and Sisiopiku, 1997), in which lower AADT increases crash rates while higher AADT decreases crash rates. The same literature finds that the presence of heavy vehicle traffic decreases crash rates, yet the present study finds that the presence of heavy vehicle traffic increases crash rates for principal arterials and interstates. A possible explanation could be that such a finding is exclusive to the state of Idaho.

As mentioned previously, two variables were found to be random on major collectors based on the statistical significance of the mean and standard deviation. A speed limit less than or equal to 40 miles per hour was found to be random and normally distributed with a mean of 0.42 and standard deviation of 0.73. This suggests that for 28.25% of heavy vehicles the estimated parameter mean is less than zero and greater than zero for 71.75% of heavy vehicles. In other words, lower speed limits on major collectors increases crash rates for the majority of drivers and decreases crash rates for just over one-quarter of the drivers. Chen et al. (2014), however, found that lower speed limits increase crash rates for all observations using the random-effects Tobit model and may indicate that the random-effects approach is not accounting for all the heterogeneity in their dataset. On the other hand, high speed limits decrease crash rates for principal arterials and interstates. Speed limits of 65 miles per hour decrease crash rates on principal arterials and partial effects show a reduction of 0.041 crashes per

MVMT with such a speed limit. Interstates with a speed limit equal to 75 miles per hour see a decrease in crash rates and partial effects indicate a marginal decrease of 0.008 heavy vehicle crashes per MVMT.

As for road configuration, a specific configuration was found to be random and normally distributed, 2-way major collectors with a double-yellow painted divider. With a mean of 0.36 and standard deviation of 0.99, the estimated parameter mean is less than zero for 35.8% of heavy vehicle and greater than zero for 64.2% of heavy vehicles. This indicates that this road configuration for major collectors increases the crash rate for 64.2% of heavy vehicles and decreases the crash rate for 35.8%. Interstates see a reduction in crash rates due to 2-way interstates with a raised/depressed divider and have a small-scale partial effect of -0.014. Road configuration, however, on principal arterials increases crash rates—partial effects suggest that 2-way roads with a 2-way left-turn lane results in an increase of 0.041 heavy vehicle crashes per MVMT.

With regard to horizontal geometrics, straight and curved conditions increase crash rates for major collectors and interstates. Major collectors experience an increase in crash rates due to straight horizontal geometrics and partial effects show an increase of 0.210 heavy vehicle crashes per MVMT. Horizontal curves increase crash rates for interstates, though only increases the number of heavy vehicle crashes per MVMT by 0.010. Curved geometrics were found to increase crash risk by Yu et al. (2015), while the percent of curvature and degree of curvature were found to increase crash likelihood and crash rate by Qi et al. (2007) and Chen et al. (2014), respectively.

Other notable contributing crash rate factors are traffic control devices and surface defects. No traffic control devices on principal arterials decrease crash rates and partial effects indicate a reduction of 0.047 heavy vehicle crashes per MVMT. On the contrary, the presence of stop signs on major collectors increases heavy vehicle crashes by 0.116 per MVMT. Interstates with no surface defects (e.g. potholes, cracks, etc.) decrease crash rates. This variable has the largest effect on interstate crash rates, as partial effects suggest a decrease of 0.023 heavy vehicle crashes per MVMT.

3.6.2 Latent Class Model

The presence of latent classes indicate that several explanatory variables have differing effects on heavy vehicle crash rates. For example, 2-way roads with a 2-way left turn lane on principal arterials was positively significant in latent class 2, but negative and not significant in latent classes 1 and 3. These results indicate that there is heterogeneity in this variable and that such road configurations can have a negative and positive impact on crash rates—the high significance of the probability that a variable belongs to a specific class informs the analyst that the respective variable has varying effects. Similar findings are presented in each latent class specification and exist for each variable.

Looking at principal arterials, the partial effects of the Tobit model are significantly greater than those of the latent class model. The partial effect for PAADT greater than 10,500 using the Tobit model was -0.084, but according to the latent class model decreases the number of heavy vehicle crashes by 0.007 per MVMT. Overall, the partial effects for the latent class model were much less than the Tobit model.

Moving to major collectors, the latent class partial effects were substantially less when compared to the Tobit model. For example, PAADT greater than 2,500 has a partial effect of -0.143 for the Tobit model, yet the same variable based on the latent class model results in a reduction of 0.059 heavy vehicle crashes per MVMT.

Interstates, however, experienced a decrease in partial effects for some variables and an increase in others, even a change in signs for one variable. For instance, 2-way interstates with a raised/depressed divider has a partial effect of -0.014 for the Tobit model while for the latent class model has the opposite effect and results in an increase of 0.002 heavy vehicle crashes per MVMT. The partial effect of the Tobit model for PAADT greater than 15,000 is -0.013, but increases to -0.008 for the latent class model. Interstates with no surface defects decrease the number of heavy vehicle crashes per MVMT by 0.023 according to the Tobit model and increases the reduction based on the latent class model to 0.108 crashes per MVMT.

3.6.3 Model Comparison

To determine the best fit model for the Idaho crash data, three metrics were assessed: overall model fit, partial effect inferences and the rate of prediction to actual crash rate values. To illustrate, the Tobit regression method for principal arterial crash rates converges to a log-likelihood of -355.22 and the latent class approach converges to 195.60—the latter is markedly closer to zero[6]. To visualize, first take a look at the predicted values versus actual values for the Tobit regression and latent class regression estimates for principal arterials. The fit is much better for the latent class model and is quantified by the shown R^2 value[7]. Looking at the actual values versus the predicted values for the other two classifications, major collectors and interstates, the best fit model is certainly apparent.

---

[6] Although log-likelihoods are typically negative, it is possible to see positive values for regression with a continuous dependent variable. In such a case, the greater the value (if positive), the better the fit of the model.

[7] The R^2 value seen with the plots is the Pearson product moment correlation coefficient. This is ***not*** the same as the Maddala R^2 and measures the fit based on the spatial location of points on the graph.

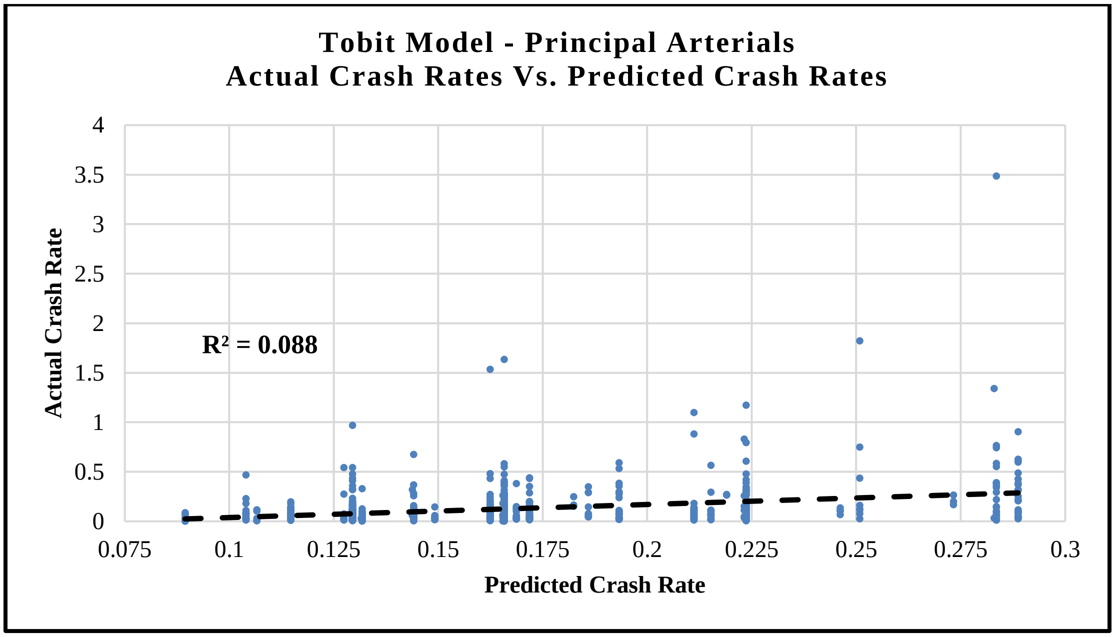

Figure 3.5: Tobit Model Actual Crash Rates vs. Predicted Crash Rates for Principal Arterials

