FHWA/IN/JTRP-2008/25

Final Report

SAFETY IMPACTS OF DESIGN EXCEPTIONS

Nataliya V. Malyshkina
Fred L. Mannering
Jose E. Thomaz

April 2009

INDOT Research Project Implementation Plan

Date: 4-30-09

Research Project Number: SPR-3220

Project Title: SAFETY IMPACTS OF DESIGN EXCEPTIONS

Principal Investigator (PI): Fred Mannering

Project Advisor (PA): Shuo Li Signature: ___________________________________

Principal Implementor (PIM): INDOT Planning Signature: ____________________________

INDOT Strategic Goal Impact Areas (select all that the implementation of this project will impact):

☐ *Mobility* ☒ *Safety* ☐ *Economic Development* ☐ *Customer Service*

☐ *Resource Management* ☐ *Training*

Summary of Implementation Plan: The statistical analysis provided in the project finds that previously granted Level One design exceptions have not had a statistically significant impact on the frequency or severity of accidents. However, the findings do suggest the caution should be exercised because the factors that determine the frequency of accidents are different between roadway segments with and without design exceptions. Although the sample of Level One design exceptions available for this study was small, some insight into potential critical roadway elements was possible (as discussed in the report). With regard to guiding future Level One design-exception decisions, using previous design exceptions as "precedents" would be the best way to proceed (broad policy statements are not yet possible given the limited number design-exceptions available for statistical analysis). Thus it is recommended that INDOT maintain a database of Level One design exceptions and that a case by case comparison with previously granted design exceptions guide future design-exception decisions. In addition, in terms of guiding decisions as to which design exceptions may be potentially problematic, the individual models estimated in this report do provide some guidance as does other recently published research using Indiana and national accident data. A case by case comparison with past research findings should also be considered when granting design exceptions

The PI is willing to assist should future analysis/assessment of design exceptions be considered.

Note: If more than one implementor recommended, please fill in the information on each implementor's implementation items:

Name of Implementor/User: *Signature:__________________________*

Responsible for Implementing:

Help or resources needed for implementation (e.g., help from PI, funding, equipment, etc.):

Name of Implementor/User: *Signature:__________________________*

Responsible for Implementing:

Help or resources needed for implementation (e.g., help from PI, funding, equipment, etc.):
Name of Implementor/User: *Signature:*_________________________ *Responsible for Implementing:* *Help or resources needed for implementation (e.g., help from PI, funding, equipment, etc.):*
Signatures of SAC members:____ Please send a copy of this form to the INDOT Research & Development Office and FHWA with the final report.

TECHNICAL *Summary*

INDOT Research

Technology Transfer and Project Implementation Information

TRB Subject Code: 51-6 Safety-Compatibility of Standards *April 2009*
Publication No. FHWA/IN/JTRP-2008/25, SPR-3220 *Final Report*

Safety Impacts of Design Exceptions

Introduction

Indiana Department of Transportation's (INDOT) highway design criteria are considered to be essential to ensure the safety of the motoring public. However, for a variety of reasons, situations arise where exceptions to standard-design criteria are requested and accepted after review. Although these decisions are carefully thought out, the safety impacts of various design-criteria exceptions are not well understood. The intent of this research is to rigorously study design exceptions in Indiana and to perform a careful statistical analysis of the impact that such exceptions have on roadway safety.

INDOT currently has a hierarchy of three levels of highway design criteria. Level One includes those highway design elements which have been judged to be the most critical indicators of highway safety and serviceability. There are 14 Level-One design criteria with minimum standards being met for: design speed; lane widths; shoulder widths; bridge width; bridge structural capacity; horizontal curvature; superelevation transition lengths, stopping-sight distance on horizontal and vertical curves; maximum grade; superelevation rate; minimum vertical clearance; accessibility for the handicapped; and bridge rail safety. Level-Two design criteria are judged to be important to safety and serviceability but are not considered as critical as Level One. Factors in Level Two criteria include: roadside safety elements; the obstruction-free zone; median and side slopes; access control; acceleration lane length; deceleration lane length; shoulder cross slope; auxiliary lane and shoulder widths; minimum grade for drainage; minimum level-of-service criteria; parking lane width; two-way left-turn width; and critical length of grade. Finally, Level Three design criteria include all other design criterion not listed in levels one and two. This research focuses on the impact of design exceptions within the most important Level-One category, which includes the most critical indicators of highway safety and serviceability.

To conduct the study, detailed information was gathered on 36 Level One design exceptions granted by INDOT between 1998 and 2003. Of these design exceptions, 32 were near bridges and 4 were along regular roadway intervals. To compare with similar roadways that were not granted design exceptions, 71 “control” roadway segments (those containing no design exceptions) were carefully chosen for their proximity and design similarities to those roadway segments that were granted design exceptions (63 control bridges and 8 control roadway intervals). Accident data were then meticulously matched (using location information) such that all police-reported accidents that occurred from January 1, 2003 to December 31, 2007 (a 5-year period). A total of 5,889 accidents occurred on these 107 roadway segments over the 5-year period (roughly 11 reported accidents per roadway segment per year).

Using these data, detailed statistical analyses of the frequency and severity of accidents were undertaken (using negative binomial regression and multinomial logit models) to determine if the design exceptions had any significant impact on the frequency or severity of accidents.

Findings

For the analysis of the severity of accidents, the injury level sustained by the most severely injured individual in the accident is used. Three options are considered: no injury (property damage only), injury and fatality. Detailed accident data are used to estimate multinomial logit models that estimate the probability of the three injury outcomes. The use of such a multivariate analysis is necessary to control for all factors that may affect injury severity (age of driver, gender of driver, number of occupants, etc.). A simple statistical comparison of the average accident severities on roadway segments with and without design exceptions would mask differences in driver and vehicle characteristics that may occur from one roadway segment to the next and potentially produce erroneous conclusions.

The multivariate analysis of accident severity (using both standard multinomial logit models and mixed multinomial logit models) found that the presence of a design exception had no statistically significant effect on the severity of accidents. In addition, a statistical test was conducted that showed that when separate severity models were estimated for roadway segments with and without design exceptions, no statistically significant difference was found. It is therefore concluded that previously granted design exceptions have not statistically affected safety in terms of accident severity.

For the analysis of the frequency of accidents, a negative binomial count model is used to estimate the number of accidents occurring over the five year period (2003-2007 inclusive) on individual roadway segments. It is found that the presence of a design exception had no statistically significant effect (this time on the likelihood of an accident) on the frequency of accidents. However, the statistical assessment showed that the process generating accident frequencies on segments with and without design exceptions was statistically different.

Implementation

The statistical analysis provided in the project finds that previously granted Level One design exceptions have not had a statistically significant impact on the frequency or severity of accidents. Although the sample of Level One design exceptions available for this study was small, some insight into potential critical roadway elements was possible (as discussed in the report). With regard to guiding future Level One design-exception decisions, using previous design exceptions as "precedents" would be the best way to proceed (broad policy statements are not yet possible given the limited number design-exceptions available for statistical analysis). Thus it is recommended that INDOT maintain a database of Level One design exceptions and that a case by case comparison with previously granted design exceptions guide future design-exception decisions. In addition, in terms of guiding decisions as to which design exceptions may be potentially problematic, the individual models estimated in this report do provide some guidance as does other recently published research using Indiana and national accident data. A case by case comparison with past research findings should also be considered when granting design exceptions.

Contacts

For more information:

Prof. Fred L. Mannering
Principal Investigator
School of Civil Engineering
Purdue University
West Lafayette IN 47907
Phone: (765) 496-7913
Fax: (765) 496-7996
E-mail: flm@purdue.edu

Indiana Department of Transportation
Division of Research
1205 Montgomery Street
P.O. Box 2279
West Lafayette, IN 47906
Phone: (765) 463-1521
Fax: (765) 497-1665

Purdue University
Joint Transportation Research Program
School of Civil Engineering
West Lafayette, IN 47907-1284
Phone: (765) 494-9310
Fax: (765) 496-7996
E-mail: jtrp@ecn.purdue.edu
http://www.purdue.edu/jtrp

Final Report

FHWA/IN/JTRP-2008/25

SAFETY IMPACTS OF DESIGN IMPACTS

by

Nataliya V. Malyshkina
Graduate Research Assistant

Fred L. Mannering
Professor

and

Jose E. Thomaz
Data Analyst
Center for Roadway Safety

School of Civil Engineering
Purdue University

Joint Transportation Research Program
Project No: C-36-56DDD
File No: 8-5-56

Prepared in Cooperation with the
Indiana Department of Transportation and
The U.S. Department of Transportation
Federal Highway Administration

Purdue University
West Lafayette, Indiana
April 30, 2009

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No. FHWA/IN/JTRP-2008/25	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Safety Impacts of Design Exceptions		**5. Report Date** April 2009
		6. Performing Organization Code
7. Author(s) Nataliya V. Malyshkina, Fred L. Mannering, and Jose E.Thomaz		**8. Performing Organization Report No.** FHWA/IN/JTRP-2008/25
9. Performing Organization Name and Address Joint Transportation Research Program 550 Stadium Mall Drive Purdue University West Lafayette, IN 47907-2051		**10. Work Unit No.**
		11. Contract or Grant No. SPR-3220
12. Sponsoring Agency Name and Address Indiana Department of Transportation State Office Building 100 North Senate Avenue Indianapolis, IN 46204		**13. Type of Report and Period Covered** Final Report
		14. Sponsoring Agency Code

15. Supplementary Notes

Prepared in cooperation with the Indiana Department of Transportation and Federal Highway Administration.

16. Abstract

Compliance to the Indiana Department of Transportation's (INDOT) highway design criteria is considered essential to ensure the roadway safety. However, for a variety of reasons, situations arise where exceptions to standard-design criteria are requested and accepted after review. This research explores the impact that design exceptions have on the accident severity and accident frequency in Indiana. Data on accidents at 36 roadway sites with design exceptions and 71 without design exceptions are used in this research, and appropriate statistical models are estimated for the severity and frequency of these accidents. The results of the detailed statistical modeling show that presence of design exceptions, approved by INDOT, do not have a statistically significant adverse effect on the frequency or severity of accidents. While the data are too limited to investigate the effect of specific design exceptions (the number of Level One design exceptions granted is a modest number), the research herein shows that INDOT procedures for granting design exceptions have been sufficiently strict to avoid adverse safety consequences and that current practices should be continued. To guide future Level One design exceptions, the detailed statistical findings of this research effort suggest that using previous design exceptions as "precedents" would be the best way to proceed. To this end, it is recommended that INDOT maintain a database of Level One design exceptions.

17. Key Words Design exceptions, accident severity, accident frequency, highway safety	**18. Distribution Statement** No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161

19. Security Classif. (of this report)	**20. Security Classif. (of this page)**	**21. No. of Pages**	**22. Price**
Unclassified	Unclassified	65	

Form DOT F 1700.7 (8-69)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Compliance to the Indiana Department of Transportation's (INDOT) highway design criteria is considered essential to ensure the roadway safety. However, for a variety of reasons, situations arise where exceptions to standard-design criteria are requested and accepted after review. This research explores the impact that design exceptions have on the accident severity and accident frequency in Indiana. Data on accidents at 36 roadway sites with design exceptions and 71 without design exceptions are used in this research, and appropriate statistical models are estimated for the severity and frequency of these accidents. The results of the detailed statistical modeling show that presence of design exceptions, approved by INDOT, do not have a statistically significant adverse effect on the frequency or severity of accidents. While the data are too limited to investigate the effect of specific design exceptions (the number of Level One design exceptions granted is a modest number), the research herein shows that INDOT procedures for granting design exceptions have been sufficiently strict to avoid adverse safety consequences and that current practices should be continued. To guide future Level One design exceptions, the detailed statistical findings of this research effort suggest that using previous design exceptions as “precedents” would be the best way to proceed. To this end, it is recommended that INDOT maintain a database of Level One design exceptions.

CHAPTER 1. INTRODUCTION

Indiana Department of Transportation's (INDOT) highway design criteria are considered to be essential to ensure the safety of the motoring public. However, for a variety of reasons, situations arise where exceptions to standard-design criteria are requested and accepted after review. Common reasons for considering design exceptions include: impact to the natural environment; social or right-of-way impacts; preservation of historic or cultural resources; sensitivity to context or accommodating community values; and construction or right-of-way costs (Federal Highway Administration, 1999; American Association of State Highway and Transportation Officials, 2004). Because of the potential of serious safety consequences and tort liability, the process for granting design exceptions is very closely monitored by state and federal highway agencies, although practices and standards for granting design exceptions can vary significantly from state to state (National Cooperative Research Program, 2003).

Although these design-exception decisions are carefully thought out, the safety impacts of various design-criteria exceptions are not well understood. Over the years, there have been numerous research efforts that have attempted to evaluate the safety impacts of design exceptions. For example, Agent et al. (2002) studied the effect of design exceptions on crash rates in the state of Kentucky. They found that the most common design exception was for a design speed lower than the posted speed limit followed by a lower than standard sight distance, curve radius or shoulder width. With an average of about 39 design exceptions per year in Kentucky, they concluded (based on observations of crash rates) that design exceptions did not result in projects with high crash rates relative to average statewide rates. Unfortunately, in this and many other studies, the amount of data available (which is limited because of the small number of design exceptions granted per year and the highly detailed roadway and accident information required) has made it difficult to develop statistically defensible models to assess the safety impacts of design exceptions in a multivariate framework.

