INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road. Ann Arbor. MI 48106-1346 USA
313/761-4700 800/521-0600

LIMITED DEPENDENT VARIABLE AND STRUCTURAL EQUATIONS MODELS: EMPIRICAL APPLICATIONS TO TRAFFIC OPERATIONS AND SAFETY

by

Venkataraman N. Shankar

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

1997

Approved by ______________________________
Chairperson of Supervisory Committee

____

____

____

Program Authorized
to Offer Degree Civil Engineering

Date JUNE 6, 1997

UMI Number: 9736377

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

Doctoral Dissertation

In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to University Microfilms, 1490 Eisenhower Place, P.O. Box 975, Ann Arbor, MI 48106, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature

Date June 6, 1997

University of Washington

Abstract

LIMITED DEPENDENT VARIABLE AND STRUCTURAL EQUATIONS MODELS: EMPIRICAL APPLICATIONS TO TRAFFIC OPERATIONS AND SAFETY

by Venkataraman N. Shankar

Chairperson of the Supervisory Committee: Professor Fred L. Mannering
Department of Civil Engineering

This dissertation presents empirical applications of econometric and statistical methodologies to traffic operations and safety. The motivation of this research was to develop methodological frameworks to investigate factors affecting rural freeway safety and operations in an intelligent transportation system (ITS) setting. The research effort focuses on methodological frameworks for evaluating pre-ITS conditions but is applicable to post-ITS settings as well.

Methodologies used in this dissertation include limited dependent variable models such as Poisson and negative binomial regressions as well as nested logit structures. and structural models involving simultaneous equations. The Poisson and negative binomial models investigate accident likelihoods on roadway sections. while the nested logit structure examines the conditional likelihood of accident severity. Simultaneous equations models examine factors affecting the cross-sectional endogeneity between lane mean speeds and lane-speed deviations. Accounting for cross-sectional endogeneity captures traffic flow dynamics which critically affect safety. These techniques are applied to an empirical setting where ITS infrastructure is being installed by the Washington State Department of Transportation (WSDOT) on a 61-kilometer section of rural Interstate 90 (I-90) located some 50 kilometers east of Seattle. Variable message signs at critical roadway locations coupled with in-vehicle signing in a selected number of vehicles will be provided to inform travelers of adverse driving conditions. The study area includes the Snoqualmie Pass summit, and experiences significant climatic interactions coupled with challenging roadway geometrics.

Findings from this dissertation provide significant insights into the complex interactions and contemporaneous effects of spatial, temporal, environmental, geometric and traffic flow factors affecting accident causality and cross-sectional traffic flow dynamics. The approaches embodied in this dissertation, while being local in model specification, have broader implications beyond ITS, encompassing critical regional and national infrastructure design, programming and investment issues relating to traffic safety and operations. They afford greater flexibility in decision making through enhancement of the design strategy identification and definition process.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Number *Page*

LIST OF ABBREVIATIONS

ITS. Intelligent Transportation Systems

WSDOT. Washington State Department of Transportation

ILS. Indirect Least Squares

2SLS. Two-Stage Least Squares

3SLS. Three-Stage Least Squares

GLS. Generalized Least Squares

LIML. Limited Information Maximum Likelihood

ACKNOWLEDGMENTS

My sincere appreciation goes to Professors Mahoney and Rutherford with the Department of Civil Engineering, Professor Ward Helms with the Electrical Engineering Department and Professor Woodrow Barfield with the Industrial Engineering Department at Virginia Polytechnic University, Blacksburg, Virginia for their availability and advice on crucial issues of this effort. Special thanks go to my colleague John Milton for his fellowship and valuable assistance in data collection needed for this effort.

Most importantly, I owe a lifetime of indebtedness to Professor Mannering for his unlimited availability, inspiration, guidance, knowledge sharing and fellowship. His challenges to bring the best out of this dissertation served as great motivating forces in this endeavor.

DEDICATION

Dedicated to my family. Their unlimited support. confidence. availability. genuine interest and fellowship made this undertaking all the more meaningful and worthwhile. Thanks for being.