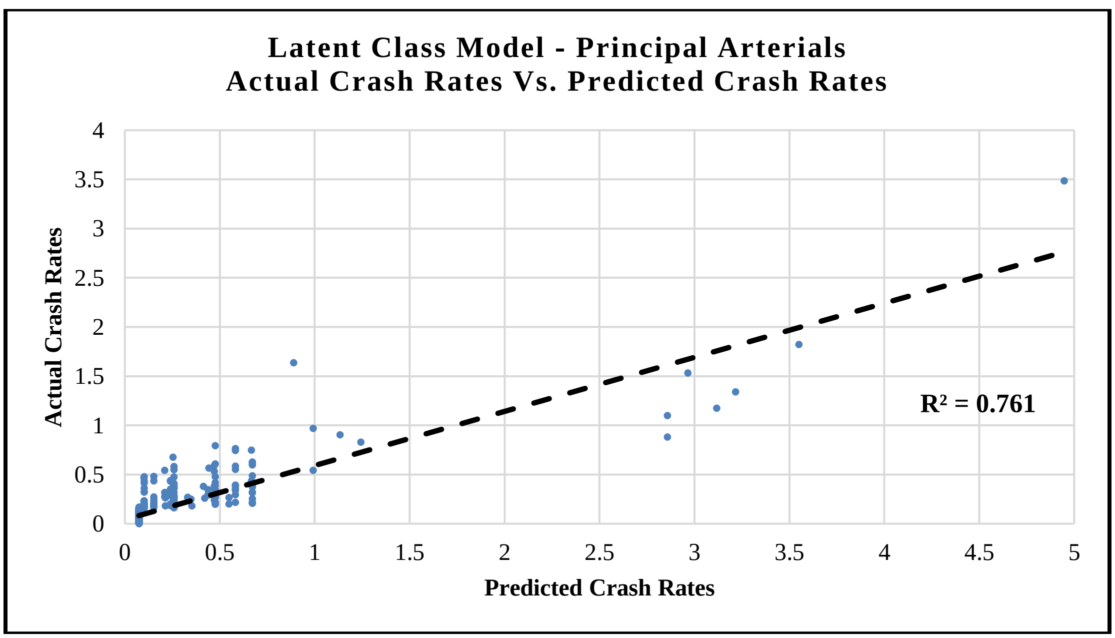

Figure 3.6: Latent Class Model Actual Crash Rates vs. Predicted Crash Rates for Principal Arterials

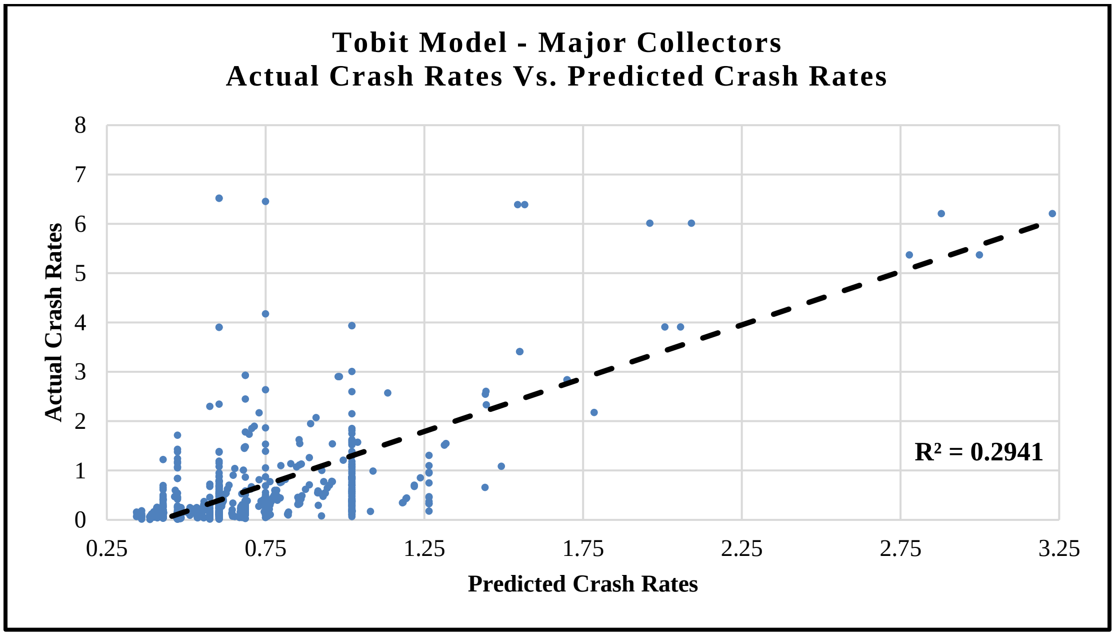

Figure 3.7: Tobit Model Actual Crash Rates vs. Predicted Crash Rates for Major Collectors

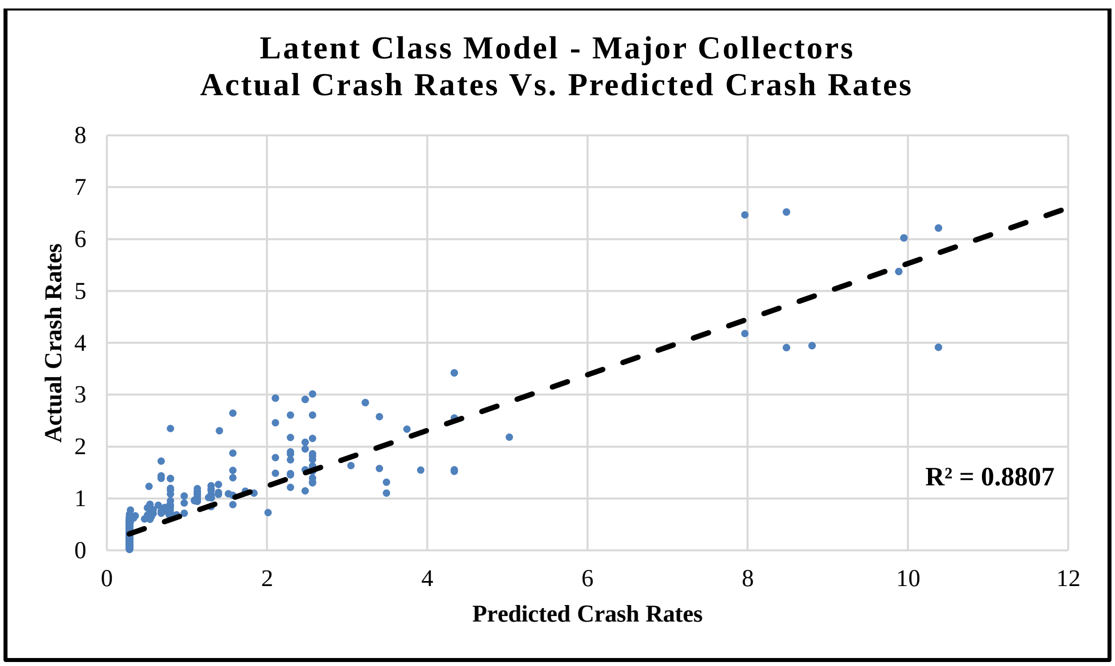

Figure 3.8: Latent Class Model Actual Crash Rates vs. Predicted Crash Rates for Major Collectors