Given the scarcity of design-exception data and associated accident data, some have attempted to infer the effects of design exceptions from statistical models that have been estimated on a simple cross section of roadway segments in an effort to uncover the impact of specific design features (shoulder width, median presence, etc.) on the frequency of accidents and the severity of accidents in terms of resulting injuries. Common statistical approaches to determine the relationship between roadway characteristics and accident frequencies include: Poisson and negative binomial models (Jones et al., 1991; Shankar et al., 1995; Hadi et al., 1995; Poch and Mannering, 1996; Milton and Mannering, 1998; Abdel–Aty and Radwan, 2000; Savolainen and Tarko, 2005; Lord, 2006; Wang and Abdel-Aty, 2008; Lord and Park, 2008); zero–inflated negative binomial models (Shankar et al., 1997; Carson and Mannering, 2001; Lee and Mannering, 2002); negative binomial with random effects models (Shankar et al., 1998); Conway–Maxwell–Poisson generalized linear models (Lord et al., 2008); negative binomial with random parameters (Anastasopoulos and Mannering, 2009) and dual-state negative binomial Markov switching models (Malyshkina et. al, 2009a). For the severity of accidents, quantifying the effects of roadway characteristics on vehicle-occupant injuries have been undertaken using a wide variety of models including multinomial logit models, dual-state multinomial logit models, nested logit models, mixed logit models and ordered probit models (O'Donnell and Connor, 1996; Shankar and Mannering, 1996; Shankar et al., 1996; Duncan et al., 1998; Chang and Mannering, 1999; Carson and Mannering, 2001; Khattak, 2001; Khattak et al., 2002; Kockelman and

Kweon, 2002; Lee and Mannering, 2002; Abdel-Aty, 2003; Kweon and Kockelman, 2003; Ulfarsson and Mannering, 2004; Yamamoto and Shankar, 2004; Khorashadi et al., 2005; Lee and Abdel-Aty, 2005; Eluru and Bhat, 2007;Savolainen and Mannering, 2007; Milton et al., 2008; Malyshkina and Mannering, 2009).

However, attempting to infer the impact of design exceptions from general roadway-segment data is potentially problematic because roadway segments that are granted design exceptions are likely to be a non-random sample of the roadway-segment population (segments may have common special features that make them more likely to require a design exception). If this is the case, roadway segments prone to design exceptions will share unobserved effects and the relationship of their characteristics to the frequency and severity of accidents may be significantly different than the relationship on the non-design-exception roadway-segment sample. One way of resolving this problem is to gather a sample of sufficient size that includes roadway segments with design exceptions and similar roadway segments without design exceptions (not a random sample of roadway segments without design exceptions), and to use random parameter models to account for possible unobserved heterogeneity. The intent of this current study is to use such a sample and modeling approach to closely assess the effect of design exceptions on the frequency and severity of accidents.

INDOT currently has a hierarchy of three levels of highway design criteria. Level One includes those highway design elements which have been judged to be the most critical indicators of highway safety and serviceability. There are 14 Level-One design criteria with minimum standards being met for: design speed; lane widths; shoulder widths; bridge width; bridge structural capacity; horizontal curvature; superelevation transition lengths, stopping-sight distance on horizontal and vertical curves; maximum grade; superelevation rate; minimum vertical clearance; accessibility for the handicapped; and bridge rail safety. Level-Two design criteria are judged to be important to safety and serviceability but are not considered as critical as Level One. Factors in Level Two criteria include: roadside safety elements; the obstruction-free zone; median and side slopes; access control; acceleration lane length; deceleration lane length; shoulder cross slope; auxiliary lane and shoulder widths; minimum grade for drainage; minimum level-of-service criteria; parking lane width; two-way left-turn width; and critical length of grade. Finally, Level Three design criteria include all other design criterion not listed in levels one and two.

In the current study we focus on the impact of (Level-One) design exceptions. Among the questions that we answer is whether or not design exceptions have significantly affected the frequency and severity of accidents. We consider data on individual accidents and use the methodologies of statistical modeling within the framework of count data and discrete outcome models. In this study we use the following two statistical modeling approaches:

1. In the first approach we will focus on severity of accidents. The idea is to study a relationship between the presence of design exceptions and the probability of various accident severity levels (determined by the injury level sustained by the most critically injured individual in the accident). This will be done by estimation of standard and mixed multinomial statistical models for accident severity.
2. In the second approach we will undertake a study of accident frequency study. We will estimate standard and mixed negative binomial statistical models for the five-year accident frequency (which is the cumulative number of accidents occurred over the considered five-year period). Then we will test whether the presence of design exceptions has any effect on accident frequency.

To reveal the effect of design exceptions on safety, while modeling accident severity and frequency, we will control for other possible confounding effects, such as road characteristics, weather conditions, driver characteristics, and so on. The use of the above two accident modeling approaches will provide important new insights and sufficient statistical evidence on the effect of design exceptions on roadway safety.

This report is organized as follows. In the next chapter we will briefly describe the methodology of statistical modeling used in our study. Detailed descriptions and simple descriptive statistics of the accident data used are given in CHAPTER 3. In CHAPTER 4 we consider influence of design exceptions on accident severity. In CHAPTER 5 we consider influence of design exceptions on accident frequency. Finally, in CHAPTER 6 we summarize and discuss the main results of our study.

CHAPTER 2. METHODOLOGY OF STATISTICAL MODELING

2.1. Multinomial logit models of accident severity

First, let us consider accident severity, which is a non-quantitative discrete outcome of traffic accidents. The most widely used statistical models for non-count data that is composed of discrete outcomes are the multinomial logit model and the ordered probit model. However, there are two potential problems with applying ordered probability models to accident severity outcomes (Savolainen and Mannering 2007). The first is related to the fact that non-injury accidents are likely to be under-reported in accident data because they are less likely to be reported to authorities. The presence of under-reporting in an ordered probability model will result in biased and inconsistent model coefficient estimates. In contrast, the coefficient estimates of an unordered multinomial logit probability model are consistent except for the constant terms (Washington et. al. 2003, page 279). The second problem is related to undesirable restrictions that ordered probability models place on influences of the explanatory variables (Washington et. al. 2003, page 294). As a result, in our research study we use and estimate multinomial logit models for accident severity.

The simple standard multinomial logit model can be introduced as follows. Let there be N available data observations and I possible discrete outcomes in each observation. Then in the multinomial logit model the probability $P_n^{(i)}$ of the i^{th} outcome in the n^{th} observation is specified by equation (Washington et al., 2003, page 263)

$$P_n^{(i)} = \frac{\exp(\boldsymbol{\beta}_\mathbf{i}' \mathbf{X}_\mathbf{in})}{\sum_{j=1}^{I} \exp(\boldsymbol{\beta}_\mathbf{j}' \mathbf{X}_\mathbf{jn})}, \qquad i = 1,2,3,...,I, \quad n = 1,2,3,...,N. \qquad \text{Eq. 2.1}$$

Here $\mathbf{X}_\mathbf{in}$ is the vector of explanatory variables for the n^{th} observation and $\boldsymbol{\beta}_\mathbf{i}$ is the vector of model coefficients to be estimated ($\boldsymbol{\beta}_\mathbf{i}'$ is the transpose of $\boldsymbol{\beta}_\mathbf{i}$). We use a conventional assumption that the first component of vector $\mathbf{X}_\mathbf{in}$ is equal to unity, and therefore, the first component of vector $\boldsymbol{\beta}_\mathbf{i}$ is the intercept in linear product $\boldsymbol{\beta}_\mathbf{i}'\mathbf{X}_\mathbf{in}$. Note that $P_n^{(i)}$, given by Equation (2.1), is a valid probability set for I discrete outcomes because the necessary and sufficient conditions $P_n^{(i)} \geq 0$ and $\sum_{i=1}^{I} P_n^{(i)} = 1$ are obviously satisfied[1].

We can multiply the numerator and denominator of the fraction in Equation (2.1) by an arbitrary number without any change of the probabilities. As a result, without any loss of generality we can set one of the intercepts to zero. We choose the first component of vector $\boldsymbol{\beta}_I$ to be zero in this case. Moreover, if the vector of explanatory variables does not depend on discrete outcomes, i.e. if $\mathbf{X}_\mathbf{in} \equiv \mathbf{X}_\mathbf{n}$, then without any loss of generality we can set one of vectors of model coefficients to zero. We choose vector $\boldsymbol{\beta}_I$ to be zero in this case.

Because accidents are independent events, the likelihood function for the set of probabilities given in Equation (2.1) is

$$L(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, ...\boldsymbol{\beta}_I) = \prod_{n=1}^{N} \prod_{i=1}^{I} [P_n^{(i)}]^{\delta_{in}}, \qquad \text{Eq. 2.2}$$

where δ_{in} is defined to be equal to unity if the i^{th} discrete outcome is observed in the n^{th} observation and to zero otherwise.

---

[1] Equation (2.1) can formally be derived by using a linear specification $U_{in} \equiv \boldsymbol{\beta}_\mathbf{i}'\mathbf{X}_\mathbf{in} + \tilde{\varepsilon}_{in}$, by defining $P_n^{(i)} = \mathrm{Prob}\left\{ U_{in} \geq \max_{\forall j \neq i}(U_{jn}) \right\}$ and by choosing the Gumbel (Type I) extreme value distribution for the i.i.d. random error terms $\tilde{\varepsilon}_{in}$. For details see Washington et al., 2003.

Now we assume that the explanatory variables vector is independent of the discrete outcomes, $\mathbf{X_{in}} \equiv \mathbf{X_n}$, and consider two simple special cases of the multinomial logit model. If there are three possible discrete outcomes, $I = 3$ and $i = 1,2,3$, then in this case Equation (2.1) simplifies to

$$P_n^{(1)} = \frac{\exp(\boldsymbol{\beta}_1'\mathbf{X_n})}{\exp(\boldsymbol{\beta}_1'\mathbf{X_n}) + \exp(\boldsymbol{\beta}_2'\mathbf{X_n}) + 1},$$
$$P_n^{(2)} = \frac{\exp(\boldsymbol{\beta}_2'\mathbf{X_n})}{\exp(\boldsymbol{\beta}_1'\mathbf{X_n}) + \exp(\boldsymbol{\beta}_2'\mathbf{X_n}) + 1}, \qquad \text{Eq. 2.3}$$
$$P_n^{(3)} = \frac{1}{\exp(\boldsymbol{\beta}_1'\mathbf{X_n}) + \exp(\boldsymbol{\beta}_2'\mathbf{X_n}) + 1},$$

where there are two coefficient vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ to be estimated. We will use these special-case logit models in the next two chapters.

It is customary to use the maximum likelihood method to estimate unknown vectors of coefficients $\boldsymbol{\beta}_\mathbf{i}$ in the logit models given by Equations (2.1) and (2.3). Namely, one finds such values of the unknown coefficients that the likelihood function (and correspondingly the log-likelihood function) given by Equation (2.2) reaches its global maximum. In the present study we use econometric software package LIMDEP/NLOGIT for all model estimations by means of the maximum likelihood method[2]. We also use MATLAB software package for initial processing of data.

Next, we describe how the magnitude of the influence of specific explanatory variables on the discrete outcome probabilities can be measured. This is done by elasticity computations (Washington et al., 2003, page 271). Elasticities $E_{X_{jn,k}}^{P_n^{(i)}}$ are computed from the partial derivatives of the outcome probabilities for the n^{th} observation as

[2] LIMDEP/NLOGIT can be found at http://www.limdep.com, we use NLOGIT 4.0.

$$E^{P_n^{(i)}}_{X_{jn,k}} = \frac{\partial P_n^{(i)}}{\partial X_{jn,k}} \cdot \frac{X_{jn,k}}{P_n^{(i)}}, \quad i,j = 1,...,I, \quad n = 1,...,N, \quad k = 1,...,K. \qquad \text{Eq. 2.4}$$

Here $P_n^{(i)}$ is the probability of outcome i given by Equation (2.1), $X_{jn,k}$ is the k^{th} component of the vector of explanatory variables $\mathbf{X_{jn}}$ that enters the formula for the probability of outcome j, and K is the length of this vector. If $j = i$, then the elasticity given by Equation (2.4) is called *direct* elasticity, otherwise, if $j \neq i$, then the elasticity is called *cross* elasticity. The direct elasticity of the outcome probability $P_n^{(i)}$ with respect to variable $X_{in,k}$ measures the percent change in $P_n^{(i)}$ that results from an infinitesimal percentage change in $X_{in,k}$. Note that $X_{in,k}$ directly enters the numerator of the formula for $P_n^{(i)}$, as given by Equation (2.1). The cross elasticity of $P_n^{(i)}$ with respect to variable $X_{jn,k}$ measures the percent change in $P_n^{(i)}$ that results from an infinitesimal percentage change in $X_{jn,k}$. Note that $X_{jn,k}$ enters the numerator of the formula for the probability $P_n^{(j)}$ of the outcome j, which is different from outcome i. Thus, cross elasticities measure indirect effects that arise from the fact that the outcome probabilities must sum to unity, $\sum_{i=1}^{I} P_n^{(i)} = 1$. If the absolute value of the computed elasticity $E^{P_n^{(i)}}_{X_{jn,k}}$ of explanatory variable $X_{jn,k}$ is less than unity, then this variable is said to be inelastic, and the resulting percentage change in the outcome probability $P_n^{(i)}$ will be less (in its absolute value) than a percentage change in the variable. Otherwise, the variable is said to be elastic. It is customary to report averaged elasticities, which are the elasticities averaged over all observations (i.e. averaged over $n = 1,2,3,...,N$). Let us consider the case of three possible discrete outcomes, given by Equation (2.3). In this case $i = 1,2,3$ and we have the following formulas for the averaged direct and cross elasticities (Washington et al., 2003, pages 271-272):

$$\overline{E}_{1;X_k}^{(1)} = \left\langle E_{X_{1n,k}}^{P_n^{(1)}} \right\rangle_n = \left\langle \left[1 - P_n^{(1)}\right] \cdot \beta_{1,k} X_{n,k} \right\rangle_n$$

$$\overline{E}_{2;X_k}^{(2)} = \left\langle E_{X_{2n,k}}^{P_n^{(2)}} \right\rangle_n = \left\langle \left[1 - P_n^{(2)}\right] \cdot \beta_{2,k} X_{n,k} \right\rangle_n$$

averaged direct elasticities;

Eq. 2.5

$$\overline{E}_{1;X_k}^{(2)} = \overline{E}_{1;X_k}^{(3)} = \left\langle E_{X_{1n,k}}^{P_n^{(2,3)}} \right\rangle_n = -\left\langle P_n^{(1)} \cdot \beta_{1,k} X_{n,k} \right\rangle_n$$

$$\overline{E}_{2;X_k}^{(1)} = \overline{E}_{2;X_k}^{(3)} = \left\langle E_{X_{2n,k}}^{P_n^{(1,3)}} \right\rangle_n = -\left\langle P_n^{(2)} \cdot \beta_{2,k} X_{n,k} \right\rangle_n$$

averaged cross elasticities.