INTRODUCTION

RESEARCH OBJECTIVE AND MOTIVATION

This dissertation presents empirical applications of econometric and statistical methodologies to traffic operations and safety. The motivation of this research was to develop methodological frameworks to investigate factors affecting rural freeway safety and operations in an intelligent transportation system (ITS) setting. The research effort focuses on methodological frameworks for evaluating pre-ITS conditions but is applicable to post-ITS settings as well. Examination of post-ITS conditions is a significant exercise in itself involving considerable data collection and is beyond the scope of this dissertation.

RESEARCH APPROACH AND DESCRIPTION

Methodologies used in this dissertation include limited dependent variable models such as Poisson and negative binomial regressions as well as nested logit structures and structural models involving simultaneous equations. The Poisson and negative binomial models investigate accident likelihoods on roadway sections, while the nested logit structure examines the conditional likelihood of accident severity. Simultaneous equations models examine factors affecting the cross-sectional endogeneity between lane-by-lane mean speeds and lane-by-lane speed deviations. Accounting for cross-sectional endogeneity captures traffic flow dynamics which critically affect safety. These techniques are applied to an empirical setting where ITS infrastructure is being installed by the Washington State Department of Transportation (WSDOT) on a 61-kilometer section of rural Interstate 90 (I-90) located some 50 kilometers east of Seattle. Variable message signs at critical roadway locations coupled with in-vehicle signing in a selected number of vehicles will be provided to inform travelers of adverse weather conditions. The I-90 study area includes the Snoqualmie Pass summit, and experiences significant climatic interactions with challenging roadway geometrics.

Chapter 1 discusses the effect of roadway geometrics and environmental factors on rural accident frequencies. Poisson and negative binomial models of accident frequencies are proposed. In addition to an overall accident frequency model, several specific accident types such as sideswipe, angle, rear-end, parked-vehicle, overturn, same-direction, and fixed-object collisions are also proposed. Departure from the regular Poisson distribution for several of these accident types was observed due to the overdispersion of the data. The negative binomial distribution which relaxes the mean-variance equality assumption of the Poisson distribution was employed. Estimation was conducted using standard maximum likelihood procedures.

Chapter 2 discusses the statistical analysis of rural accident severities. Discrete choice model structures are applied to model the distribution of accident severities as logit functions of environmental, geometric, vehicular and behavioral factors. Four discrete severity categories are modeled, and a nested logit structure is proposed for predicting the probability of severity conditioned on the fact that an accident occurred. It was observed that the milder severities namely property damage and possible injury categories shared unobservables and hence were incorporated into the lower level of the two-level nesting structure. The remaining two severity categories, evident injury, disabling injury/fatality were modeled as independent alternatives. Estimation of parameters was conducted using the sequential estimation process.

Chapter 3 discusses structural models of vehicular speed-flow-speed deviation relationships in rural freeway corridors. Structural models of speed-flow-speed deviation relationships are proposed. Spot mean speeds and spot speed deviations measured by lane and cross-sectional flows were the endogenous variables in the model. In addition, temporal factors such as time-of-day, time-of-week and seasonal factors, vehicle-mix factors are also proposed in the specification. Structural models were found to vary by direction of travel. For each direction, two structural models that recursively relate to each other were developed. The mean speed-flow model with lane-by-lane mean speed as the dependent variable was estimated first using endogenous relationships between spot mean speeds by lane, cross-sectional flows by lane and other exogenous variables. Predicted values of lane-by-lane mean speed from this model were used as exogenous variables along with endogenous lane-by-lane speed deviations in the speed deviation structural model. Estimation was conducted using the method of three-stage least squares (3SLS). Order and ranking conditions for identification of the equations was examined prior to conducting the estimation. By using the 3SLS method and accounting for contemporaneous correlation across equations, we eliminate the bias and inconsistency that would arise from modeling each of the speed-flow and speed deviation-flow-speed equations as separate linear regressions.

Chapter 4 concludes with a summary of findings and recommendations for further research.