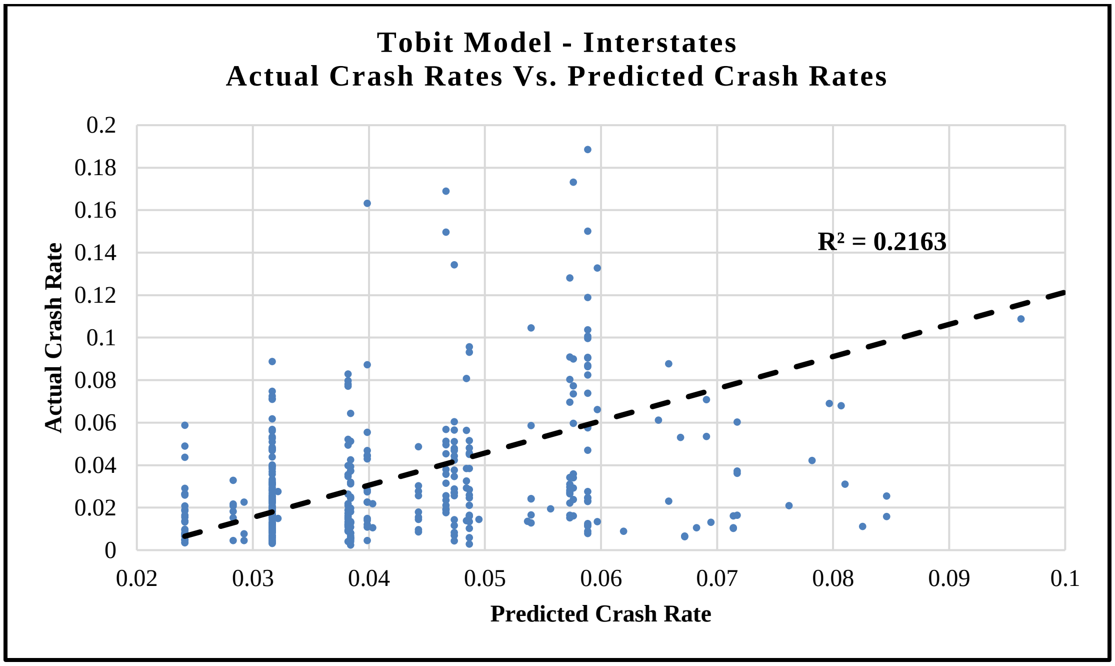

Figure 3.9: Tobit Model Actual Crash Rates vs. Predicted Crash Rates for Interstates

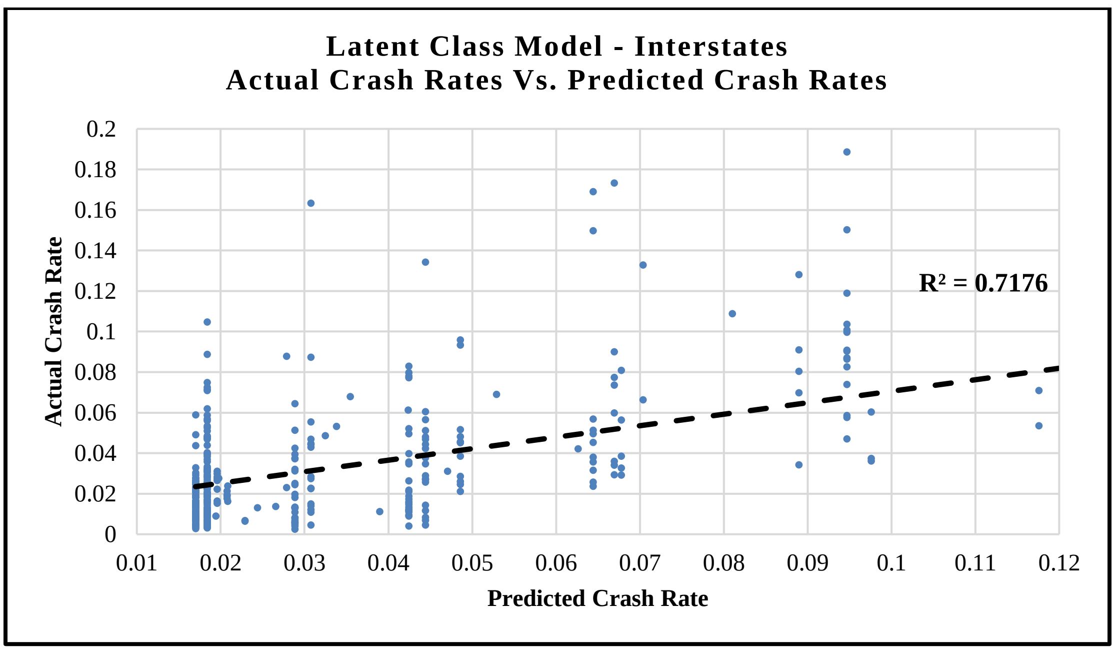

Figure 3.10: Latent Class Model Actual Crash Rates vs. Predicted Crash Rates for Interstates

3.7 Insights & Future Work

This study concentrates on two specific frameworks, random-parameter Tobit regression and latent class regression, to determine factors that contribute to the number of heavy vehicle crashes per MVMT by roadway classification while identifying a preferred method to account for unobserved heterogeneity. Crash data is often missing key variables and has variation across the existing variables, therefore employing the random-parameter Tobit method allows the analyst to define a distribution and allow a parameter to vary as a means to account for heterogeneity. The latent class approach accounts for heterogeneity differently, as no distribution is defined and the parameters are permitted to vary across the specified number of classes. Using overall model fit, partial effect inferences and correctly predicted crash rate values, the estimates of the two approaches were examined.

The Idaho case study provides new insights into heavy vehicle crash rates by roadway classification. To illustrate, different road configurations, horizontal geometrics and traffic control devices were found to be significant for each road classification. A specific road configuration was found to decrease crash rates for major collectors and interstates, but increase crash rates on principal arterials. Curved horizontal geometrics increase crash rates on interstates and straight horizontal geometrics increase crash rates on major collectors. Stop signs on major collectors increase crash rates, yet no traffic control devices on principal arterials decrease crash rates. High speed limits decrease crash crates on principal arterials and interstates, and although lower speed limits were found to mostly decrease crash rates, lower speed limits can increase crash rates for a proportion of the heavy vehicles. The most common insight from this study, as well as previous studies, is that high traffic volumes decrease crash rates and low volumes increase crash rates. With that in mind, different from past literature, this study found that the presence of heavy vehicles increases crash rates.

To assess the accuracy of the two frameworks, the actual crash rates and predicted crash rates were plotted and the Pearson product moment correlation coefficient was provided for each. In doing so, the latent class regression approach outperformed Tobit regression for each road classification and should be considered in future crash rate analyses that consider the use of Tobit regression. In addition, with regard to the best fit number of latent classes, the sample size may indicate what information criterion (AIC or BIC) to use when selecting the best fit number of latent classes.

In summary, this study exhibits two distinct methodologies to model crash rates while accounting for heterogeneity. Using road classification in Idaho as a case study to test the methodology that better estimates crash rates, it was determined that the latent class approach is superior. Moreover, factors that contribute to crash rates differ dependent on road classification and in future work should be modeled separately. Such findings can assist with safety measures in Idaho by providing agencies and policy makers with factors that influence crash rates with more precision, by road classification. The presented framework, censored latent class regression, should strongly be considered when conducting future crash rate analyses.