Here brackets $\langle ... \rangle_n$ means averaging over all observations $n = 1,2,3,...,N$.

The elasticity formulas given above are applicable only when explanatory variable $X_{jn,k}$ used in the outcome probability model is continuous. In the case when $X_{jn,k}$ takes on discrete values, the elasticities given by Equation (2.4) cannot be calculated, and they are replaced by *pseudo-elasticities* (for example, see Washington et al., 2003, page 272). The later are given by the following equation, which is an obvious discrete counterpart of Equation (2.4),

$$E_{X_{jn,k}}^{P_n^{(i)}} = \frac{\Delta P_n^{(i)}}{\Delta X_{jn,k}} \cdot \frac{X_{jn,k}}{P_n^{(i)}}, \quad i, j = 1,...,I, \quad n = 1,...,N, \quad k = 1,...,K. \qquad \text{Eq. 2.6}$$

Here $\Delta P_n^{(i)}$ denotes the resulting discrete change in the probability of outcome i due to discrete change $\Delta X_{jn,k}$ in variable $X_{jn,k}$.

In addition to simple multinomial logit models, we consider mixed multinomial logit models of accident severity. In a mixed multinomial logit model, the probability of the i^{th} outcome in the n^{th} observation is (Washington et. al. 2003, page 287)

$$\tilde{P}_n^{(i)} = \int P_n^{(i)} \, q(\boldsymbol{\beta}_\mathbf{i} \,|\, \boldsymbol{\varphi}_\mathbf{i}) \; d\boldsymbol{\beta}_\mathbf{i} \,, \qquad i = 1,2,3,...,I \,, \quad n = 1,2,3,...,N \,. \qquad \text{Eq. 2.7}$$

The right-hand-side of Equation (2.7) is a mixture of the standard multinomial probabilities $P_n^{(i)}$, given by Equation (2.1). Probability distribution $q(\boldsymbol{\beta}_\mathbf{i} \,|\, \boldsymbol{\varphi}_\mathbf{i})$ is the distribution of the multinomial logit parameters $\boldsymbol{\beta}_\mathbf{i}$, given fixed parameters $\boldsymbol{\varphi}_\mathbf{i}$.

The likelihood equation (2.2) and the elasticity equations (2.4) and (2.6) hold for mixed multinomial logit models with $P_n^{(i)}$ replaced by $\widetilde{P}_n^{(i)}$.

2.2. Negative binomial models of accident frequency

Now, let us consider accident frequency, which is a quantitative count data that is the number of accidents occurred. The most widely used statistical models for count data are the Poisson and negative binomial models. Poisson model is a particular case of negative binomial model (a negative binomial model reduces to a Poisson model when the overdispersion parameter is zero). As a result, without loss of generality, we consider only negative binomial models in this study.

The simple standard negative binomial model of five-year accident frequency can be introduced as follows. The probability of A_n accidents occurred on road segment *n* during the considered five-year time period (Washington et al., 2003, page 248)

$$P_n^{(A)} = \frac{\Gamma(A_n + 1/\alpha)}{\Gamma(1/\alpha) A_n!} \left(\frac{1}{1+\alpha\lambda_n} \right)^{1/\alpha} \left(\frac{\alpha\lambda_n}{1+\alpha\lambda_n} \right)^{A_n}, \qquad \text{Eq. 2.8}$$
$$\lambda_n = \exp(\boldsymbol{\beta}'\mathbf{X}_n), \qquad n = 1,2,...,N .$$

Here $\mathbf{X}_n$ is the vector of explanatory variables for the n th roadway segment, Γ is the gamma-function, prime means transpose ($\boldsymbol{\beta}'$ is the transpose of $\boldsymbol{\beta}$), and N is the number of roadway segments. Vector $\boldsymbol{\beta}$ and the over-dispersion parameter α are unknown estimable coefficients of the negative binomial model. Scalar λ_n is the mean five-year accident rate on roadway segment n.

Accident events are assumed to be independent. Therefore, the full likelihood function is

$$L(\boldsymbol{\beta},\alpha)=\prod_{n=1}^{N}P_n^{(A)}\ . \qquad \text{Eq. 2.9}$$

As in the case of accident severity, for accident frequency models we use the maximum likelihood estimation (MLE) with the help of LIMDEP/NLOGIT econometric software package.

With regard to the magnitude of the influence of specific explanatory variables on the expected accident frequency, instead of the elasticities used for the severity analysis we use marginal effects which are easier to interpret for count-data models. The marginal effect is computed as (see Washington et al., 2003),

$$\frac{\partial E(A_n\,|\,\mathbf{X}_n)}{\partial X_{n,k}}=\frac{\partial \lambda_n}{\partial X_{n,k}}=\frac{\partial}{\partial X_{n,k}}\left[exp(\boldsymbol{\beta}'\mathbf{X}_n)\right]=\lambda_n\boldsymbol{\beta}\ . \qquad \text{Eq. 2.10}$$

where $X_{n,k}$ is the k^{th} component of the vector of explanatory variables $\mathbf{X}_n$. The marginal effect gives the effect that a one unit change in the explanatory variable $X_{n,k}$ has on the mean accident frequency λ_n. As was the case with elasticities, because each observation generates its own marginal effect, the average across all observation will be reported in the forthcoming empirical analysis.

In this study, we also used mixed negative binomial models, which are defined similarly to the mixed multinomial logit models. In a mixed negative binomial model, the probability of A_n accidents occurred on road segment n is

$$\widetilde{P}_n^{(A)}=\int P_n^{(A)}\ q(\boldsymbol{\beta},\alpha\,|\,\boldsymbol{\varphi})\ d\boldsymbol{\beta}\ d\alpha\ , \qquad n=1,2,...,N\ . \qquad \text{Eq. 2.10}$$

The right-hand-side of Equation (2.10) is a mixture of the standard negative binomial probabilities $P_n^{(A)}$, given by Equation (2.8). Probability distribution $q(\boldsymbol{\beta},\alpha\,|\,\boldsymbol{\varphi})$ is the distribution of the negative binomial parameters $\boldsymbol{\beta}$ and α, given fixed parameters $\boldsymbol{\varphi}$. The likelihood equation (2.9) holds for mixed multinomial negative binomial models with $P_n^{(A)}$ replaced by $\widetilde{P}_n^{(A)}$.

2.3. Choice of explanatory variables for model estimation

To uncover the direct influence of design exceptions on accident severity and frequency, we need to control for other explanatory variables (factors) that might also affect severity/frequency. Examples of these other variables are weather conditions, accident time and date, vehicle and driver characteristics, roadway segment characteristics and so on. All explanatory variables can be divided into two distinct types. First, there are indicator (dummy) variables that are equal to unity if some particular conditions are satisfied, and are equal to zero otherwise. Examples of indicator variables are driver's gender indicator, weekend indicator, precipitation indicator and roadway median presence indicator. Second, there are quantitative variables that take on meaningful quantitative values, such as driver's age, speed limit, roadway segment length and AADT. In addition, one can easily define derivative indicator variables that are obtained from quantitative variables. For example, one can define a "young driver" indicator as being equal to unity if the driver's age is below 25. When estimating models, we frequently define and use the most useful (as judged by the model likelihood function) new derivative indicator variables that are based on quantitative variables.

We check statistical significance of the explanatory variables in all logit models by using 5% significance level for the two-tailed t-test of a large data sample. In other words, coefficients with t-ratios between -1.96 and +1.96 are considered to be statistically insignificant and others outside of these bounds are statistically significant. Note that the explanatory variables can be mutually dependent (e.g. a quantitative variable and its derivative indicator variable are strongly mutually dependent).

Statistical models are estimated by maximizing the model's log-likelihood function. However, one cannot rely on the log-likelihood maximization alone in order to choose the optimal number of explanatory variables to be included in the statistical model. The reason is that the log-likelihood (LL) function is always maximized when all available explanatory variables are included into the model. This is because a removal of any explanatory variable is equivalent to restricting its value to zero, which always either decreases the maximum of LL or leaves it the same. As a result, in the present study we use the Akaike Information Criterion (AIC), minimization of which ensures an optimal choice of explanatory variables in a model (Tsay, 2002, page 37; Washington et al., 2003, page 212; Wikipedia). The main idea behind the AIC is to examine the complexity of a model together with goodness of its fit to the data sample, and to find a balance between the two. A model with too few explanatory variables will provide a poor fit to the data sample. A model with too many variables will provide a very good fit, but will lack necessary robustness and will perform poorly in out-of-the-sample data. The preferred model with the optimal number of explanatory variables is the model with the lowest AIC value, which is given by equation

$$AIC = -2LL + 2K \, , \qquad \text{Eq. 2.11}$$

where LL is the log-likelihood value of a model, and K is the number of estimable coefficients in the model (one coefficient for each explanatory variable, including the intercepts).

In our research we estimate all logit models by using one of the two procedures A and B shown in Figure 2.1. Procedure A is as follows:

I. We start with all explanatory variables initially included into a logit model. Note that, when estimating a model, we have to exclude observations that are missing for any of the included variables. Next, we obtain the final model by using three steps of model estimation. The first step is

We remove the least statistically significant explanatory variables (as judged by their t-ratios) one by one if *both* of the following two conditions are satisfied: the removal of a variable decreases the AIC value *and* the removed variable is statistically insignificant (under the 5% confidence level)[3]. Note that while using the Akaike information criterion, we always keep the number of data sample observations constant in order to calculate the changes of the AIC value correctly. Each time when we have removed several (usually four) least significant explanatory variables from a model, we include some of the previously excluded observations back into the data sample because now the model includes fewer variables with missing observations. We keep removing insignificant explanatory variables one by one, periodically including previously excluded observations back into the data sample, until we cannot remove any additional variable under the two conditions listed above.

II. We start with all explanatory variables initially included into a logit model. Note that, when estimating a model, we have to exclude observations that are missing for any of the included variables. Next, we obtain the final model by using three steps of model estimation. The first step is

1. We remove the least statistically significant explanatory variables (as judged by their t-ratios) one by one if *both* of the following two conditions are satisfied: the removal of a variable decreases the

[3] If the asymptotic normality of maximum likelihood estimates holds, then the AIC value does not change with removal (addition) of a variable whose coefficient has 15.73% p-value for the two-tailed test (15.73% p-value corresponds to $\pm\sqrt{2}$ t-ratio for a normal variate). In this case the 5% confidence level test of the variable is redundant, and the AIC test alone can be used for removal and addition of variables in model estimation steps 1 and 2. Nevertheless, we use both tests to make our estimation procedures more robust in case the normality of maximum likelihood estimates does not hold.