CHAPTER 1: EFFECT OF ROADWAY GEOMETRICS AND ENVIRONMENTAL FACTORS ON RURAL FREEWAY ACCIDENT FREQUENCIES

1.1: RESEARCH OBJECTIVE

Beginning in the winter 1996, the Washington State Department of Transportation proposed to implement an Intelligent Vehicle Highway System (ITS) on Interstate 90 (I-90) to reduce the likelihood of vehicular accidents in the vicinity of Snoqualmie Pass (located about 50 kilometers east of Seattle, WA, see Figure 1). This ITS system was to consist of variable message signs, variable speed-limit signs and in-vehicle signing. The intent of the signing is to reduce accident risk by warning drivers of adverse weather conditions, ice on the roadway, and lane blockages caused by vehicle accidents and/or disablements. However, to evaluate the effectiveness of such a signing system, it is important to first develop an understanding of the factors that have historically contributed to the likelihood of an accident in the study area (i.e., before the signing system is implemented). With this in hand, a before and after statistical comparison of the data can be conducted and the effectiveness of the signing system can be assessed. The intent of this chapter is to investigate and report the "before" conditions to isolate the factors that are contributing to accident likelihoods[1].

---

[1] The study of historic data is important even if the evaluation of some future accident reduction strategy is not being undertaken. This is because an analysis of historic accident data can provide valuable information relating to the nature of accidents, factors affecting their frequency, and the likely effectiveness of alternative accident mitigation strategies.

The section of Interstate 90 shown in Figure 1 experiences a high number of vehicular accidents as a result of challenging roadway geometrics (i.e., small horizontal curve radii and steep grades) and adverse weather conditions. The climate in the vicinity of the Snoqualmie Pass summit is severe. At an elevation of over 900 meters above sea level, the area receives an average of over 100 centimeters of rainfall and over 1700 centimeters of snowfall, annually. Snowfall occurs during every month except July and August. During a large portion of the year, residual snow and ice can accumulate on the ground contribute to adverse driving conditions. Factors that contribute to the accidents include driver behavior, geometric characteristics (e.g., grade and curve radii), weather-related variables (e.g., rainfall and snowfall, intensity of snowfall and rainfall), interactions between geometrics and weather elements, and seasonal effects such as traffic volume, precipitation and ambient temperature-related variations.

The intent of this chapter is to focus on factors affecting accident risk, specifically roadway geometrics, weather conditions and their complex interactions.

1.2: PREVIOUS RESEARCH

Previous research, for the most part, has dealt with modeling relationships between accident occurrences and geometric elements. Examples of this include the work of Wong and Nicholson (1992). They observed that modifications to roadway geometrics were important because of the strong association between adverse geometric elements and high-accident locations. This association has been confirmed in studies by Boughton (1975), National Cooperative Highway Research Program (1978), and the