REFERENCES

Aguero-Valverde, J., & Jovanis, P. P. (2008). Analysis of Road Crash Frequency with Spatial Models. *Transportation Research Record: Journal of the Transportation Research Board*, *2061*(1), 55–63. doi:10.3141/2061-07

Anastasopoulos, P. C., & Mannering, F. L. (2009). A Note on Modeling Vehicle Accident Frequencies With Random-Parameters Count Models. *Accident Analysis and Prevention*, *41*(1), 153–159. doi:10.1016/j.aap.2008.10.005

Anastasopoulos, P. C., Mannering, F. L., Shankar, V. N., & Haddock, J. E. (2012). A Study of Factors Affecting Highway Accident Rates Using the Random-Parameters Tobit Model. *Accident Analysis and Prevention*, *45*, 628–633. doi:10.1016/j.aap.2011.09.015

Anastasopoulos, P. C., Shankar, V. N., Haddock, J. E., & Mannering, F. L. (2012). A Multivariate Tobit Analysis of Highway Accident-Injury-Severity Rates. *Accident Analysis and Prevention*, *45*, 110–119. doi:10.1016/j.aap.2011.11.006

Anastasopoulos, P. C., Tarko, A. P., & Mannering, F. L. (2008). Tobit Analysis of Vehicle Accident Rates on Interstate Highways. *Accident Analysis & Prevention*, *40*(2), 768–775. doi:10.1016/j.aap.2007.09.006

Bhat, C. R. (2003). Simulation Estimation of Mixed Discrete Choice Models Using Randomized and Scrambled Halton Sequences. *Transportation Research Part B: Methodological*, *37*(9), 837–855. doi:10.1016/S0191-2615(02)00090-5

Bhat, C. R., Born, K., Sidharthan, R., & Bhat, P. C. (2014). A Count Data Model with Endogenous Covariates: Formulation and Application to Roadway Crash Frequency at Intersections. *Analytic Methods in Accident Research*, *1*, 53–71. doi:10.1016/j.amar.2013.10.001

Blincoe, L., Miller, T. R., Zaloshnja, E., & Lawrence, B. A. (2015). *The Economical and Societal Impact of Motor Vehicle Crashes, 2010 (Revised May 2015)* (Vol. 2015). Retrieved from http://www-nrd.nhtsa.dot.gov/pubs/812013.pdf

Brown, S., Greene, W. H., Harris, M. N., & Taylor, K. (2015). An Inverse Hyperbolic Sine Heteroskedastic Latent Class Panel Tobit Model: An Application to Modelling Charitable Donations. *Economic Modelling*, *50*, 228–236. doi:10.1016/j.econmod.2015.06.018

Bureau of Transportation Statistics. (2015). *Freight Facts and Figures*. (W. H. Moore, Ed.) (11th ed.). Washington, D.C.: United States Department of Transportation. Retrieved from http://ops.fhwa.dot.gov/freight/freight_analysis/nat_freight_stats/docs/12factsfigures/table3_8.htm

Carlsson, F., & Martinsson, P. (2007). Willingness to Pay Among Swedish Households to Avoid Power Outages: A Random Parameter Tobit Model Approach. *Energy Journal*, *28*(1), 75–89. doi:10.5547/ISSN0195-6574-EJ-Vol28-No1-4

Castro, M., Paleti, R., & Bhat, C. R. (2012). A Latent Variable Representation of Count Data Models to Accommodate Spatial and Temporal Dependence: Application to Predicting Crash Frequency at Intersections. *Transportation Research Part B: Methodological*, *46*(1), 253–272. doi:10.1016/j.trb.2011.09.007

Chen, E., & Tarko, A. P. (2012). Analysis of Crash Frequency in Work Zones with Focus on Police Enforcement. *Transportation Research Record: Journal of the Transportation Research Board*, *2280*, 127–134. doi:10.3141/2280-14

Chen, F., Ma, X., & Chen, S. (2014). Refined-Scale Panel Data Crash Rate Analysis Using Random-Effects Tobit Model. *Accident Analysis and Prevention*, *73*, 323–332. doi:10.1016/j.aap.2014.09.025

Collins, L. M., Fidler, P. L., Wugalter, S. E., & Long, J. D. (1993). Goodness-of-Fit Testing for Latent Class Models. *Multivariate Behavioral Research*, *28*(3), 375–389.

Dickerson, A., Peirson, J., & Vickerman, R. (2000). Road Accidents and Traffic Flows: An Econometric Investigation. *Economica*, *67*(265), 101–121.

Federal Motor Carrier Safety Administration. (2013). Commercial Motor Vehicle Facts. Retrieved February 11, 2015, from https://www.fmcsa.dot.gov/safety/data-and-statistics/commercial-motor-vehicle-facts-—-march-2013

Federal Motor Carrier Safety Administration. (2016). *Large Truck and Bus Crash Facts 2014*.

Golembiewski, G. A., & Chandler, B. (2011). *Roadway Departure Safety: A Manual for Local Rural Road Owners*. Washington, D.C.

Greene, W. H. (2012a). *Econometric Analysis* (7th ed., Vol. 97). New York: Pearson Education. doi:10.1198/jasa.2002.s458

Greene, W. H. (2012b). *Econometric Modeling Guide* (10th ed.). Plainview, NY: Econometric Software, Inc.

Hagenaars, J. A., & McCutcheon, A. L. (Eds.). (2002). *Applied Latent Class Analysis*. New York: Cambridge University Press.

Halton, J. H. (1960). On The Efficiency Of Certain Quasi-Random Sequences of Points in Evaluating Multi-Dimension Integrals. *Numerische Mathematik*, *2*, 84–90.

Idaho Office of Highway Safety. (2014). *Idaho Traffic Crashes*.

Islam, M., & Hernandez, S. (2015). Fatality Rates for Crashes Involving Heavy Vehicles on Highways: A Random Parameter Tobit Regression Approach. *Journal of Transportation Safety & Security*, *8*(3), 247–265. doi:10.1080/19439962.2015.1027071

Islam, S., Jones, S. L., & Dye, D. (2014). Comprehensive Analysis of Single and Multi-Vehicle Large Truck At-Fault Crashes on Rural and Urban Roadways In Alabama. *Accident Analysis & Prevention*, *67*, 148–158. doi:10.1016/j.aap.2014.02.014

Jedidi, K., Ramaswamy, V., & Desarbo, W. S. (1993). A Maximum Likelihood Method for Latent Class Regression Involving a Censored Dependent Variable. *Psychometrika*, *58*(3), 375–394. doi:10.1007/BF02294647

Khorashadi, A., Niemeier, D., Shankar, V., & Mannering, F. (2005). Differences in Rural and Urban Driver-Injury Severities in Accidents Involving Large-Trucks: An Exploratory Analysis. *Accident Analysis and Prevention*, *37*(5), 910–921. doi:10.1016/j.aap.2005.04.009

Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). *Stated Choice Methods: Analysis and Applications*. New York: Cambridge University Press. doi:10.1002/jae.701

Magidson, J., & Vermunt, J. K. (2004). Latent Class Models, (617), 11.

Mannering, F. L., Shankar, V., & Bhat, C. R. (2016). Unobserved Heterogeneity and the Statistical Analysis of Highway Accident Data. *Analytic Methods in Accident Research*, *11*, 1–16. doi:10.1016/j.amar.2016.04.001

Milton, J. C., Shankar, V. N., & Mannering, F. L. (2008). Highway Accident Severities and the Mixed Logit Model: An Exploratory Empirical Analysis. *Accident Analysis and Prevention*, *40*(1), 260–266. doi:10.1016/j.aap.2007.06.006

Moore, D. N., Schneider IV, W. H., Savolainen, P. T., & Farzaneh, M. (2011). Mixed Logit Analysis of Bicyclist Injury Severity Resulting From Motor Vehicle Crashes at Intersetion and Non-Intersection Locations. *Accident Analysis and Prevention*, *43*(3), 621–630. doi:10.1016/j.aap.2010.09.015

NHTSA. (n.d.). NCSA DATA RESOURCE WEBSITE.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling*, *14*(4), 535–569.