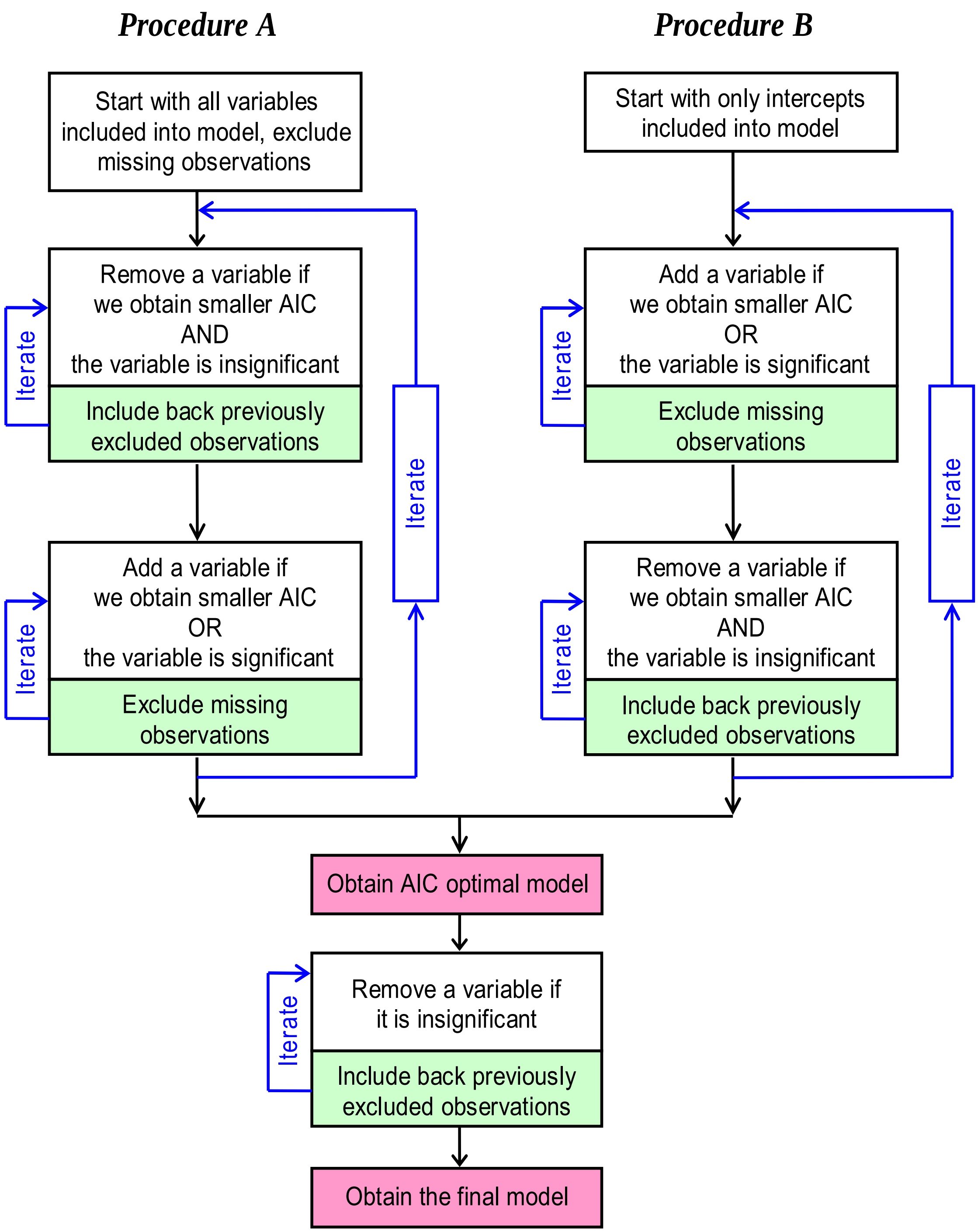

Figure 2.1 Model estimation procedures

AIC value *and* the removed variable is statistically insignificant (under the 5% confidence level)[4]. Note that while using the Akaike information criterion, we always keep the number of data sample observations constant in order to calculate the changes of the AIC value correctly. Each time when we have removed several (usually four) least significant explanatory variables from a model, we include some of the previously excluded observations back into the data sample because now the model includes fewer variables with missing observations. We keep removing insignificant explanatory variables one by one, periodically including previously excluded observations back into the data sample, until we cannot remove any additional variable under the two conditions listed above.

After we removed all variables that we could, we need to check if any of the removed variables can be added back into the model. This is because variables are mutually dependent and “interact” in the model. Therefore, we proceed to the second step of model estimation:

2. We add explanatory variables one by one if *at least one* of the following two conditions is satisfied: *either* the addition of a variable decreases the AIC value *or* the added variable is significant[5]. As usual, the AIC values are compared under the condition that the number of observations is kept constant. As the number of the explanatory variables included into the model

---

[4] If the asymptotic normality of maximum likelihood estimates holds, then the AIC value does not change with removal (addition) of a variable whose coefficient has 15.73% p-value for the two-tailed test (15.73% p-value corresponds to $\pm\sqrt{2}$ t-ratio for a normal variate). In this case the 5% confidence level test of the variable is redundant, and the AIC test alone can be used for removal and addition of variables in model estimation steps 1 and 2. Nevertheless, we use both tests to make our estimation procedures more robust in case the normality of maximum likelihood estimates does not hold.

[5] We first search for and add AIC decreasing variables, and afterwards we add significant variables if there are any.

grows, the data sample size shrinks because of a larger number of missing observations associated with the included variables. We add explanatory variables one by one until no any additional variable can be added to the model.

Next we return back to the first estimation step given above and remove variables that can be removed. We iterate between steps 1 and 2 until we can neither remove nor add any more variables. At this point we arrive at the model that we call the "AIC optimal model" (refer to Figure 2.1). Next, we proceed to the third and final step of model estimation:

3. To make our final results more robust, we drop from the AIC optimal model all remaining statistically insignificant variables (judged by the 5% significance level for the two-tailed t-test). As a result, we obtain the final model, which is our best model (according to the estimation procedures chosen by us).

Now we describe procedure B:

I. In this procedure we start with only intercepts (constant terms) initially included into a logit model (refer to Figure 2.1). Next, we proceed in a way very similar to that used in procedure A. We run step 2 of model estimation and add explanatory variables into the model. Then, we iterate between steps 1 and 2 until we can neither remove nor add any more variables, at which point we arrive at the AIC optimal model. Finally, we run step 3 of model estimation and obtain the best final model.

By default we always use procedure A for model estimation, and only if we cannot use it (usually when the available data sample is too small for the initial model estimation with all explanatory variables included), then we resort to procedure B.

2.4. Likelihood ratio test

In the forthcoming chapters we will need to compare several estimated models in order to infer if there are statistically significant differences among these models. As a result, here we would like to demonstrate how model comparisons are done by using a likelihood ratio test. Assume that we have divided a data sample into different data bins. The likelihood ratio test uses the model estimated for the whole data sample and the models separately estimated for each data bin. The test statistic is (Washington et al., 2003, page 244)

$$X^2 = -2\left[LL(\boldsymbol{\beta}) - \sum_{m=1}^{M} LL(\boldsymbol{\beta}_{\mathbf{m}})\right] \sim \chi^2_{\mathrm{df}=(M-1)\times K}, \qquad \text{Eq. 2.12}$$

where $LL(\boldsymbol{\beta})$ is the log-likelihood of the model estimated for the whole data sample and $\boldsymbol{\beta}$ is the vector of coefficients estimated for this model; $LL(\boldsymbol{\beta}_{\mathbf{m}})$ is the log-likelihood of the model estimated for observations in the m^{th} data bin and $\boldsymbol{\beta}_{\mathbf{m}}$ is the vector of coefficients estimated for this model ($m = 1,2,3,...,M$); M is the number of the data bins; K is the number of coefficients estimated for each model (i.e. K is the length of vectors $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_{\mathbf{m}}$)[6]; $\chi^2_{\mathrm{df}=(M-1)\times K}$ is the chi-squared distribution with $(M-1)\times K$ degrees of freedom (df). The zero-hypothesis for the test statistic given by Equation (2.12) is that the model estimated for the whole data sample and the combination of the M models separately estimated for the data bins, are statistically the same. In other words, for a chosen confidence level π if the left-hand-side of Equation (2.12) is between zero and the $(1-\pi)^{\text{th}}$ percentile of the chi-squared distribution given on the right-hand-side, then we conclude that the division of the data into different bins makes no statistically significant difference for the model estimation. We conclude that there is a difference otherwise.

---

[6] Note that the left-hand-side of Equation (2.12) is always non-negative because a combination of models separately estimated for data bins always provides a fit which is at least as good as the fit for the whole data sample.

In the case when data is divided into two bins “a” and “b” one can conduct an alternative likelihood ratio test (Washington et al., 2003, page 282)

$$X^2 = -2\left[LL(\boldsymbol{\beta}_{b,a}) - LL(\boldsymbol{\beta}_a)\right] \sim \chi^2_{\mathrm{df}=K}\ , \qquad \text{Eq. 2.14}$$

where $LL(\boldsymbol{\beta}_a)$ is the log-likelihood of data “a” given a model estimated by MLE using data “a”, and $LL(\boldsymbol{\beta}_{b,a})$ is the log-likelihood of data “a” given the same model estimated by MLE using data “b”. In Equation (2.) the degrees of freedom is equal to K, which is the number of coefficients in each model. The test given by Equation (2.) can also be reversed using $LL(\boldsymbol{\beta}_b)$ and $LL(\boldsymbol{\beta}_{a,b})$ for data “b”.

Note that the likelihood ratio tests given by Equations (2.12) and (2.) are theoretically justified only for large data samples (Gourieroux and Monfort, 1996).

CHAPTER 3. DATA DESCRIPTION

3.1. Design exception and control sites

Our data consists of 32 bridges and 4 roadway intervals with level-one design exceptions (DEs). For a control data sample, we chose 63 control bridges and 8 control roadway intervals without any design exceptions. The control sites were chosen so that their characteristics are similar to those of the design exception (DE) sites. The list of all sites with design exception is given in Table 3.1. In Appendix C we give a map of the Indiana State, which shows the numbers of design exceptions requested and approved in each county in 1998-2003. A list of all control sites is given in Table 3.2.

It is important to note that all bridges are geographically localized sites (they are points on the map). As a result, below we will introduce an “effective radius of influence”, and we will consider accidents that occurred within this radius from the localized sites (bridges). In the next chapter we will estimate the effective radius of influence to be $R_{eff} = 0.55\ \text{miles}$. In contrast, roadway intervals are non-localized sites, and we will consider all accident that occurred in the intervals.

3.2. Accident severity data

The data on individual accidents used in the present study is from the Indiana Electronic Vehicle Crash Record System (EVCRS). The EVCRS was launched in 2004 and includes available information on all accidents investigated by Indiana police starting from January 1, 2003.

Table 3.1 A list of sites with design exceptions

Design #	Location	Structure related	County	Design exception	Road related	Approval date (y/m)
0002230	75-th St. over I-465	bridge	Marion	minimum vertical clearance	I-465	2001/02
9982320	CR580 (Hamiltom Rd.) over Kilmore creek	bridge	Clinton	design speed	CR580	-
0006743	CR600 (MT Comfort Rd.) over I-70	bridge	Hancock	stopping-sight distance	I-70	2003/03
9980560	CR900 (Hunters Creek Rd.) over Little Salt creek	bridge	Lawrence	superelevation transition length, superelevation rate, bridge rail safety	CR900	2001/07
9702290	Ditch Rd. over I-465	bridge	Marion	minimum vertical clearance	I-465	-
0100316	Green Valley Rd. over I-265	bridge	Floyd	bridge rail safety	Green Valley Rd.	2002/07
9614710	I-64 from 0.4 mile West of SR 165 to 0.2 mile West of Owenswille Rd.	interval	Posey, Vanderburgh	stopping-sight distance, horizontal obstruction clearance	I-64	2000/12
9709094	I-64 over abandoned R/R; 7.71 km West of SR 165	bridge	Posey	cross slope	I-64	-
9709095	I-64 over Flat run ditch; 0.95 km West of SR 165	bridge	Posey	cross slope	I-64	-
9614701	I-64 over the Big Bayou river	bridge	Posey	bridge width	I-64	-
9709091	I-64 over the Black river; 6.21 mile West of SR165	bridge	Posey	shoulder width, cross slope	I-64	-
9709093	I-64 over Wilsey Rd.; 5.67 mile West of SR165	bridge	Posey	cross slope	I-64	-
9241915 9241916	I-65 over SR311	bridge	Clark	superelevation transition length	SR311, I-65	-
0100294	I-65 over railroad; 0.06 mile south of SR38	bridge	Tippecanoe	minimum vertical clearance	I-65	2002/06
9884890	Lena Rd. over Conrail railroad	bridge	Clay	vertical alignment	Lena Rd.	-

Table 3.1 (Continued)

Design #	Location	Structure related	County	Design exception	Road related	Approval date (y/m)
9900930	SR5 over Wabash River at Huntington reservoir	bridge	Huntington	shoulder width, cross slope	SR5	-
9803620	SR9 over Loon creek; 2.51 km South of US 24	bridge	Huntington	superelevation rate	SR9	-
9900900	SR9 over Conrail R/R; 0.15 km South of SR 18	bridge	Grant	minimum vertical clearance	SR9	-
9620250	SR56 over French Lick creek, 0.12 miles East of SR145	bridge	Orange	bridge rail safety	SR56	-
9241925	SR60 over I-65	bridge	Clark	stopping-sight distance	SR60, I-65	2003/03
9611950	SR62 over Stinking Fork; 3.69 mile east of SR37	bridge	Crawford	bridge rail safety	SR62	-
9900570	SR63 over CSX railroad	bridge	Vermillion	superelevation transition length	SR63	2000/08
9620230	SR66 over Deer creek; 0.14 mile East of SR166	bridge	Perry	stopping-sight distance, super-elevation transition length	SR66	2001/06
9800320	SR101 over Dubois creek	bridge	Union	vertical alignment, maximum grade	SR101	2000/03
9800310	SR168 over creek; 2.54 mile East of US41	bridge	Gibson, Knox	bridge width	SR168	-
9137955	SR250 over Wades creek; 1.09 km west of SR156	bridge	Switzerland	horizontal curvature, superelevation rate, horizontal obstruction clearance	SR250	2002/08
9244245	SR267 over I-74	bridge	Hendricks	shoulder width	SR267	1998/09
9702050	SR267 over abandoned railroad	bridge	Hendricks	bridge rail safety	SR267	-
9803520	SR327 underpass at CSX railroad	bridge	Decalb	vertical alignment	SR327	1998/03
8915240	US12 over Munson ditch; 1.1 mile East of SR49	bridge	Porter	superelevation rate	US12	2001/01

Table 3.1 (Continued)

Design #	Location	Structure related	County	Design exception	Road related	Approval date (y/m)
8915285	US12 over Unnamed channel; 1.3 mile West of SR49	bridge	Porter	shoulder width	US12	2001/02
8900433	US24 road relocation from 0.8 mile West of CR400 to 0.2 mile East of CR600	interval	Wabash	stopping-sight distance	US24	1999/01
9707490	US41 (Calumet Ave.) from 1.5 mile North of I-80 to 2.8 mile	interval	Lake	vertical alignment	US41	-
9707150	US41 from Sr64 to 1 mile South of SR441	interval	Gibson, Knox	shoulder width	US41	2002/06
0101270	US52 NB exit ramp over I-65	bridge	Boone	horizontal curvature, stopping-sight distance	US52, I-65	2002/08
9702150	US231 over Big Wea creek; 1.56 km North of SR28	bridge	Tippecanoe	superelevation transition length	US231	-