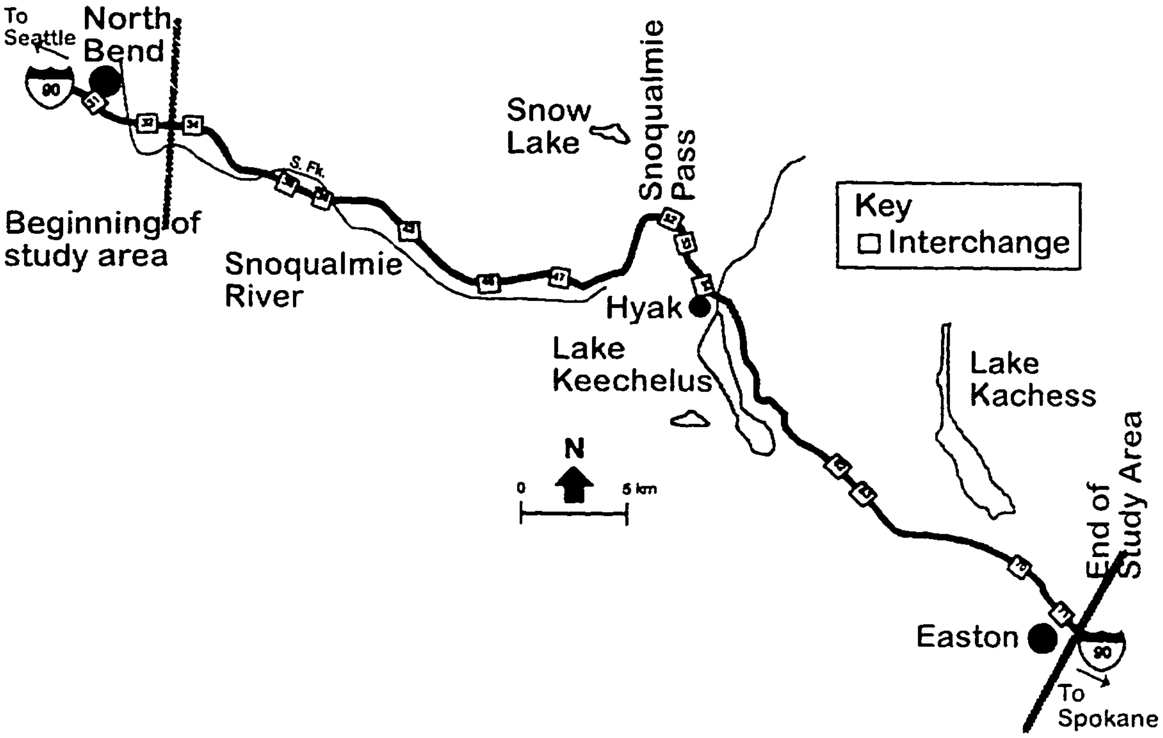

Figure 1. Map of I-90 Study Area.

Federal Highway Administration (1982). Other empirical relationships between vehicle accidents and highway geometrics have been studied through the use of statistical models to investigate accident involvement rate, accident probability, geometric design variables critical to safety, and the accident reduction potential of geometric improvements (National Cooperative Highway Research Program 1978; Hammerslag, Roos, and Kwakernaak 1982; Okamoto and Koshi 1981; Miaou, Hu, Wright, Davis, and Rathi 1991).

In terms of the relationship between accidents and weather elements, a number of important studies have been conducted (Ivey, Griffin, Newton, Lytten, Hankins 1981; Jovanis and Delleur 1981; Mori and Uematsu 1967; Snyder 1974). This past work studied the effect of rainfall and snowfall on accident occurrences and attempted to quantify the contribution of these environmental elements to increasing accident likelihoods. Other types of methodologies have also been applied to the problem of accident analysis. For example, risk-based approaches, applied to the prediction of wet-weather accidents, have also been documented (Brodsky and Hakkert 1988) and, recently, the effect of winter pavement maintenance on accident rates has been investigated (Hanbali 1992). Finally, seasonal variations in weather elements coupled with corresponding variations in traffic volumes have been examined (Jones, Janssen and Mannering 1991) in a multi-variate framework.

Although past research work has provided insight into the effect of weather on accident rates and frequencies, efforts to investigate the interaction of weather and geometric elements, and their consequent impact on accident likelihoods, have been minimal. The study of such interactions is important because it could shed light on the impact of weather on critical geometric design elements and serve as a guide in the design of roadway geometrics so as to minimize accident likelihoods in the presence of varying climatic conditions. This concept contrasts with present roadway geometric design practice which applies a uniform nationwide standard in terms of assumed weather impacts on geometric design (Mannering and Kilareski. 1990). Presumably, much could be gained by adjusting this standard to account for weather conditions that deviate greatly from the norm.