Pahukula, J., Hernandez, S., & Unnikrishnan, A. (2015). A Time of Day Analysis of Crashes Involving Large Trucks in Urban Areas. *Accident Analysis & Prevention*, *75*, 155–163. doi:10.1016/j.aap.2014.11.021

Park, E. S., & Lord, D. (2007). Multivariate Poisson-Lognormal Models for Jointly Modeling Crash Frequency by Severity. *Transportation Research Record: Journal of the Transportation Research Board*, *2019*, 1–6. doi:10.3141/2019-01

Qi, Y. G., Smith, B. L., & Guo, J. (2007). Freeway Accident Likelihood Prediction Using a Panel Data Analysis Approach. *Journal of Transportation Engineering*, *133*(3), 149–156. doi:10.1061/(ASCE)0733-947X(2007)133:3(149)

Shaheed, M. S., & Gkritza, K. (2014). A Latent Class Analysis of Single-Vehicle Motorcycle Crash Severity Outcomes. *Analytic Methods in Accident Research*, *2*, 30–38. doi:10.1016/j.amar.2014.03.002

Shea, M. S., Le, T. Q., & Porter, R. J. (2015). Combined Crash Frequency–Crash Severity Evaluation of Geometric Design Decisions. *Transportation Research Record: Journal of the Transportation Research Board*, *2521*, 54–63. doi:10.3141/2521-06

Shen, J. (2009). Latent Class Model or Mixed Logit Model? A Comparison by Transport Mode Choice Data. *Applied Economics*, *41*(22), 2915–2924. doi:10.1080/00036840801964633

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica*, *26*(1), 24–36.

Train, K. (1999). Halton Sequences for Mixed Logit. *Working Papers*, 1–18.

Ukkusuri, S., Hasan, S., & Aziz, H. M. A. (2011). Random Parameter Model Used to Explain Effects of Built-Environment Characteristics on Pedestrian Crash Frequency. *Transportation Research Record: Journal of the Transportation Research Board*, *2237*, 98–106. Retrieved from http://trrjournalonline.trb.org/doi/pdf/10.3141/2237-11

Veall, M. R., & Zimmerman, K. F. (1996). Pseudo-R2 Measures for Some Common Limited Dependent Variable Models. *Journal of Economic Surveys*, *10*(3).

Washington, S. P., Karlaftis, M. G., & Mannering, F. L. (2010). *Statistical and Econometric Methods for Transportation Data Analysis*. *Chapman & Hall/CRC* (2nd ed.). Chapman & Hall/CRC. doi:10.1198/tech.2004.s238

Xie, K., Wang, X., Ozbay, K., & Yang, H. (2014). Crash Frequency Modeling for Signalized Intersections in a High-Density Urban Road Network. *Analytic Methods in Accident Research*, *2*, 39–51. doi:10.1016/j.amar.2014.06.001

Xie, Y., & Zhang, Y. (2008). Crash Frequency Analysis with Generalized Additive Models. *Transportation Research Record: Journal of the Transportation Research Board*, (2061), pp 39–45. doi:10.3141/2061-05

Xu, P., & Huang, H. (2015). Modeling Crash Spatial Heterogeneity: Random Parameters Versus Geographically Weighting. *Accident Analysis and Prevention*, *75*, 16–25. doi:10.1016/j.aap.2014.10.020

Yang, C. C. (2004). Evaluating Latent Class Analysis Models in Qualitative Phenotype Identification. *Computational Statistics and Data Analysis*, *50*(4), 1090–1104. doi:10.1016/j.csda.2004.11.004

Ye, X., Pendyala, R. M., Washington, S. P., Konduri, K., & Oh, J. (2009). A Simultaneous Equations Model of Crash Frequency by Collision Type for Rural Intersections. *Safety Science*, *47*(3), 443–452. doi:10.1016/j.ssci.2008.06.007

Yu, R., Xiong, Y., & Abdel-Aty, M. (2015). A Correlated Random Parameter Approach to Investigate the Effects of Weather Conditions on Crash Risk for a Mountainous Freeway. *Transportation Research Part C: Emerging Technologies*, *50*, 68–77. doi:10.1016/j.trc.2014.09.016

Zhang, Y., Xie, Y., & Li, L. (2012). Crash Frequency Analysis of Different Types of Urban Roadway Segments Using Generalized Additive Model. *Journal of Safety Research*, *43*(2), 107–114. doi:10.1016/j.jsr.2012.01.003

Zhou, M., & Sisiopiku, V. (1997). Relationship Between Volume-to-Capacity Ratios and Accident Rates. *Transportation Research Record*, *1581*, 47–52. doi:10.3141/1581-06

Zhu, X., & Srinivasan, S. (2011). A Comprehensive Analysis of Factors Influencing the Injury Severity of Large-Truck Crashes. *Accident Analysis and Prevention*, *43*(1), 49–57. doi:10.1016/j.aap.2010.07.007

4.0 CONCLUSION

The preceding two chapters are comprised of individual articles that together advanced the literature on heavy vehicle involved crash analysis by exploring and comparing recent analytical methods in crash research, namely the *random-parameter* and *latent class* frameworks. Furthermore, these econometric techniques provide a mechanism to account for unobserved heterogeneity and allows the analyst to better understand injury severity and crash rates as a result of roadway classification for heavy vehicle related crashes. This chapter will summarize the key findings of both manuscripts, discuss the practical applications of this research and consider future research as a result of this study.

4.1 Summary of Key Findings

Using Idaho as a case study, Chapter 2 compares the estimates of two advanced econometric frameworks while statistically validating if road classifications should be modeled separately. In the process, several significant injury severity contributing factors were determined and the most statistically significant modeling framework for the Idaho dataset was identified. The model separation test indicates with 99.99% confidence that road classifications should be considered separately and was further illustrated by the large number of variables that were exclusive to a specific classification. In addition, the latent class framework provided better estimates and correctly predicted more injury severity outcomes.

Two variables were significant in predicting injury severities for each road classification, cloudy weather conditions and horizontal curves. Cloudy weather was also found to be significant in previous studies, but differed in impact depending on geographical location. For instance, Kim et al. (2013) found cloudy conditions to increase the probability of fatal crashes in California, yet Mohamed et al. (2013) found cloudy weather to decrease the probability of fatal crashes in New York and Montreal. In Idaho, cloudy weather decreases the probability of an injury on each road classification and provides further evidence that drivers experienced in driving in cloudy conditions are not prone to sustain a serious injury in a crash. A possible explanation could be that drivers in such locations are practiced at driving in cloudy conditions and drive with more caution. In regard to horizontal curvature, this was found to decrease the outcome probability of no injury crashes on major collectors and interstates, but increase the probability of major injury crashes on interstates.

In Chapter 3, the study was extended to focus on another common safety metric, crashes per vehicle-miles-traveled, and by applying two advanced econometric frameworks, specifically the *random-parameter* and *latent class* frameworks to a Tobit regression approach. Based on the findings from Chapter 2, road classifications should be considered independently, hence the same three road classifications were studied. Chapter 3 took an alternate approach to the traditional frequency/count models by using a censoring method to predict crash rates, the Tobit regression framework. As with Chapter 2, this method was compared to the estimates of its latent class equivalent to identify the most statistically significant approach. The latent class method was found to better predict crash rates, as it better predicted injury severities in Chapter 2. Unlike Chapter 2, similar variables were found to be significant for each classification.