Table 3.2 A list of control sites (without design exceptions)

Related design number	Location	Structure related	County	Road related
0002230	West Hanna Ave. over I-74.	bridge	Marion	I-74
	South eastern Ave. over I-465	bridge	Marion	I-465
9982320	CR150 over creek	bridge	Huntington	CR150
	CR70 over creek	bridge	Warren	CR70
0006743	11A Rd. over I-69	bridge	DeKalb	I-69
	North Little Point Rd. over I-70	bridge	Morgan	I-70
9980560	CR300 (Old Scotland Rd.) over creek	bridge	Greene	CR300
	CR180 over creek	bridge	Greene	CR180
9702290	East 46th St. over I-465	bridge	Marion	I-465
	West 46th St. over I-465	bridge	Marion	I-465
0100316	Klerner Lane over I-265	bridge	Floyd	Klerner Lane
	Payne Koehler Rd. over I-265	bridge	Floyd	Payne Koehler Rd.
9614710	I-64 from 0.4 mile West of SR68 to 0.8 mile East of CR700	interval	Warrick, Spencer	I-64
	I-64 from 0.4 mile West of SR165 to 0.2 mile West of OwensvilleRd.	interval	Perry	I-64
9709094	I-64 over creek	bridge	Warrick	I-64
	I-70 over creek	bridge	Wayne	I-70
9709095	I-70 over ditch	bridge	Hancock	I-70
	I-74 over ditch	bridge	Montgomery	I-74
9614701	I-70 over river	bridge	Vigo	I-70
	I-74 over river	bridge	Vermillion, Fountain	I-74
9709091	I-65 over river	bridge	Huntington	I-69
	I-70 over creek	bridge	Putnam	I-70
9709093	I-65 over West Manson Colfax Rd.	bridge	Clinton	I-65
	I-74 over Whites Hill Rd.	bridge	Dearborn	I-74
9241915	I-65 over US24	bridge	Bartholomew	I-65, SR31
	I-69 over US24	bridge	Allen	I-69, US24
0100294	I-65 over railroad	bridge	Tippecanoe	I-65
	I-69 over railroad	bridge	Delaware	I-69

Table 3.2 (Continued)

Related design number	Location	Structure related	County	Road related
9884890	CR200 (Meridian Rd.) over railroad	bridge	Clay	CR200
	CR500 (South Whitney Rd.) over R/R	bridge	Delaware	CR500
9900930	East Salamonie Dam Rd. at Salamonie reservoir	bridge	Wabash	East Salamonie Dam Rd.
	East Mississinewa Dam Rd. at Salamonie reservoir	bridge	Miami	East Mississinewa Dam Rd.
9803620	US30 over creek	bridge	Kosciusko	US30
	US30 over creek	bridge	Whitley	US30
9900900	SR24 over railroad	bridge	Allen	SR24
	East Raymond Str. over railroad	bridge	Marion	East Raymond Str.
9620250	SR56 over creek	bridge	Orange	SR56
	SR54 West over creek	bridge	Lawrence	SR54
9241925	SR334 over I-65	bridge	Boone	SR334, I-65
	SR39 over I-70	bridge	Hendricks	SR39, I-70
9611950	SR62 over creek	bridge	Harrison	SR62
	SR9 over creek	bridge	Shelby	SR9
9900570	SR37 North over ditch	bridge	Morgan	SR37
9620230	SR66 over creek	bridge	Perry	SR66
	SR129 over creek	bridge	Switzerland	SR129
9800320	US35 over creek	bridge	Pulaski	US35
	SR16 over creek	bridge	White	SR16
9800310	SR67 over creek	bridge	Greene	SR67
	US36 over creek	bridge	Parke	US36
9137955	SR62 over creek	bridge	Jefferson	SR62
	SR121 over creek	bridge	Fayette	SR121
9244245	US52 under I-465	bridge	Marion	US52
	East Main Str. Under I-65	bridge	Johnson	East Main Str.
9702050	Bluff Rd. over creek	bridge	Marion	Bluff Rd.
	SR238 over creek	bridge	Hamilton	SR238
9803520	SR331 underpass at R/R	bridge	Saint Joseph	SR331
	Benham Ave. railroad underpass	bridge	Elkhart	Benham Ave.

Table 3.2 (Continued)

Related design number	Location	Structure related	County	Road related
8915240	US33 over ditch	bridge	Whitley	US33
	US136 over ditch	bridge	Montgomery	US136
8915285	US136 over channel	bridge	Montgomery	US136
	US421 over channel	bridge	Pulaski	US421
8900433	US24 from 0.36 mile West of CR200 to 0.48 mile West of CR750	interval	Huntington	US24
	US30 from 2.23 mile West of SR101 to 0.56 mile East of SR101	interval	Allen	US30
9707490	US41 (Calumet Ave.) from 0.2 mile South of I-80 to Fisher Str.	interval	Lake	Calumet Ave.
	Kennedy Ave. from 0.58 mile South of I-90 to 0.61 mile North of I-80	interval	Lake	Kennedy Ave.
9707150	US41 from 4.3 mile South of SR54 to 0.55 mile North of SR67	interval	Knox, Sullivan	US41
	US41 from 0.55 mile South of SR10 to 4.55 mile North of US24	interval	Newton	US41
0101270	SR37 NB exit ramp over I-69	bridge	Hamilton	SR37, I-69
	SR 912 (Cline Ave.) NB exit ramp over I-90	bridge	Lake	SR912, I-90
9702150	US50 over creek	bridge	Jackson	US50
	US150 over creek	bridge	Washington	US150

Finally, detailed descriptions of all design exceptions considered in Tables 3.1 and 3.2 (along with aerial photos of the design exception sites) are available on request in .pdf computer files.

The information on accidents included into the EVCRS can be divided into three major categories[7]:

1. An Environmental Record – it includes information on circumstances related to an accident. For example, weather, roadway and traffic conditions, number of dead and injured people involved, etc.
2. A Vehicle and Driver Record – it includes information on all vehicles involved into an accident and on all drivers of these vehicles. For example, accident contributing factors by each vehicle, type and model of each vehicle, posted speed limit for each vehicle, driver's injury status, driver's age and gender, driver's name and address, etc.
3. Non-driver Individual Record – it includes information on all people who are involved into an accident but are not drivers. This record includes only the name and address of those people, but it does not include any information on their injuries (if any).

In our study we use only information from the first two categories above. These two categories include 127 variables for each accident, which is an abundance of data. However, we do not need to consider all these variables. Indeed, because our study focuses on accident causation and severity, we choose all information and all data variables that can reasonably be related to the subject of our study, and we consider only these variables. For example, we do not consider the name of the road where an accident took place and the license plate numbers of the vehicles involved because we can reasonably expect that

[7] Note that accident data is subject to missing observations and typos. In addition, there can be misidentification errors on police crash reports due police officers' mistakes and prejudices. We eliminate obvious typos during initial data processing and exclude missing observations, but we do not correct for concealed typos and unobserved misidentification errors. We assume that police misidentification errors are sufficiently small not to affect our final results.

these variables do not contribute to the accident cause and severity. The list of all variables that we consider for accident severity is given in Appendix A.

In the present study, we consider data on 5889 accidents occurred from 2003 to 2007. Among these, 3429 accidents occurred in a $R_{eff} = 0.55$ mile proximity of all bridges (R_{eff} is the effective radius of influence estimated below), and 2460 accidents occurred on all roadway intervals. Of the 3429 accidents occurred near bridges, 1192 accidents occurred in the proximity of design exception bridges and 2237 accidents occurred in the proximity of control bridges. Of the 2460 accidents occurred on roadway intervals, 739 accidents occurred on design exception intervals and 1721 accidents occurred on control intervals.

The percentage distributions of the 5889 accidents that we consider by accident type are given in Figure 3.1[8]. The percentage distribution of the accidents by their severity level is given in Figure 3.2.

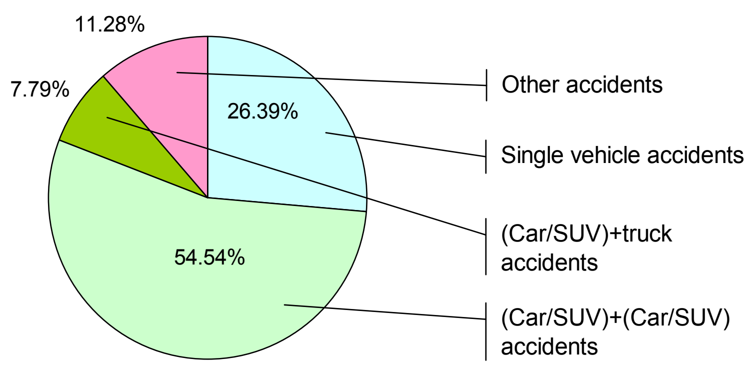

Figure 3.1 Percentage distribution of accidents by their type

[8] For convenience, from each of the percentage distribution plot we exclude accidents for which the considered descriptive variable (e.g. accident type) is unknown.

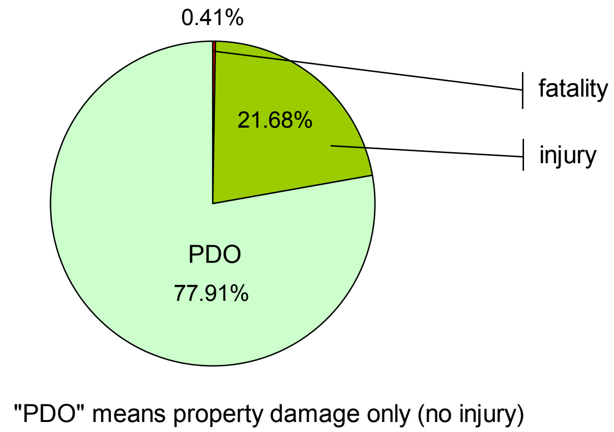

Figure 3.2 Percentage distribution of accidents by their severity level

3.3. Accident frequency data

In our accident frequency study, we estimate negative binomial models for five-year accident frequencies. These are numbers of accidents that occurred on roadway segments over the five-year period 2003-2007. Thus, we need to choose roadway segments, and we make this choice as follows. For all bridges (with and without design exceptions) we choose 1.1-mile long roadway segments around the bridges ($R_{eff} = 0.55$ each way from the bridges). As far as the roadway intervals (with and without design exceptions) are concerned, we divide them into smaller segments that have roughly homogeneous properties (e.g. homogeneous AADT). As a result, we end up with 143 roadway segments. Among those 104 segments are at bridges (35 segments with design exceptions and 69 control segments).[9] The remaining 39 segments are on roadway intervals (13 segments with design exceptions and 26 control segments).

[9] The number of segments at bridges, 104, is larger than the number of all bridges, 95. This is because in some cases we consider accidents on two roads that cross at the bridge sites.

After the roadway segments are chosen, we need to collect information on characteristic properties of these segments. Segment length, locality of the road (rural/urban), number of lanes, median surface type, median width (in feet), interior shoulder presence and width, outside shoulder presence and width, number of bridges, number of horizontal curves, number of ramps, horizontal curve lengths and radii are determined by using the Google Earth software.[10] Average annual daily traffic (AADT) volumes are taken from the Indiana Department of Transportation (INDOT) website (WWW.IN.GOV/INDOT/3238.HTM), most of the AADT volumes are adjusted by using the appropriate growth factors, which are also defined by the INDOT. Road class (interstate, US route, state route, county road, street), rumble strips, median type, road surface type, speed limit value, road type (one-lane, two-lane, multi-lane, one-way, two-way, undivided, divided, alley, private drive) are taken from the available data on individual accidents data (the same data that is used in the severity study). Accident frequencies on roadway segments are found by matching locations of segments and individual accidents for the five-year period considered (2003-2007). The list of all explanatory variables that we consider for accident frequency is given in Appendix B.

---

[10] For each segment, the horizontal curve length and radius are calculated by using the length of the segment curve chord and the maximum perpendicular distance between the chord and the segment curve.

CHAPTER 4. ACCIDENT SEVERITY STUDY

In this chapter we study the severity of accidents and its dependence on the presence/absence of design exceptions and other factors. Below, we first explain how we use the available accident data and estimate statistical models of accident severity. Then, we present the results obtained from the estimation of these models for accidents that happened in Indiana in 2003-2007.

4.1. Modeling Procedures: accident severity

For each accident, the severity level is determined by the injury level sustained by the most severely injured individual (if any) involved into the accident. By using the available individual accident data, we are able to distinguish between three levels of accident severity. Listed in increasing order, these are (refer to Figure 3.2 for injury proportions)

1. no-injury or property damage only (PDO),
2. injury,
3. fatality.

As a result, for the statistical modeling of accident severity we use a multinomial logit model with three possible outcomes that correspond to these three levels of accident severity. This multinomial logit model is given by Equation (2.3), where the outcomes “1”, “2” and “3” correspond to “fatality”, “injury” and “PDO” levels of accident severity respectively. To find important explanatory variables and the best multinomial logit models, we use the model outcome methodology described in detail in Section 2.3 of CHAPTER 2.