In addition to weather and geometrics, it may be argued that human factors contribute significantly to accident occurrences and hence warrant inclusion in the modeling effort. Previous research (Treat 1980; Sabey and Taylor 1980) indicates that human factors are involved in 95% of all traffic accidents, either alone or in combination with other factors. However, other research (Massie, Campbell, Blower 1993) tempers the criticism of research excluding human factors by pointing out that the human factors approach ignores the problem associated with classifying collisions and their related causes, be it human or otherwise. The authors add that such an approach fails to address the issue of helping drivers avoid collisions. Identification of geometric and weather-related factors and their interrelationship can be used to assist the driver in reducing the chances of a collision by offsetting the ignorance factor caused by unanticipated changes in roadway geometrics and their interrelation with adverse weather conditions.

From a methodological perspective, attempts to model accident frequencies have varied from the use of least squares regression techniques to methods involving exponential distribution families including the Poisson and negative binomial models. Previous research on Poisson and least squares (Jovanis and Chang 1986; Joshua and Garber 1990; Miaou and Lum 1993) indicates the inappropriateness of least squares techniques to modeling of accident frequencies, and recommends the employment of the Poisson distribution. The Poisson distribution, however, suffers from an important limitation, namely that the mean and variance are constrained to be equal. Over-dispersion (variance greater than the mean) or under-dispersion (variance less than the mean) of data violates this constraint and leads to biased coefficient estimates. A more general distribution, such as the negative binomial, has been employed in such situations (Engel 1984; Lawless 1987; Manton, Woodbury and Stallard 1981) to relax this constraint. Negative binomial distributions have been employed frequently in physics, medical sciences and marketing. Documented use of the negative binomial distribution in the field of traffic engineering includes applications in trip generation (Frisbie 1980) and transportation economics (Hellerstein 1991). In terms of applying the negative binomial distribution to model accident occurrences, research has been conducted on accident proneness (Bates and Neyman 1952), accident migration (Maher 1987; 1990), accident "blackspots" identification (Senn and Collie 1988) and accident frequencies (Miaou 1994; Maher 1991; Poch and Mannering 1994).

The body of extant literature provides important methodological direction for our study of the interrelationship between roadway geometrics and weather and accident frequencies. Details of this methodological direction are discussed in the following section.

1.3: METHODOLOGY

Count data are often modeled by assuming Poisson distributions (Cameron and Trivedi 1986). The Poisson distribution is a useful starting point because; (1) it lends itself well to the modeling of count data by virtue of its discrete, non-negative, integer-distribution characteristics, and (2) can be generalized to more flexible distributional forms. In terms of accident frequencies, in this study we will focus on modeling the number of accidents occurring on a specified section of roadway in a one month time period. In such a case, the Poisson distribution gives

$$P(n_{ij}) = e^{-\lambda_{ij}} \lambda_{ij}^{n_{ij}} / n_{ij}! \qquad (1.3.1)$$

where $P(n_{ij})$ is the probability of n accidents occurring on roadway section i in month j and λ_{ij} is the expected number of accidents on roadway section i in month j. Given a vector of geometric, traffic and weather data, λ_{ij} can be estimated by the equation,

$$\ln \lambda_{ij} = \mathbf{X}_{ij}\beta \qquad (1.3.2)$$

where **X** is a vector of geometric, traffic and weather data for roadway section i in month j and β is a vector of estimable coefficients. As mentioned in our review of previous research, the Poisson distribution constrains the mean and variance to be equal, (i.e., $E[n_{ij}] = Var[n_{ij}]$). As previously mentioned, estimation using a Poisson distribution violating this assumption (i.e., when data are overdispersed or underdispersed) results in biased estimates of β. It is well known, based on the findings of many previous research efforts, that accident frequency data tend to be over-dispersed, with the variance being significantly greater than the mean. Consequently, the Poisson distribution can lead to erroneous coefficient estimates and erroneous inferences can be drawn. To overcome this, the negative binomial distribution, which includes a gamma-distributed error term, is appropriate because it relaxes the Poisson's mean-variance equality constraint. The negative binomial model is derived by re-writing equation 2 as,

$$\ln \lambda_{ij} = \mathbf{X}_{ij}\beta + \varepsilon_{ij} \qquad (1.3.3)$$

where $\exp(\varepsilon_{ij})$ is a gamma-distributed error term with mean one and variance α. This results in the mean-variance relationship,

$$\mathrm{Var}[n_{ij}] = E[n_{ij}][1 + \alpha E[n_{ij}]] \qquad (1.3.4)$$

If α is significantly different from zero, the data are over-dispersed or underdispersed. If α is equal to zero the negative binomial reduces to the Poisson distribution.