Low total traffic volumes (volumes that include both passenger vehicles and heavy vehicles) were found to increase crash rates on each classification and high traffic volumes were found to decrease crash rates. Previous studies also found that high traffic volume decreases crash rates and low traffic volumes increase crash rates (Anastasopoulos and Mannering, 2009; Anastasopoulos et al., 2012a, 2012b, 2008; Dickerson et al., 2000; Qi et al., 2007; Zhou and Sisiopiku, 1997). However, the same studies find that the presence of heavy vehicles decreases crash rates, but the present research finds that any volume of heavy vehicles increases crash rates for all road classifications. This could be attributed to the location of the case study, Idaho, in which the presence of heavy vehicles increases crash rates for this region of the United States.

Road configuration and posted speed limits impact crash rates by road classification. Specifically, 2-way interstates with a depressed/raised divider reduce crash rates and 2-way principal arterials with a 2-way left-turn lane increase crash rates. Major collectors, however, experience differing effects due to 2-way, double-yellow painted divider configurations. In regards to posted speed limits, higher speed limits reduce crash rates on principal arterials and interstates, yet lower speed limits reduce crash rates for some heavy vehicles and increase crash rates for others on major collectors.

In summary, this thesis has identified statistically significant frameworks for injury severity and crash rate analyses while determining if roadway classifications need to be considered separately. By comparing predicted injury severity outcomes, overall model fit and marginal effects for the random-parameter and latent class methods, the latent class method proved to be the more statistically significant framework for modeling heavy vehicle injury severity in Idaho. Likewise, by comparing predicted crash rates to actual crash rates for the random-parameter Tobit and latent class Tobit methods, the latent class method performed substantially better. In using these methods, this thesis was able to capture both observable and unobservable factors that contribute to heavy vehicle injury severity and crash rates. In regard to roadway classification, a model separation test statistically determined that heavy vehicle crashes should be considered by roadway classification with 99.99% confidence. Ultimately, the framework provided in this thesis can shed light on the contributing factors to heavy vehicle crashes when taking into account heterogeneity, and as a result, will strengthen current and future transportation safety tools.

4.2 Real-World Applications

In general, advanced crash analyses will contribute to the evolution of transportation safety and assist transportation engineers, planners and agencies in mitigating future crashes. The current thesis expands on the traditional crash analysis frameworks and generates new, insightful information regarding heavy vehicle crashes. Specifically, findings in this thesis can provide guidance to transportation engineers, planners, agencies and policy makers to direct their safety efforts with more precision to augment monetary investments and overall benefits. The contributing factors identified in this research can provide guidance to the aforesaid entities to enhance heavy vehicle safety by roadway classification.

Although the methodologies may not be explicitly used by agencies, the results obtained through the modeling framework presented in this work can assist in identifying appropriate countermeasures to help mitigate heavy vehicle crashes. For example, no traffic control devices on principal arterials increases the probability of sustaining a minor injury during a crash, hence an economical solution could be to place traffic signs along principal arterial routes where warranted to control for excessive speeds. In regards to crash rates, specific configurations increase the number of crashes per MVMT on principal arterials and major collectors and striping could be a possible solution to reduce heavy vehicle crashes on these classifications.

By identifying factors by roadway classification, city, county, state and federal agencies can appropriate funding by prioritizing transportation safety projects adequately. Injury severity results can provide guidance by identifying high risk roadway segments with factors that lead to more severe injuries, while the crash rate results can assist in identifying roadway classifications and factors that increase the number of heavy vehicle crashes. In using these results, Idaho and the trucking industry and utilize their assets to mitigate major injury crashes while decreasing the total number of crashes.

4.3 Future Work

The methodological approach to this research has the potential to open research doors in several directions. Until a uniform approach that is capable of capturing each and every variable needed to predict injury severity outcomes and crashes, heterogeneity is going to be present. The latent class framework, based on the empirical results from the Idaho case study, produce the most statistically significant results by addressing more of the heterogeneity. Furthermore, the latent class framework can be extended to include random parameters within classes as an attempt to capture more of the unobserved heterogeneity. The results of this research are promising, but it is highly recommended that future work build upon the latent class framework in a united pursuit of the most accurate and effective method to analyze transportation safety data.

In addition to the latent class method, futures studies are needed that focus on crash analysis by roadway classification. If classifications should be investigated separately in Idaho, it is likely to see similar results in alternate regions. Moreover, each road classification in Idaho has several differing variables that effect injury severity and/or crash rates. To further crash analysis research, in general, disaggregating datasets by key aspects will provide guidance at a higher accuracy and assist agencies, engineers and policy makers to pin-point specific issues. Horizontal geometrics were found to impact crash rates, yet horizontal curves are present to reduce costs. Another avenue for future work would be to conduct a study on the construction costs, lifecycle costs and costs due to crashes to determine what is best financially and for mitigating crashes.

To fill the notable gap in literature regarding roadway classification, heavy vehicle injury severity and crash rates, this thesis explored two advanced econometric techniques used to capture unobserved heterogeneity. This study illustrates the significance of disaggregating data to conduct crash analyses and provides the groundwork needed to continue to extend heavy vehicle crash research.

5.0 REFERENCES

Abdel-Aty, M.A., Radwan, A.E., 2000. Modeling Traffic Accident Occurrence and Involvement. Accid. Anal. Prev. 32, 633–642.

Aguero-Valverde, J., Jovanis, P.P., 2008. Analysis of Road Crash Frequency with Spatial Models. Transp. Res. Rec. J. Transp. Res. Board 2061, 55–63.

Anastasopoulos, P.C., Mannering, F.L., 2009. A Note on Modeling Vehicle Accident Frequencies With Random-Parameters Count Models. Accid. Anal. Prev. 41, 153–159.

Anastasopoulos, P.C., Mannering, F.L., 2011. An Empirical Assessment of Fixed and Random Parameter Logit Models Using Crash- and Non-Crash-Specific Injury Data. Accid. Anal. Prev. 43, 1140–1147.

Anastasopoulos, P.C., Mannering, F.L., Shankar, V.N., Haddock, J.E., 2012a. A Study of Factors Affecting Highway Accident Rates Using the Random-Parameters Tobit Model. Accid. Anal. Prev. 45, 628–633.

Anastasopoulos, P.C., Shankar, V.N., Haddock, J.E., Mannering, F.L., 2012b. A Multivariate Tobit Analysis of Highway Accident-Injury-Severity Rates. Accid. Anal. Prev. 45, 110–119.

Anastasopoulos, P.C., Tarko, A.P., Mannering, F.L., 2008. Tobit Analysis of Vehicle Accident Rates on Interstate Highways. Accid. Anal. Prev. 40, 768–775.

Behnood, A., Roshandeh, A.M., Mannering, F.L., 2014a. Latent Class Analysis of the Effects of Age, Gender, and Alcohol Consumption on Driver-Injury Severities. Anal. Methods Accid. Res. 3-4, 56–91.

Behnood, A., Roshandeh, A.M., Mannering, F.L., 2014b. Latent Class Analysis of the Effects of Age, Gender and Alcohol Consumption on Driver-Injury Severities. Anal. Methods Accid. Res. 3-4, 56–91.

Bhat, C.R., 2003. Simulation Estimation of Mixed Discrete Choice Models Using Randomized and Scrambled Halton Sequences. Transp. Res. Part B Methodol. 37, 837–855.

Bhat, C.R., Born, K., Sidharthan, R., Bhat, P.C., 2014. A Count Data Model with Endogenous Covariates: Formulation and Application to Roadway Crash Frequency at Intersections. Anal. Methods Accid. Res. 1, 53–71.

Blincoe, L., Miller, T.R., Zaloshnja, E., Lawrence, B.A., 2015. The Economical and Societal Impact of Motor Vehicle Crashes, 2010 (Revised May 2015).

Brown, S., Greene, W.H., Harris, M.N., Taylor, K., 2015. An Inverse Hyperbolic Sine Heteroskedastic Latent Class Panel Tobit Model: An Application to Modelling Charitable Donations. Econ. Model. 50, 228–236.