Because bridges are geographically localized sites (points on the map) we introduce an effective radius of influence R_{eff}. It is the radius of a circular area around a bridge within which occurring accidents are influenced by the presence of the bridge. We find R_{eff} as follows:

- We estimate a multinomial logit (MNL) model for severity of all accidents inside the 2-mile areas around all bridges (with and without design exceptions). We use model choice methodology described in detail in Section 2.3 to construct this model.
- We define an auxiliary distance variable $\hat{r}_n$

$$\hat{r}_n = \min\{r_n, R\} = \begin{cases} r_n, & r_n \le R \\ R, & r_n > R \end{cases}, \qquad \text{Eq. 4.1}$$

  where r_n is the distance between the n^{th} accident and the nearby bridge, and R is a radius of influence of the bridge on the neighboring accidents. Equation (4.1) implies that the value of the geographical distance r_n (between the bridge and the n^{th} accident) matters and is used only if $r_n \le R$, otherwise this value does not matter and is replaced by R.
- We use variable $\hat{r}_n$, given by Equation 4.1, as an explanatory variable in the multinomial logit (MNL) model for severity of all accidents inside the 2-mile areas around all bridges. We estimate this model by the maximum likelihood estimation (MLE) method for different values of R, $R \in (0,2)$, and find the dependence of the resulting log-likelihood (LL) at the MLE convergence on the value of R.[11] The resulting dependence of LL on the value of R is shown in Figure 4.1. Finally, we define the effective radius R_{eff} of influence to be equal to the value of R that maximizes LL.

[11] Note that the number of accident severity observations is independent of the value of R and is equal to the number of accidents occurred inside the 2-mile areas around all bridges.

We use the above procedure and find that the effective radius of influence is $R_{eff} = 0.55 \text{ miles}$. Hereafter, we consider only those accidents around bridges that occurred within this radius from the bridges. On the other hand, because roadway intervals are non-localized sites (curves on the map), we consider all accidents that occurred on the intervals.

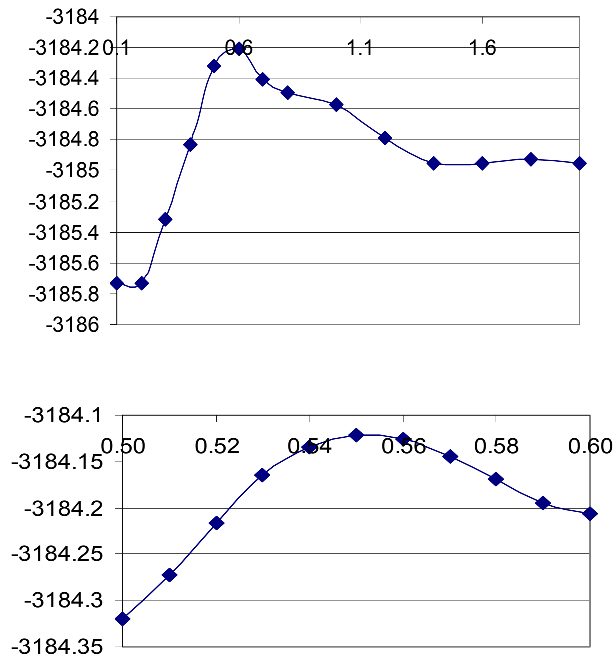

Figure 4.1 Dependence of the log-likelihood of MNL model of accident severity on the radius of influence R

4.2. Results: standard MNL models of accident severities

In this section we consider and estimate standard multinomial logit (MNL) models of accident severities. These models are specified by Equation (2.3) in CHAPTER 2.

First, we test whether severity of accidents near bridges (within $R_{eff} = 0.55$ mile radius) and severity of accidents on roadway intervals should be considered together or must be considered separately. We divide accident data into two bins: the first bin includes accidents occurred near bridges, and the second bin includes accidents on roadway intervals. We then use the likelihood ratio test, explained in Section 2.4, to test whether two MNL models estimated for severity of accidents in the two bins are statistically different. The test result, presented in Table 4.1, shows that the two models are statistically the same. Thus, the severity of accidents occurred near bridges and on roadway intervals should be considered together.

Table 4.1 Likelihood ratio tests for standard MNL models of accident severity

Test purpose	M	K	$LL(\boldsymbol{\beta}_\mathbf{m})$	$\sum LL(\boldsymbol{\beta}_\mathbf{m})$	**df**	**p-value**	**conclusion**
compare bridges and intervals	2	19	-1840.03	-1827.22	19	0.141	the same
compare DE and control sites	2	19	-1840.03	-1831.90	19	0.640	the same

Next, by using the model choice methodology described in Section 2.3, we construct and estimate the best standard multinomial logit (MNL) model of severity of all accidents occurred near bridges (within $R_{eff} = 0.55$ mile radius) and on roadway intervals. Summary statistics for explanatory variables used in the accident severity models is given in Table 4.2. The estimation results for this MNL model are given in Table 4.3.

To determine whether the presence of design exceptions has any effect on accident severity, we carry out the following two tests:

1. First, we include “design exception (DE) presence” indicator variable into the MNL model. We find that this variable is statistically insignificant (see Table 4.3).

2. Second, we divide data into two bins – the first bin contains all accidents that are near bridges and on roadway intervals with design exceptions (DEs), while the second bin contains all accidents near control bridges and on control intervals (without DEs). We estimate the best MNL model separately for severity of the accidents in the two bins. Then, we carry out the likelihood ratio test (see Section 2.4) to determine whether there is a statistically significant difference between the two MNL models estimated for severity of the accidents in the two bins. The test result, given in Table 4.1, shows that there is no significant difference, and we do not need to distinguish between DE sites and control sites, while estimating accident severity.

Thus, we find that design exception presence does not have any statistically significant effect on accident severity.

Finally, as a test, we include “bridge” indicator variable into the best MNL model. We find that this variable is statistically insignificant (see Table 4.3). This confirms the result that severity of accidents occurred near bridges and on roadway intervals should be considered together, as earlier found by the likelihood ratio test.

Table 4.2 Summary statistics for variables used in models of accident severity

Variable	Mean	Standard deviation	Minimum	Maximum
“Age of the oldest driver is ≥ 30 and < 40 years” indicator variable	0.189	0.391	0	1.00
“Primary cause of accident is driver-related” indicator variable	0.858	0.349	0	1.00
“Help arrived in 10 minutes or less after the crash” indicator variable	0.590	0.492	0	1.00
“License state of the vehicle at fault is Indiana” indicator variable	0.794	0.405	0	1.00
“Road classification is "interstate"” indicator variable	0.269	0.443	0	1.00
“The largest number of occupants in all vehicles involved” quantitative variable	1.72	1.39	0	48.0
“No road junction at the accident location” indicator variable	0.620	0.485	0	1.00
“Age (in years) of the oldest vehicle involved” quantitative variable	8.44	5.24	-1.00	40.0
“Traffic control device for the vehicle at fault is a signal” indicator variable	0.127	0.333	0	1.00
“Roadway surface is covered by snow/slush” indicator variable	0.0529	0.224	0	1.00
“Two vehicles are involved” indicator variable	0.705	0.456	0	1.00
“Age (in years) of the vehicle at fault” quantitative variable	7.08	5.18	-1.00	38.0
“Urban locality of the accident” indicator variable	0.680	0.466	0	1.00
“Number of occupants in the vehicle at fault” quantitative variable	1.42	1.11	0	48.0
“Posted speed limit (if the same for all vehicles involved)” quantitative variable	44.4	13.8	5.00	70.0
“Gender of the driver at fault: 1 – female, 0 – male” indicator variable	0.389	0.487	0	1.00

Table 4.3 Estimation results for the standard MNL model of accident severities[12]

Variable	Coefficient (t-ratio)		Averaged elasticities of variables			
	fatality	injury	$\overline{E}^{(1)}_{1;SL}$	$\overline{E}^{(2)}_{1;SL} = \overline{E}^{(3)}_{1;SL}$	$\overline{E}^{(2)}_{2;SL}$	$\overline{E}^{(1)}_{2;SL} = \overline{E}^{(3)}_{2;SL}$
Constant	-6.01 (-10.5)	-3.49 (-9.46)	-	-	-	-
“Two vehicles are involved” indicator variable	-1.87 (-2.92)	-.597 (-4.89)	-1.32	.0020	-.333	.0887
“Roadway surface is covered by snow/slush” indicator variable	-.545 (-2.49)	-.545 (-2.49)	-.0298	.0001	-.0258	.0041
“Help arrived in 10 minutes or less after the crash” indicator variable	.367 (4.03)	.367 (4.03)	.216	-.0007	.166	-.0511
“Number of occupants in the vehicle at fault” quantitative variable	-.159 (-2.66)	-.159 (-2.66)	-.233	.0007	-.185	.0480
“The largest number of occupants in all vehicles involved” quantitative variable	.149 (3.00)	.149 (3.00)	.260	-.0007	.203	-.0579
“Age (in years) of the vehicle at fault” quantitative variable	.140 (3.43)	-	.977	-.0048	-	-
“No road junction at the accident location” indicator variable	-	-.227 (-2.39)	-	-	-.114	.0261
“License state of the vehicle at fault is Indiana” indicator variable	-	.230 (2.09)	-	-	.142	-.0396
“Road classification is "interstate"” indicator variable	-	-.343 (-2.64)	-	-	-.0710	.0165
“Urban locality of the accident” indicator variable	-	-.344 (-2.76)	-	-	-.184	.0470
“Primary cause of accident is driver-related” indicator variable	-	1.58 (8.92)	-	-	1.02	-.308
“Posted speed limit (if the same for all vehicles involved)” quantitative variable	-	.0178 (3.27)	-	-	.631	-.167
“Traffic control device for the vehicle at fault is a signal” indicator variable	-	.399 (3.22)	-	-	.0361	-.0145
“Gender of the driver at fault: 1 – female, 0 – male” indicator variable	-	.178 (2.07)	-	-	.0537	-.0157
“Age (in years) of the oldest vehicle involved” quantitative variable	-	.0265 (3.18)	-	-	.176	-.0508
“Age of the oldest driver is ≥ 30 and < 40 years” indicator variable	-	.317 (3.10)	-	-	.0462	-.0151

[12] Refer to Equations (2.3)–(2.5), where outcomes “1”, “2”, “3” correspond to “fatality”, “injury”, “PDO”.

Table 4.3 (Continued)

Variable	Coefficient (t-ratio)		Averaged elasticities of variables			
	fatality	injury	$\overline{E}^{(1)}_{1;SL}$	$\overline{E}^{(2)}_{1;SL} = = \overline{E}^{(3)}_{1;SL}$	$\overline{E}^{(2)}_{2;SL}$	$\overline{E}^{(1)}_{2;SL} = = \overline{E}^{(3)}_{2;SL}$
"DE presence" indicator variable	.497 (.815)	-.0399 (-.434)	.159	-.0007	-.0101	.0027
"Bridge site" indicator variable	-.235 (-.383)	.0634 (.653)	-.136	.0004	.0288	-.0079
Log-likelihood at MLE convergence	-1840.03					
Restricted log-likelihood	-1951.34					
Number of parameters	19					
Number of observations	3666					
R^2	0.057					

4.3. Results: mixed MNL models of accident severities

In this section we consider and estimate mixed multinomial logit (MNL) models of accident severities. These models are given by Equation (2.7) in CHAPTER 2.

We follow the same procedures as in the previous section, where we considered standard MNL models. First, we use the likelihood ratio test in order to test whether severity of accidents near bridges (within $R_{eff} = 0.55 \text{ mile}$ radius) and severity of accidents on roadway intervals should be considered together or must be considered separately. The test result, presented in the first row in Table 4.4, shows that the severity of accidents occurred near bridges and on roadway intervals should be considered together during estimation of mixed MNL models.

Table 4.4 Likelihood ratio tests for mixed MNL models of accident severity

Test purpose	M	K	$LL(\boldsymbol{\beta}_{\mathbf{m}})$	$\sum LL(\boldsymbol{\beta}_{\mathbf{m}})$	**df**	**p-value**	**conclusion**
compare bridges and intervals	2	21	-1829.57	-1816.28	21	0.185	the same
compare DE and control sites	2	21	-1830.53	-1816.92	21	0.164	the same

Next, we construct and estimate the best mixed multinomial logit (MNL) model of severity of all accidents occurred near bridges (within $R_{eff} = 0.55 \text{ mile}$ radius) and on roadway intervals. The estimation results for this mixed MNL model are given in Table 4.5. To find whether the presence of design exceptions has any effect on accident severity, we again carry out two tests:

1. First, we include “design exception (DE) presence” indicator variable into the mixed MNL model. We find that this variable is statistically insignificant (see Table 4.5).

2. Second, we carry out the likelihood ratio test to determine whether there is a statistically significant difference between the two mixed MNL models estimated for severity of the accidents occurred at DE sites and control sites. The test result, given in the second row in Table 4.4, shows that there is no significant difference, and we do not need to distinguish between DE sites and control sites, while estimating mixed MNL models for accident severities.

Thus, we again find that design exception presence does not have any statistically significant effect on accident severity.

As a test, we again include "bridge" indicator variable into the best mixed MNL model. We find that this variable is statistically insignificant (see Table 4.5). This confirms the result that the severity of accidents that occurred near bridges and on roadway intervals should be considered together, as found by the likelihood ratio test.

Turning to the specific model results shown in Table 4.5, The findings in this table show that the severity model has a very good overall fit (McFadden ρ^2 statistic above 0.5) and that the parameter estimates are of plausible sign, magnitude and average elasticity. We find that two variables produce random parameters (in the mixed-logit formulation). The indicator variable for having two vehicles involved in the crash was found to be normally distributed in the injury-crash outcome with a mean -1.85 and standard deviation of 2.65. This means that for 75.7% of the observations having two vehicles involved in the crash reduced the probability of the injury outcome and for 24.3% of the observations having two vehicles involved increased the probability of an injury outcome. Also, the parameter for the interstate-highway indicator variable is uniformly distributed with a mean of -2.26 and a standard deviation of 6.03.