The resulting probability distribution under the negative binomial assumption is,

$$P(n_{ij}) = \frac{\Gamma(\theta + n_{ij})}{\Gamma(\theta) n_{ij}!} u_{ij}^{\theta}(1 - u_{ij})^{n_{ij}} \qquad (1.3.5)$$

where $u_{ij} = \theta/(\theta+\lambda_{ij})$, $\theta = 1/\alpha$, and $\Gamma(.)$ is a value of the gamma function. Estimation of λ_{ij} can be conducted through standard maximum likelihood (ML) procedures (see Greene

1993). Using equation 5, the likelihood function (the product of probabilities) for the negative binomial is,

$$L(\lambda_{ij}) = \prod_{i=1}^{N}\prod_{j=1}^{T}\frac{\Gamma(\theta + n_{ij})}{\Gamma(\theta) n_{ij}!}\left[\frac{\theta}{\theta + \lambda_{ij}}\right]^{\theta}\left[\frac{\lambda_{ij}}{\theta + \lambda_{ij}}\right]^{n_{ij}} \qquad (1.3.6)$$

where T is the last month of accident data and N is the total number of roadway sections. This function is maximized to obtain coefficient estimates for β and α.

Careful attention must be paid to the appropriateness of the negative binomial distribution in the case of overdispersed data. For example, equation (3) may hold while the distribution of n_{ij} conditioned on $\mathbf{X}_{ij}$ may not be negative-binomial distributed. In such a case, the coefficient estimates will be consistent though less efficient than those for the correct distribution. Importantly, the asymptotic variance-covariance matrix will be incorrect and likely underestimated. However, in practice, this underestimation is not likely to affect substantive conclusions drawn from model estimation (see Lawless 1987).

In addition to maximum likelihood estimation procedures, other methods such as quasilikelihood, weighted least squares (McCullagh and Nelder 1983), moment estimation techniques (Breslow 1984) and regression-based estimation (Cameron and Trivedi 1986, 1990) are available. Examples of the application of the moment method and regression-based estimation in accident modeling indicates that these methods should be used with caution (Miaou 1994). Indications from statistical research on the estimation of α, the dispersion coefficient, suggest that for large samples (N > 20) the quasilikelihood and maximum likelihood methods perform best (Piegorsch 1990).

1.4: EMPIRICAL SETTING

The study area consists of a 61 kilometer portion of I-90 located some 48 kilometers east of Seattle (see Figure 1). This portion of Interstate 90 generally consists of a three-lane (3.66 meter lanes) cross-section, in each direction, with 3.05 meter shoulders and a 104.6 kph speed limit. Virtually no variation in travel lane and shoulder widths exists in the study area.

Data from a number of sources were gathered over the period from January 1988 to May 1993. Precipitation data were assembled from the Desert Research Institute and Western Regional Climate Center and the geometric attributes of the roadway and accident data were obtained from the Washington State Department of Transportation. The available precipitation data consisted of information relating to monthly rainfall and snowfall including average monthly snowfall and rainfall, maximum daily snowfall and rainfall and number of snowy and rainy days per month. Three weather stations located at Snoqualmie Falls, Stevens Pass and Cle Elum (all in Washington State near the study area) were used as the sources of climatic data. Weather data were assigned to sections based on their geographic proximity and elevation levels[2].

---

[2] As a result, several contiguous sections shared the same weather information (this can be seen in Table 1). Shared weather data raises the issue of serial correlation of model error terms. Weather information shared by contiguous sections causes any shocks in data to propagate through sections common to that data, thereby causing spatial correlation. To date, the effect of such spatial correlation has not been specifically investigated in a count data model context. However, based on experiences in linear regression contexts, it can be reasonably assumed that spatial correlation could cause some loss of efficiency of parameter estimates. Studies have shown that, in most practical contexts, this is not a major concern (Mannering 1995).