Bureau of Transportation Statistics, 2015. Freight Facts and Figures, 11th ed. United States Department of Transportation, Washington, D.C.

Carlsson, F., Martinsson, P., 2007. Willingness to Pay Among Swedish Households to Avoid Power Outages: A Random Parameter Tobit Model Approach. Energy J. 28, 75–89.

Carson, J., Mannering, F., 2001. The Effect of Ice Warning Signs on Ice-Accident Frequencies and Severities. Accid. Anal. Prev. 33, 99–109.

Castro, M., Paleti, R., Bhat, C.R., 2012. A Latent Variable Representation of Count Data Models to Accommodate Spatial and Temporal Dependence: Application to Predicting Crash Frequency at Intersections. Transp. Res. Part B Methodol. 46, 253–272.

Cerwick, D.M., Gkritza, K., Shaheed, M.S., Hans, Z., 2014. A Comparison of the Mixed Logit and Latent Class Methods for Crash Severity Analysis. Anal. Methods Accid. Res. 3-4, 11–27.

Chen, E., Tarko, A.P., 2012. Analysis of Crash Frequency in Work Zones with Focus on Police Enforcement. Transp. Res. Rec. J. Transp. Res. Board 2280, 127–134.

Chen, F., Ma, X., Chen, S., 2014. Refined-Scale Panel Data Crash Rate Analysis Using Random-Effects Tobit Model. Accid. Anal. Prev. 73, 323–332.

Chin, H.C., Quddus, M.A., 2003. Applying the Random Effect Negative Binomial Model to Examine Traffic Accident Occurrence at Signalized Intersections. Accid. Anal. Prev. 35, 253–259.

Chu, H.C., 2014. Assessing Factors Causing Severe Injuries in Crashes of High-Deck Buses in Long-Distance Driving on Freeways. Accid. Anal. Prev. 62, 130–136.

Collins, L.M., Fidler, P.L., Wugalter, S.E., Long, J.D., 1993. Goodness-of-Fit Testing for Latent Class Models. Multivariate Behav. Res. 28, 375–389.

Dickerson, A., Peirson, J., Vickerman, R., 2000. Road Accidents and Traffic Flows: An Econometric Investigation. Economica 67, 101–121.

Eluru, N., Bagheri, M., Miranda-Moreno, L.F., Fu, L., 2012. A Latent Class Modeling Approach for Identifying Vehicle Driver Injury Severity Factors at Highway-Railway Crossings. Accid. Anal. Prev. 47, 119–127.

Federal Motor Carrier Safety Administration, 2013. Commercial Motor Vehicle Facts [WWW Document]. URL https://www.fmcsa.dot.gov/safety/data-and-statistics/commercial-motor-vehicle-facts-—-march-2013 (accessed 2.11.15).

Federal Motor Carrier Safety Administration, 2016. Large Truck and Bus Crash Facts 2014.

Geedipally, S., Turner, P., Patil, S., 2011. Analysis of Motorcycle Crashes in Texas with Multinomial Logit Model. Transp. Res. Rec. J. Transp. Res. Board 2265, 62–69.

Gkritza, K., Mannering, F.L., 2008. Mixed Logit Analysis of Safety-Belt Use in Single- and Multi-Occupant Vehicles. Accid. Anal. Prev. 40, 443–451.

Golembiewski, G.A., Chandler, B., 2011. Roadway Departure Safety: A Manual for Local Rural Road Owners. Washington, D.C.

Greene, W.H., 2012a. LIMDEP Version 10 Reference Guide, 10th ed. Econometric Software, Inc., New York.

Greene, W.H., 2012b. NLOGIT 5 Reference Guide. Econometric Software, Inc., New York.

Greene, W.H., 2012. Econometric Analysis, 7th ed. Pearson Education, New York.

Greene, W.H., 2012c. Econometric Modeling Guide, 10th ed. Econometric Software, Inc., Plainview, NY.

Greene, W.H., Hensher, D.A., 2003. A Latent Class Model for Discrete Choice Analysis: Contrasts with Mixed Logit. Transp. Res. Part B Methodol. 37, 681–698.

Hagenaars, J.A., McCutcheon, A.L. (Eds.), 2002. Applied Latent Class Analysis. Cambridge University Press, New York.

Halton, J.H., 1960. On The Efficiency Of Certain Quasi-Random Sequences of Points in Evaluating Multi-Dimension Integrals. Numer. Math. 2, 84–90.

Idaho Office of Highway Safety, 2014. Idaho Traffic Crashes.

Islam, M., Hernandez, S., 2013a. Modeling Injury Outcomes of Crashes Involving Heavy Vehicles on Texas Highways. Transp. Res. Rec. 2388, 28–36.

Islam, M., Hernandez, S., 2013b. Large Truck–Involved Crashes: Exploratory Injury Severity Analysis. J. Transp. Eng. 139, 596–604.

Islam, M., Hernandez, S., 2015. Fatality Rates for Crashes Involving Heavy Vehicles on Highways: A Random Parameter Tobit Regression Approach. J. Transp. Saf. Secur. 8, 247–265.

Islam, S., Jones, S.L., Dye, D., 2014a. Comprehensive Analysis of Single and Multi-Vehicle Large Truck At-Fault Crashes on Rural and Urban Roadways In Alabama. Accid. Anal. Prev. 67, 148–158.

Islam, S., Jones, S.L., Dye, D., 2014b. Comprehensive Analysis of Single- and Multi-Vehicle Large Truck At-Fault Crashes on Rural and Urban Roadways in Alabama. Accid. Anal. Prev. 72, 267–276.

Jedidi, K., Ramaswamy, V., Desarbo, W.S., 1993. A Maximum Likelihood Method for Latent Class Regression Involving a Censored Dependent Variable. Psychometrika 58, 375–394.

Kashani, A.T., Mohaymany, A.S., 2011. Analysis of the Traffic Injury Severity on Two-Lane, Two-Way Rural Roads Based on Classification Tree Models. Saf. Sci. 49, 1314–1320.

Khorashadi, A., Niemeier, D., Shankar, V., Mannering, F., 2005. Differences in Rural and Urban Driver-Injury Severities in Accidents Involving Large-Trucks: An Exploratory Analysis. Accid. Anal. Prev. 37, 910–921.

Kim, J.K., Ulfarssom, G.F., Shankar, V.N., Mannering, F.L., 2010. A Note on Modeling Pedestrian-Injury Severity in Motor-Vehicle Crashes with the Mixed Logit Model. Accid. Anal. Prev. 42, 1751–1758.

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

Lee, J., Mannering, F., 2002. Impact of Roadside Features on the Frequency and Severity of Run-Off-Roadway Accidents: An Empirical Analysis. Accid. Anal. Prev. 34, 149–161.

Louviere, J.J., Hensher, D.A., Swait, J.D., 2000. Stated Choice Methods: Analysis and Applications. Cambridge University Press, New York.

Magidson, J., Vermunt, J.K., 2004. Latent Class Models 11.

Malyshkina, N. V., Mannering, F.L., Tarko, A.P., 2009. Markov Switching Negative Binomial Models: An Application to Vehicle Accident Frequencies. Accid. Anal. Prev. 41, 217–226.

Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved Heterogeneity and the Statistical Analysis of Highway Accident Data. Anal. Methods Accid. Res. 11, 1–16.

McFadden, D., Train, K., 2000. Mixed MNL Models for Discrete Response. J. Appl. Econom. 15, 447–470.

Milton, J.C., Shankar, V.N., Mannering, F.L., 2008. Highway Accident Severities and the Mixed Logit Model: An Exploratory Empirical Analysis. Accid. Anal. Prev. 40, 260–266.