Some other interesting results included the age of the at-fault vehicle (where elasticity values show that a 1% increase in at-fault vehicle age increases the probability of a fatal injury by 0.972%) and the age of the oldest vehicle involved in the accident (which also increased the probability of an injury). These two variables may be capturing improvements in safety technologies on newer vehicles.

The presence of snow and slush was found to reduce the probability of fatality and injury, likely due to lower levels of friction which may increase collision time and, therefore, allow energy to be more easily dissipated during a crash. Accidents that did not occur at an intersection and those that occurred in urban areas were less likely to result in an injury (by an average of 12.9% and 21% respectively as indicated by the average elasticities). Finally, accidents involving female drivers who were at fault, having the at-fault vehicle under signal control, having higher posted speed limits, and having driver-related causes indicated as the primary cause of the accident all resulted in a higher likelihood of an injury accident.

Table 4.5 Estimation results for the mixed MNL model of accident severities[13]

Variable	Coefficient (t-ratio)		Averaged elasticities of variables					
	fatality	injury	$\overline{E}^{(1)}_{1;SL}$	$\overline{E}^{(2)}_{1;SL}$	$\overline{E}^{(3)}_{1;SL}$	$\overline{E}^{(1)}_{2;SL}$	$\overline{E}^{(2)}_{2;SL}$	$\overline{E}^{(3)}_{2;SL}$
Fixed parameters								
Constant	-6.09 (-10.4)	-4.59 (-7.26)	-	-	-	-	-	-
“Two vehicles are involved” indicator variable	-2.41 (-3.63)	-	-1.45	.0012	.0030	-	-	-
“Roadway surface is covered by snow/slush” indicator variable	-.843 (-2.41)	-.843 (-2.41)	-.0460	.0001	.0002	.0038	-.0255	.0038
“Help arrived in 10 minutes or less after the crash” indicator variable	.609 (3.69)	.609 (3.69)	.358	-.0008	-.0014	-.0488	.1576	-.0488
“Number of occupants in the vehicle at fault” quantitative variable	-.328 (-2.54)	-.328 (-2.54)	-.479	.0010	.0016	.0574	-.2301	.0574
“The largest number of occupants in all vehicles involved” quantitative variable	.303 (2.70)	.303 (2.70)	.526	-.0013	-.0040	-.0666	.243	-.0666
“Age (in years) of the vehicle at fault” quantitative variable	.139 (3.38)	-	.972	-.0033	-.0055	-	-	-
“No road junction at the accident location” indicator variable	-	-.409 (-2.43)	-	-	-	.0285	-.129	.0285
“License state of the vehicle at fault is Indiana” indicator variable	-	.390 (2.11)	-	-	-	-.0390	.145	-.0390
“Urban locality of the accident” indicator variable	-	-.686 (-2.78)	-	-	-	.0533	-.210	.0533
“Primary cause of accident is driver-related” indicator variable	-	2.27 (8.28)	-	-	-	-.255	.814	-.255
“Posted speed limit (if the same for all vehicles involved)” quantitative variable	-	.0239 (2.66)	-	-	-	-.129	.511	-.129
“Traffic control device for the vehicle at fault is a signal” indicator variable	-	.724 (2.90)	-	-	-	-.0146	.0353	-.0146

[13] Refer to Equations (2.3)–(2.5), where outcomes “1”, “2”, “3” correspond to “fatality”, “injury”, “PDO”.

Table 4.5 (Continued)

Variable	Coefficient (t-ratio)		Averaged elasticities of variables					
	fatality	injury	$\overline{E}_{1;SL}^{(1)}$	$\overline{E}_{1;SL}^{(2)}$	$\overline{E}_{1;SL}^{(3)}$	$\overline{E}_{2;SL}^{(1)}$	$\overline{E}_{2;SL}^{(2)}$	$\overline{E}_{2;SL}^{(3)}$
"Gender of the driver at fault: 1 – female, 0 – male" indicator variable	-	.308 (2.06)	-	-	-	-.0155	.0543	-.0155
"age (in years) of the oldest vehicle involved" quantitative variable	-	.0417 (2.96)	-	-	-	-.0457	.160	-.0457
"age of the oldest driver is ≥ 30 and < 40 years" indicator variable	-	.577 (3.12)	-	-	-	-.0162	.0501	-.0162
"DE presence" indicator variable	.460 (.752)	-.0974 (-.622)	.147	-.0004	-.0007	.0038	-.0149	.0038
"bridge site" indicator variable	-.244 (-.395)	.175 (1.11)	-.141	.0003	.0005	-.0119	.0443	-.0119
Random parameters								
"Two vehicles are involved" indicator variable	-	-1.85 (-3.67)	-	-	-	.0223	-.0185	.0223
"Road classification is "interstate"" indicator variable	-	-2.26 (-2.59)	-	-	-	-.0141	.104	-.0141
Standard deviations of parameter distributions								
"Two vehicles are involved" indicator variable	-	2.65 (3.86) normal	-	-	-	-	-	-
"Road classification is "interstate"" indicator variable	-	6.03 (3.74) uniform	-	-	-	-	-	-
Log-likelihood at MLE convergence	-1828.38							
Restricted log-likelihood	-1951.34							
Number of parameters	21							
Number of observations	3666							
R^2	0.546							

CHAPTER 5. ACCIDENT FREQUENCY STUDY

In this chapter we study five-year frequencies of accidents and their dependence on the presence/absence of design exceptions and other factors. Below, we first explain how we use the available accident data and estimate statistical models of accident frequencies. Then, we present the results obtained from the estimation of these models for accidents that happened in Indiana in 2003-2007.

5.1. Modeling Procedures: accident frequency

A five-year accident frequency A_n is the number of accidents occurred on the n^{th} roadway segment during a given five-year time period. In this study, we use negative binomial models for modeling of five-year accident frequencies (see Section 2.2). In addition, we also considered negative binomial models of annual accident frequencies. However, in this case we encountered likelihood convergence problems, which were due to the presence of repeat observations (each roadway segment is observed during five years) and a resulting correlation of error terms in the estimated models.[14] As a result, in this study we focus only on the results of modeling of five-year accident frequencies. All major findings and conclusions reported below for five-year accident frequencies were found to hold for annual accident frequencies as well.

As was the case with accident severity, we use the effective radius of influence around bridges to be $R_{eff} = 0.55$ miles. (Keep in mind that R_{eff} is the radius of a

---

[14] The presence of this correlation was confirmed by estimation of negative monomial models with random effects for annual accident frequencies.

circular area around a bridge within which occurring accidents are influenced by the presence of the bridge.) As a result, for DE and control bridges we choose 1.1-mile long roadway segments around the bridges ($R_{eff} = 0.55$ each way from each bridge), and then consider these segments. As far as DE and control roadway intervals are concerned, we choose roadway segments by dividing these intervals into smaller segments that have the same AADT (averaged annual daily traffic), same number of lanes, etc. This gives 143 roadway segments from the original sample of 26 design exceptions and 71 control segments.

In order to find important explanatory variables and best negative binomial models, we use model choice methodology described in detail in Section 2.3 of CHAPTER 2.

5.2. Results: negative binomial models of accident frequencies

In this section we consider and estimate standard negative binomial (NB) and mixed negative binomial models of five-year accident frequencies. These models are given by Equations (2.8) and (2.10) in CHAPTER 2 respectively.

We attempted the estimation of a random parameters negative binomial model as shown in Equation (2.10). Trying various distributions, all estimated parameters were determined to be fixed at the likelihood convergence (standard deviations of parameter estimates across the population were not significantly different from zero implying that the parameters were fixed across observations). Thus, standard negative binomial models are estimated on five-year accident frequencies, and 122 of the 143 road segments had complete data for use in the accident-frequency model estimation. For these 122 road segments, the average 5-year accident frequency was 34.84 with a standard deviation of 65.51.

The negative binomial estimation results are given in Table 5.1 along with the marginal effects as previously discussed. The results show that the parameter estimates are of plausible sign and magnitude and the overall statistical fit is quite good (McFadden ρ^2 statistic above 0.75).

Table 5.1 Estimation results for the standard NB model of 5-year accident frequencies (at bridges and on roadway intervals)

Variable	Coefficient	t-ratio	Marginal effects
Constant	3.12	7.23	-
"Locality of the road: 1 – urban, 0 – rural" indicator variable	1.80	4.43	71.9
"Degree of curvature of the sharpest horizontal curve on the road segment" quantitative variable	-0.0562	-2.08	-2.24
"Average annual daily traffic per lane in thousands" quantitative variable	0.0509	2.28	2.04
"Logarithm of a roadway segment length" quantitative variable	0.937	2.83	37.5
"Total number of ramps" quantitative variable	0.163	2.00	6.52
"Roadway surface is "blacktop"" indicator variable	-1.08	-3.13	-43.4
"Interior shoulder presence" indicator variable	-1.25	-3.10	-50.1
"Median width is less than 30 feet" indicator variable	-0.905	-2.55	-36.2
Over-dispersion parameter (alpha)	1.37	7.94	-
"DE presence" indicator variable	0.0601	0.204	-
"bridge site" indicator variable	-0.155	-0.414	-
Log-likelihood at MLE convergence	-472.77		
Restricted log-likelihood	-1963.29		
Number of parameters	10		
Number of observations	122		
R^2	0.759		

Table 5.1 shows that the design exception parameter is statistically insignificant again suggesting that design exceptions do not have a statistically significant impact on the frequency of accidents.[15]

[15] The bridge-segment indicator variable was also statistically insignificant suggesting no difference between bridge and non-bridge segments.

Turning to the specific model results shown in Table 5.1, we find that urban roads have a significantly higher number of accidents and that the higher the degree of curvature (defined as 18000 divided by π times the radius of the curve in feet), the lower the accident risk. This second finding seems counterintuitive (sharper curves result in fewer accidents) but this could be reflecting the fact that drivers may be responding to sharp curves by driving more cautiously and/or that such curves are on lower design-speed segments with inherently lower accident risk. Other results in Table 5.1 show that: increases in average annual daily traffic per lane increase accident frequencies (the marginal effect shows that for every 1000 vehicle increase in AADT per lane the 5-year accident frequency goes up by 2.04 accidents); longer road-segment lengths increase accident frequencies (this is an exposure variable because it is related to the number of miles driven on the roadway segment); and for interstates the higher the number of ramps the greater the number of accidents (with marginal effects indicating that each ramp increases the 5-year accident rate by 6.52 accidents).

The asphalt surface indicator was found to result in fewer accidents. This is likely capturing unobserved information relating to pavement friction and condition (as measured by the International Roughness Index, rutting measurements, and so on) because other studies with detailed pavement-condition information have found the type of roadway surface (concrete or asphalt) to be statistically insignificant (see Anastasopoulos et al., 2008 and Anastasopoulos and Mannering, 2009). Finally, for multilane highways, the presence of an interior shoulder and medians widths of less than 30 feet were found to decrease accident frequency. This latter finding is likely capturing unobserved characteristics associated with highway segments that had medians of 30ft or more (which was about 55% of the sample).

We also conducted likelihood ratio tests as was done for the mixed-logit severity analysis. The test statistic X^2, given by Equation (2.14) was 23.00 with 10 degrees of freedom. The corresponding p-value based on the χ^2 distribution, is 0.0107 (the critical χ^2 value at the 90% confidence level is 15.99). However, because we have only a limited number of accident-frequency observations (equal to 122), the parameter estimates of the separate frequency models (for design exception and non-design exception segments) are not necessarily statistically reliable (high standard errors) and the asymptotic χ^2 distribution is likely to be a poor approximation for the test statistic X^2. To resolve this problem, Monte Carlo simulations can be undertaken to find the true distribution of the test statistic X^2. This is done by first generating a large number of artificial data sets under the null hypothesis that the model is the same for segments with and without design exceptions. Then the test statistic values X^2, given by Equation (2.14), for each of the simulated data sets are computed, and these values are used to find the true probability distribution of X^2. This distribution is then used for determining the p-value that corresponds to the X^2 calculated for the actual observed data. The p-value is then used for the inference. This Monte-Carlo-simulations-based approach to the likelihood ratio test is universal, it works for any number of observations (Cowen, 1998).

The true p-value, calculated by using the simulations-based distribution of X^2 is 0.0311, which is about three times larger than the approximate χ^2-based value 0.0107. However, both these values are below 5%. Therefore, the hypothesis that design exception and non-design exception sites were statistically the same is rejected, and it can be concluded that design exceptions have a statistically significant effect on accident frequencies. This is an extremely important finding. The fact that the indicator variable for design exceptions was found to be statistically insignificant suggests that the difference between design exception and non-design exception segments in terms of higher accident frequencies is not significant. However, the likelihood ratio test results suggest that the process (estimated parameters) generating the accident frequencies of the design exception and non-design exception segments are significantly different. This has important implications in that potential changes in explanatory variables **X** could produce significantly different accidents frequencies between design exception and non-design exception segments. While more data would be needed to completely uncover these effects, this finding indicates that caution needs to be exercised even when granting design exceptions that appear to have been acceptable based on historical data.

Finally for additional background information on the frequency data, Table 5.2 presents summary statistics for the full 143 segments in the roadway sample.