Geometric characteristics included: number of horizontal curves, number of horizontal curves underdesigned (those curves with design speeds less than 112.6 kph, less than 96.5 kph, and less than 80.45 kph), maximum and minimum horizontal radii, number of vertical curves and maximum and minimum grades.

With this data in hand, the issue of dividing the study area into manageable sections of roadway must be addressed. The existing literature addresses several important issues relating to roadway section length determination in a linear regression context (Okamoto and Koshi 1989). The findings of these studies show that great care must be taken in determining roadway section lengths because of two model estimation concerns; (1) the possibility of heteroskedasticity (i.e., error terms are not identically distributed), and (2) the possibility of biased model coefficients. Heteroskedasticity, especially in the context of a negative binomial specification (as opposed to a Poisson specification), is an important issue due to the incorporation of the gamma-distributed error term.

The most popular alternatives for determining roadway section lengths are the use of fixed-length sections or homogeneous sections (i.e., sections with homogeneous geometric characteristics, see Miaou, Hu, Wright, Davis, and Rathi 1991). With regard to homogeneous sections (both in terms of geometrics and weather), several important problems arise. One of these problems is that roadways with numerous horizontal curves and grades tend to produce sections that are less than 1 kilometer in length (i.e., to ensure homogeneity in geometrics). This can result in locational error problems because accidents, in most states, are locationally reported to the nearest milepost (1.609

kilometers). Potential bias resulting from such accident-reporting locational error is clearly undesirable.

Homogeneity of weather data presents a different problem. Weather data, by virtue of their geographic characteristics, usually encompass much larger areas and, if allowed to govern section lengths, are likely to result in long geometrically diverse sections, thus violating geometric homogeneity.

Finally, the unequal length of sections that will result from the homogeneity requirement may exacerbate potential heteroskedasticity problems (i.e., unequal sample sizes, see Mannering 1995) and lead to a loss in estimation efficiency. The resulting increase in the standard errors of model coefficients could lead the analyst to draw erroneous inferences with regard to the effects of model covariates.

The disadvantages of using fixed-length sections, relative to homogeneous sections, are far less severe. In fact, most potential disadvantages can be overcome by accounting for the non-homogeneity of geometric and weather-related variables by including detailed measures of the variability across sections in the model specification (e.g., number of curves, maximum grade and number of underdesigned curves, and so on). If such data are available, there is little need to constrain the analysis to homogeneous sections. Moreover, fixed-length sections may offer other advantages such as being able to mitigate the effects of the accident migration which is a phenomenon involving the migration of accidents to a different portion of a hazardous roadway section after corrective measures have been taken on some other portion of the roadway (see

Boyle and Wright 1984; McGuigan 1985; Maher 1987). If one were to use geometrically homogeneous sections, it would be exceedingly difficult to account for the effect that changes in accident likelihoods on one section would have on others (due to accident migration). However, the use of fixed-length, non-homogeneous sections accounts for the possibility of accident migration, to some extent, because the migration across the homogeneous "sub-sections" that comprise the fixed length section is internalized.

As a result of the above discussion, the sections considered in this study were determined to be fixed, equal-length sections. Thus, accident frequencies and associated geometric and weather data were compiled along ten sections, of equal length, over the 61 kilometer study area (i.e., each section is 6.1 kilometers in length). Accident frequencies and roadway geometrics for both roadway directions (eastbound and westbound) were used[3]. A total of 2,225 reported accidents occurred in the study area between January 1988 and May 1993[4]. Accidents were sorted by year and month and integrated with geometric and monthly weather data into one database. The consolidated database, after accounting for some missing weather data (which resulted when weather stations were not functioning due to mechanical failures) consisted of 464 observations with some sections experiencing zero accidents in some months[5]. The implicit

---

[3] Interstate 90 has divided cross-sections with different grades and horizontal curve attributes in three of the ten study sections. By combining both east and west directions, we constrain the β's to be the same. An empirical test of this assumption revealed that this constraint is statistically valid.