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

Moore, D.N., Schneider IV, W.H., Savolainen, P.T., Farzaneh, M., 2011. Mixed Logit Analysis of Bicyclist Injury Severity Resulting From Motor Vehicle Crashes at Intersetion and Non-Intersection Locations. Accid. Anal. Prev. 43, 621–630.

Morgan, A., Mannering, F.L., 2011. The Effects of Road-Surface Conditions, Age, and Gender on Driver-Injury Severities. Accid. Anal. Prev. 43, 1852–1863.

NHTSA, n.d. NCSA DATA RESOURCE WEBSITE [WWW Document].

Nylund, K.L., Asparouhov, T., Muthén, B.O., 2007. Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. Struct. Equ. Model. 14, 535–569.

Obeng, K., 2011. Gender Differences in Injury Severity Risks in Crashes at Signalized Intersections. Accid. Anal. Prev. 43, 1521–1531.

Pahukula, J., Hernandez, S., Unnikrishnan, A., 2015. A Time of Day Analysis of Crashes Involving Large Trucks in Urban Areas. Accid. Anal. Prev. 75, 155–163.

Park, B.J., Lord, D., Hart, J.D., 2010. Bias Properties of Bayesian Statistics in Finite Mixture of Negative Binomial Regression Models in Crash Data Analysis. Accid. Anal. Prev. 42, 741–749.

Park, E.S., Lord, D., 2007. Multivariate Poisson-Lognormal Models for Jointly Modeling Crash Frequency by Severity. Transp. Res. Rec. J. Transp. Res. Board 2019, 1–6.

Peng, Y., Boyle, L.N., 2012. Commercial Driver Factors in Run-off-Road Crashes. Transp. Res. Rec. 2281, 128–132.

Poch, M., Mannering, F., 1996. Negative Binomial Analysis of Intersection-Accident Frequencies. J. Transp. Eng. 122, 105–113.

Qi, Y.G., Smith, B.L., Guo, J., 2007. Freeway Accident Likelihood Prediction Using a Panel Data Analysis Approach. J. Transp. Eng. 133, 149–156.

Russo, B.J., Savolainen, P.T., Schneider IV, W.H., Anastasopoulos, P.C., 2014. Comparison of Factors Affecting Injury Severity in Angle Collisions by Fault Status Using a Random Parameters Bivariate Ordered Probit Model. Anal. Methods Accid. Res. 2, 21–29.

Savolainen, P., Mannering, F., 2007. Probabilistic Models of Motorcyclists' Injury Severities in Single - and Multi - Vehicle Crashes. Accid. Anal. Prev. 39, 955–963.

Savolainen, P.T., Tarko, A.P., 2005. Safety Impacts at Intersections on Curved Segments. Transp. Res. Rec. pp 130–140.

Shaheed, M.S., Gkritza, K., 2014. A Latent Class Analysis of Single-Vehicle Motorcycle Crash Severity Outcomes. Anal. Methods Accid. Res. 2, 30–38.

Shankar, V., Albin, R., Milton, J.C., Mannering, F.L., 1998. Evaluating Median Crossover Likelihoods with Clustered Accident Counts: An Empirical Inquiry Using the Random Effects Negative Binomial Model. Highw. Saf. Model. Anal. Des. 44–48.

Shankar, V., Mannering, F., Barfield, W., 1995. Effect of Roadway Geometrics and Environmental Factors on Rural Freeway Accident Frequencies. Accid. Anal. Prev. 27, 371–389.

Shankar, V., Milton, J., Mannering, F., 1997. Modeling Accident Frequencies as Zero-Altered Probability Processes: An Empirical Inquiry. Accid. Anal. Prev. 29, 829–837.

Shea, M.S., Le, T.Q., Porter, R.J., 2015. Combined Crash Frequency–Crash Severity Evaluation of Geometric Design Decisions. Transp. Res. Rec. J. Transp. Res. Board 2521, 54–63.

Shen, J., 2009. Latent Class Model or Mixed Logit Model? A Comparison by Transport Mode Choice Data. Appl. Econ. 41, 2915–2924.

Tobin, J., 1958. Estimation of Relationships for Limited Dependent Variables. Econometrica 26, 24–36.

Train, K., 1999. Halton Sequences for Mixed Logit. Work. Pap. 1–18.

Ukkusuri, S., Hasan, S., Aziz, H.M.A., 2011. Random Parameter Model Used to Explain Effects of Built-Environment Characteristics on Pedestrian Crash Frequency. Transp. Res. Rec. J. Transp. Res. Board 2237, 98–106.

Veall, M.R., Zimmerman, K.F., 1996. Pseudo-R2 Measures for Some Common Limited Dependent Variable Models. J. Econ. Surv. 10.

Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2010. Statistical and Econometric Methods for Transportation Data Analysis, 2nd ed, Chapman & Hall/CRC. Chapman & Hall/CRC.

Xie, K., Wang, X., Ozbay, K., Yang, H., 2014. Crash Frequency Modeling for Signalized Intersections in a High-Density Urban Road Network. Anal. Methods Accid. Res. 2, 39–51.

Xie, Y., Zhang, Y., 2008. Crash Frequency Analysis with Generalized Additive Models. Transp. Res. Rec. J. Transp. Res. Board pp 39–45.

Xie, Y., Zhao, K., Huynh, N., 2012. Analysis of Driver Injury Severity in Rural Single-Vehicle Crashes. Accid. Anal. Prev. 47, 36–44.

Xiong, Y., Mannering, F.L., 2013. The Heterogeneous Effects of Guardian Supervision on Adolescent Driver-Injury Severities: A Finite-Mixture Random-Parameters Approach. Transp. Res. Part B Methodol. 49, 39–54.

Xu, P., Huang, H., 2015. Modeling Crash Spatial Heterogeneity: Random Parameters Versus Geographically Weighting. Accid. Anal. Prev. 75, 16–25.

Yang, C.C., 2004. Evaluating Latent Class Analysis Models in Qualitative Phenotype Identification. Comput. Stat. Data Anal. 50, 1090–1104.

Yasmin, S., Eluru, N., 2013. Evaluating Alternate Discrete Outcome Frameworks for Modeling Crash Injury Severity. Accid. Anal. Prev. 59, 506–521.

Yasmin, S., Eluru, N., Bhat, C.R., Tay, R., 2014. A Latent Segmentation Based Generalized Ordered Logit Model to Examine Factors Influencing Driver Injury Severity. Anal. Methods Accid. Res. 1, 23–38.

Ye, F., Lord, D., 2014. Comparing Three Commonly Used Crash Severity Models on Sample Size Requirements: Multinomial Logit, Ordered Probit and Mixed Logit Models. Anal. Methods Accid. Res. 1, 72–85.

Ye, X., Pendyala, R.M., Washington, S.P., Konduri, K., Oh, J., 2009. A Simultaneous Equations Model of Crash Frequency by Collision Type for Rural Intersections. Saf. Sci. 47, 443–452.

Yu, R., Xiong, Y., Abdel-Aty, M., 2015. A Correlated Random Parameter Approach to Investigate the Effects of Weather Conditions on Crash Risk for a Mountainous Freeway. Transp. Res. Part C Emerg. Technol. 50, 68–77.

Zhang, Y., Xie, Y., Li, L., 2012. Crash Frequency Analysis of Different Types of Urban Roadway Segments Using Generalized Additive Model. J. Safety Res. 43, 107–114.

Zhou, M., Sisiopiku, V., 1997. Relationship Between Volume-to-Capacity Ratios and Accident Rates. Transp. Res. Rec. 1581, 47–52.

Zhu, X., Srinivasan, S., 2011. A Comprehensive Analysis of Factors Influencing the Injury Severity of Large-Truck Crashes. Accid. Anal. Prev. 43, 49–57.