Table 5.2 Summary statistics for variables used in models of accident frequency

Variable	Mean	Standard deviation	Minimum	Maximum
“5-year accident frequency on segment” dependent variable	41.13	101.23	0	877
“Average annual daily traffic per lane in thousands” quantitative variable	10.28	9.12	0.12	45.20
“Roadway surface is "blacktop"” indicator variable	0.238	0.426	0	1.00
“Degree of curvature of the sharpest horizontal curve on the road segment”	6.067	17.4	0	109
“Median width is less than 30 feet” indicator variable	0.566	0.496	0	1.00
“Logarithm of a roadway segment length” quantitative variable	0.274	0.463	-0.916	1.89
“Locality of the road: 1 – urban, 0 – rural” indicator variable	0.266	0.442	0	1.00
“Interior shoulder presence” indicator variable	0.566	0.496	0	1.00
“Total number of ramps” quantitative variable	1.14	2.03	0	9.00

CHAPTER 6. DISCUSSION

Overall, our results suggest that the current process used to grant design exceptions has been sufficiently strict to avoid adverse safety consequences resulting from design exceptions – although the finding that different processes may be generating the frequencies of accidents in design exception and non-design exception segments is cause for concern with regard to future granting of design exceptions.

Our specific findings (even with the limited data available to us) provide some insight into areas where caution should be exercised when granting Level One design exceptions. With regard to the severity of accidents, while most of the factors that affected severity were driver characteristics, we did find that urban-area accidents have a lower likelihood of injury and that the posted speed limit is critical (higher speed limits result in a significantly higher probability of an injury accident). Thus, urban/rural location and design exceptions on highways with higher speed limits need to be given careful scrutiny.

With regard to the frequency of accidents, we find that horizontal curvature is critical and thus special attention needs to be paid to design exceptions relating to horizontal curves. For multilane highways, the presence of interior shoulders was found to significantly reduce the frequency of accidents so this should be considered carefully when granting design exceptions. Also, higher accident frequencies were found in urban areas suggesting that special attention should be given to design exceptions that could compromise safety in these areas (as expected, urban areas have higher accident frequencies but lower severities).

Finally, the asphalt-surface indicator was found to result in fewer accidents. As stated previously, this is likely capturing unobserved information relating to pavement friction and condition (as measured by the International Roughness Index, rutting measurements, and so on), and suggests that friction and pavement conditions have to be watched closely when design exceptions are granted.

In terms of a process in the form of a decision support system for guiding future Level One design exceptions, the statistical findings of this research effort suggest that using previous design exceptions as precedents would be a good starting point. While the current study indicates that the design exceptions granted over the 1998-2003 timeframe have not adversely affected overall safety, the number of available design exceptions is too small to make broad statements with regard to policy. Thus, a case by case comparison with previously granted design exceptions is the only course of action that can be recommended.

LIST OF REFERENCES

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

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

Agent, K., Pitman, J., Stamatiadis, N., 2003. Safety implications from design exceptions. Kentucky Transportation Center, KTC-02-09/SPR230-01-1F, Lexington, KY.

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

American Association of State Highway and Transportation Officials, 2004. A guide for achieving flexibility in highway design, Washington, DC.

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

Carson, J., Mannering, F., 2001. The effect of ice warning signs on accident frequencies and severities. Accident Analysis and Prevention 33(1), 99-109.

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

Cowen, G., 1998. Statistical data analysis. Oxford University Press, USA.

Duncan C., Khattak, A., Council, F., 1998. Applying the ordered probit model to injury severity in truck-passenger car rear-end collisions. Transportation Research Record, 1635, 63-71.

Eluru, N., Bhat, C., Hensher, D. (2008). A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accident Analysis and Prevention, 40(3), 1033–1054.

Federal Highway Administration, 1997. Flexibility in highway design. United States Department of Transportation, FHWA-PD-97-062.

Gourieroux, C., and Monfort, A., 1996. Statistical Methods in Econometrics. Volume 2, Cambridge University Press, Cambridge.

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

Hadi, M. A., Aruldhas, J., Chow, L.-F., and Wattleworth, J. A., 1995. Estimating safety effects of cross-section design for various highway types using negative binomial regression. Transportation Research Record, 1500, 169

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

Jones, B., Janssen, L., Mannering, F., 1991. Analysis of the frequency and duration of freeway accidents in Seattle. Accident Analysis and Prevention 23(2), 239-255.

Khattak, A., 2001. Injury severity in multi-vehicle rear-end crashes. *Transportation Research Record*, *1746*, 59-68.

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

Khorashadi, A., Niemeier, D., Shankar, V., and 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.

Kockelman, K., Kweon, Y.-J., 2002. Driver injury severity: An application of ordered probit models. Accident Analysis and Prevention 34(4), 313-321.

Kweon, Y.-J., Kockelman, K., 2003. Overall injury risk to different drivers: Combining exposure, frequency, and severity models. Accident Analysis and Prevention, 35(3), 414-450.

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

Lee, C., Abdel-Aty, M., 2005. Comprehensive analysis of vehicle-pedestrian crashes at intersections in Florida. Accident Analysis and Prevention 37(4), 775-786.

Lord, D., 2006. Modeling motor vehicle crashes using Poisson-gamma models: Examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accident Analysis and Prevention 38(4), 751-766.

Lord, D., Park, Y.-J., 2008. Investigating the effects of the fixed and varying dispersion parameters of Poisson-gamma models on empirical Bayes estimates, Accident Analysis and Prevention, 40(4), 1441-1457.

Lord, D., Guikema, S., D., Geedipally, S., R., 2008. Application of the Conway–Maxwell–Poisson generalized linear model for analyzing motor vehicle crashes, Accident Analysis and Prevention, 40(3), 1123-1134.

Malyshkina, N., Mannering, F., 2009. Markov switching multinomial logit model: An application to accident-injury severities. Working paper.

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

Malyshkina, N., Mannering, F. and Thomaz, J., 2009b. Safety impacts of design exceptions. Prepared for the Joint Transportation Research Program, Indiana Department of Transportation.

McFadden, D. (1981). Econometric Models of probabilistic choice. In Manski & D. McFadden (Eds.), A structural analysis of discrete data with econometric applications. Cambridge, MA: The MIT Press.

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

Milton, J., and Mannering, F., 1998. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. Transportation, 25(4), 395-413.

National Cooperative Highway Research Program, 2003. Design exception practices. National Cooperative Highway Research Program Synthesis 316, Transportation Research Board, Washington, DC.

Poch, M., and Mannering, F. L., 1996. Negative binomial analysis of intersection accident frequency. Journal of Transportation Engineering, 122(2), 105-113.

Savolainen, P., and Mannering F., 2007. Additional evidence on the effectiveness of motorcycle training and motorcyclists' risk-taking behavior. Transportation Research Record 2031, 52-58.

Savolainen, P. T., and Mannering, F., 2007. Probabilistic models of motorcyclists' injury severities in single- and multi-vehicle crashes. Accident Analysis and Prevention, 39(5), 955-963.

Shankar, V., Mannering, F., and Barfield, W., 1995. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. Accident Analysis and Prevention, 27(3), 371-389.

Shankar, V., Mannering, F., and Barfield, W., 1996. Statistical analysis of accident severity on rural freeways. Accident Analysis and Prevention, 28(3), 391-401.

Shankar, V., Milton, J., Mannering, F., 1997. Modeling accident frequencies as Zero–Altered probability processes: An empirical inquiry. Accident Analysis and Prevention 29, 829–837.

Ulfarsson, G. F., and Mannering, F. L., 2004. Differences in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger car accidents. Accident Analysis and Prevention, 36(2), 135-147.

Wang, X., Abdel-Aty, M., 2008. Modeling left-turn crash occurrence at signalized intersections by conflicting patterns. Accident Analysis and Prevention, 40(1), 76-88.

Washington, S. P., Karlaftis, M. G., and Mannering, F. L., 2003. Statistical and econometric methods for transportation data analysis. Chapman & Hall/CRC, Boca Raton, Florida.

Yamamoto, T. , Shankar, V., 2004. Bivariate ordered-response probit model of driver's and passenger's injury severities in collisions with fixed objects, Accident Analysis and Prevention 36(5), 869-876.

Appendix A.

List of all explanatory variables considered for accident severity study:

X_3 – Collision date

X_4 – Day of the week

X_5 – Collision time

X_{13} – Construction
(*no; yes; buck-up of traffic outside of but due to construction zone*)

X_{14} – Light condition
(*daylight; dawn / dusk; dark with street lights on; dark with no lights*)

X_{15} – Weather condition
(*clear; cloudy; sleet/hail / freezing rain; fog / smoke / smog; rain; snow; severe cross wind*)

X_{16} – Surface condition
(*dry; wet; muddy; snow / slush; ice; loose material on roadway; water*)

X_{17} – Type of median
(*drivable; curbed; barrier wall; none*)

X_{18} – Type of roadway junction
(*no junction involved; four-way intersection; ramp T-intersection; Y-intersection; traffic circle / roundabout; five point or more; interchange*)

X_{19} – Road character
(*straight / level; straight / grade; straight / hillcrest; curve / level; curve / grade; curve / hillcrest; non roadway crash*)

X_{20} – Primary contributing circumstance
(*alcoholic beverages; illegal drugs; driver asleep or fatigue; prescription drugs; driver illness; unsafe speed; failure to yield right of way; disregard signal / red signal; left of center; improper passing; improper turning; improper lane usage; following too closely; unsafe backing; overcorrecting / oversteering; ran off road right; ran off road left; wrong way on one way; pedestrian action; passenger distraction; violation of license restriction; jackknifing; cell phone usage; other telematics in use; other (explain in narrative); driver distracted [explain in narrative]; speed too fast for weather conditions; engine failure or defective; accelerator failure or defective; brake failure or defective; tire failure or defective;*

headlight defective or not on; other lights defective; steering failure; window / windshield defective; oversize / overweight load; insecure / leaky load; tow hitch failure; other explained in narrative; glare; roadway surface condition; holes / ruts in surface; shoulder defective; road under construction; severe crosswinds; obstruction not marked; lane marking obscured; view obstructed; animal on roadway; traffic control problem; other [explained in narrative]; utility work)

X_{22} – Time when help arrived

X_{25} – Vehicle type, considered for the vehicle at fault, i.e. for the vehicle that contributed to the primary cause of an accident
(*passenger car / station wagon; pickup; van; sport utility vehicle; truck [single 2 axle, 6 tires]; truck [single 3 or more axles]; truck / trailer [not semi]; tractor / one semi trailer; tractor / double trailer; tractor / triple trailer tractor [cab only, no trailer]; motor home / recreational vehicle; motorcycle; bus/seats 9-15 persons with driver; bus / seats 15+ persons with driver; school bus; unknown type; farm vehicle; combination vehicle; pedestrian; bicycle*)

X_{26} – Vehicle year, considered for all vehicles involved

X_{27} – Number of occupants, considered for all vehicles involved

X_{28} – Vehicle license state, considered for the vehicle at fault, i.e. for the vehicle that contributed to the primary cause of an accident
(*Indiana; Indiana's neighboring states [IL, KY, OH, MI]; other US states; Canada / Mexico / U.S. Territories; other foreign countries*)

X_{29} – Speed limit, considered only if known and the same speed limit value for all vehicles involved

X_{30} – Road type, considered for the vehicle at fault, i.e. for the vehicle contributed to the primary cause of an accident
(*one lane [one way]; two lanes [one way]; multi-lanes [one way]; two lanes [two way]; multi-lane undivided [two way]; multi-lane undivided 2-way left [two way]; multi-lane divided 3 or more lanes [two way]; alley; private drive*)

X_{31} – Traffic control, considered for the vehicle at fault, i.e. for the vehicle contributed to the primary cause of an accident
(*officer / crossing guard / flagman; RR crossing gate / flagman; RR crossing flashing signal; RR crossing sign; traffic control signal; flashing signal; stop sign; yield sign; lane control; no passing zone; other regulatory sign / marking; none*)

X_{33} – Fire, considered for all vehicles involved
(*no; yes*)

X_{34} – Driver age, considered for all drivers involved

X_{35} – Driver gender, considered for all drivers involved

Appendix B.

List of all explanatory variables considered for accident frequency study:

X_2 – Number of accidents on a roadway segment during a five-year period

X_7 – Site characteristics
(design exception (DE) site ; control (not DE) site)

X_8 – Site type
(bridge; interval)

X_9 – Roadway classification
(*interstate; US route; state route; county road; city street; unknown*)

X_{10} – Rumble strips presence
(*no; yes*)

X_{11} – Type of median
(*drivable; curbed; barrier wall; none*)

X_{12} – Roadway surface
(*concrete; blacktop; brick; dirt / gravel; other*)

X_{13} – Speed limit, in miles per hour

X_{14} – Road type, considered for the vehicle at fault, i.e. for the vehicle contributed to the primary cause of an accident
(*one lane [one way]; two lanes [one way]; multi-lanes [one way]; two lanes [two way]; multi-lane undivided [two way]; multi-lane undivided 2-way left [two way]; multi-lane divided 3 or more lanes [two way]; alley; private drive*)

X_{15} – Length of the roadway segment, in miles

X_{16} – Locality of the roadway segment
(*urban; rural*)

X_{15} – Number of lanes on the travel way (in one direction)

X_{16} – Surface of the median section
(*paved; grass; grass with trees; grass with bushes; other; none*)

X_{19} – Width of the median section, in feet

X_{20} – Interior shoulder presence
(*no; yes*)

X_{21} – Interior shoulder width, in feet
(*no; yes*)

X_{22} – Outside shoulder presence
(*no; yes*)

X_{23} – Outside shoulder width, in feet
(*no; yes*)

X_{24} – Number of bridges along the roadway segment

X_{25} – Number of horizontal curves along the roadway segment

X_{28} – Length of the sharpest horizontal curve, in miles

X_{29} – Radius of the sharpest horizontal curve, in miles

X_{30} – Number of ramps along the roadway segment

X_{34} – Average Annual Daily Traffic, in vehicles per day

Appendix C.

Indiana State map with numbers of design exceptions shown in counties