[4] In this analysis we include only those accidents reported to the Washington State Highway Patrol (WSP). Although this section of highway is heavily patrolled by WSP, it is likely that some minor accidents are never reported.

[5] Note that our data has repeated observations from the same section of roadway. That is, each section produces as many as 12 observations (corresponding to 12 months) per

specification of accident frequency per month as the dependent variable allows the modeling of seasonal variations in traffic volumes, ambient temperature and other environmental data such as daylight duration.

Table 1 summarizes the averages of the variables measured in this study. Mean section accident frequency per month was 3.26 (Figure 2 shows average per-month accident frequencies by section) with an observed monthly minimum of zero and maximum of 28 (the observed monthly variance was 16.32). Other values worthy of note include the high number of horizontal curves on the ten sections. The twelve horizontal curves in section 6 (sections 6 and 7 are near the summit) suggest complex geometrics in the area (i.e., about two horizontal curves per kilometer). Also the average monthly snowfall, observed to be 145.78 centimeters in sections 5-8, is quite high and reflects the severe climate resulting from the relatively high elevation.

1.5: MODEL ESTIMATION

The negative binomial estimation of section-accident frequencies is presented in Table 2. This table shows that all variables are of plausible sign with reasonably high statistical significance.

---

year. Such data raises the possibility of error term correlation among observations produced by the same section, with observations from the same section sharing unobserved factors that may impact accident likelihoods (e.g., a scenic distraction). A likely consequence of such correlation is some loss in efficiency of coefficient estimates. However, research by Mannering and Winston (1991) indicates that the efficiency loss from this source is small, particularly if section-specific constants are included in the model specification (as will be the case in this study).

Table 1. Sample Summary Statistics (section averages).

Variable	Section 1	Section 2	Section 3	Section 4	Section 5	Section 6	Section 7	Section 8	Section 9	Section 10
Accident frequency (per month)	1.80	2.25	1.66	2.49	7.81	8.35	5.92	4.15	2.81	2.86
Number of curves with a design speed less than 128.7 kph	1	3	1	3	1	5	9	2	8	2
Number of curves with a design speed less than 96.5 kph	0	1	1	0	1	4	6	2	7	1
Number of curves with a design speed less than 80.5 kph	0	0	0	0	1	1	1	0	0	0
Number of horizontal curves in section	8	8	10	9	10	12	10	9	10	4
Maximum horizontal curve radius in section (m)	3030	3636	909	3030	1515	3030	695	1736	1736	1818
Minimum horizontal curve radius in section (m)	636	595	595	606	333	333	347	347	347	788

Table 1 (continued). Sample Summary Statistics (section averages).

Variable	Section 1	Section 2	Section 3	Section 4	Section 5	Section 6	Section 7	Section 8	Section 9	Section 10
Number of vertical curves in section	7	8	9	10	8	5	16	7	15	5
Maximum grade in section	5.00	3.00	1.76	3.63	5.29	4.22	2.00	2.60	3.83	5.00
Minimum grade in section	0.03	0.27	0.14	0.46	3.29	0.67	0.43	0.08	0.20	0.74
Average monthly rainfall (cm)	5.01	5.01	5.01	5.01	8.70	8.70	8.70	8.70	1.93	1.93
Maximum daily rainfall in the month (cm)	2.73	2.73	2.73	2.73	5.28	5.28	5.28	5.28	1.40	1.40
Number of rainy days in the month	1.09	1.09	1.09	1.09	2.11	2.11	2.11	2.11	0.56	0.56
Average monthly snowfall (cm)	1.70	1.70	1.70	1.70	145.78	145.78	145.78	145.78	8.5	8.5
Maximum daily snowfall in the month (cm)	1.28	1.28	1.28	1.28	27.65	27.65	27.65	27.65	3.30	3.30
Number of snowy days in the month	0.20	0.20	0.20	0.20	10.31	10.31	10.31	10.31	1.24	1.24

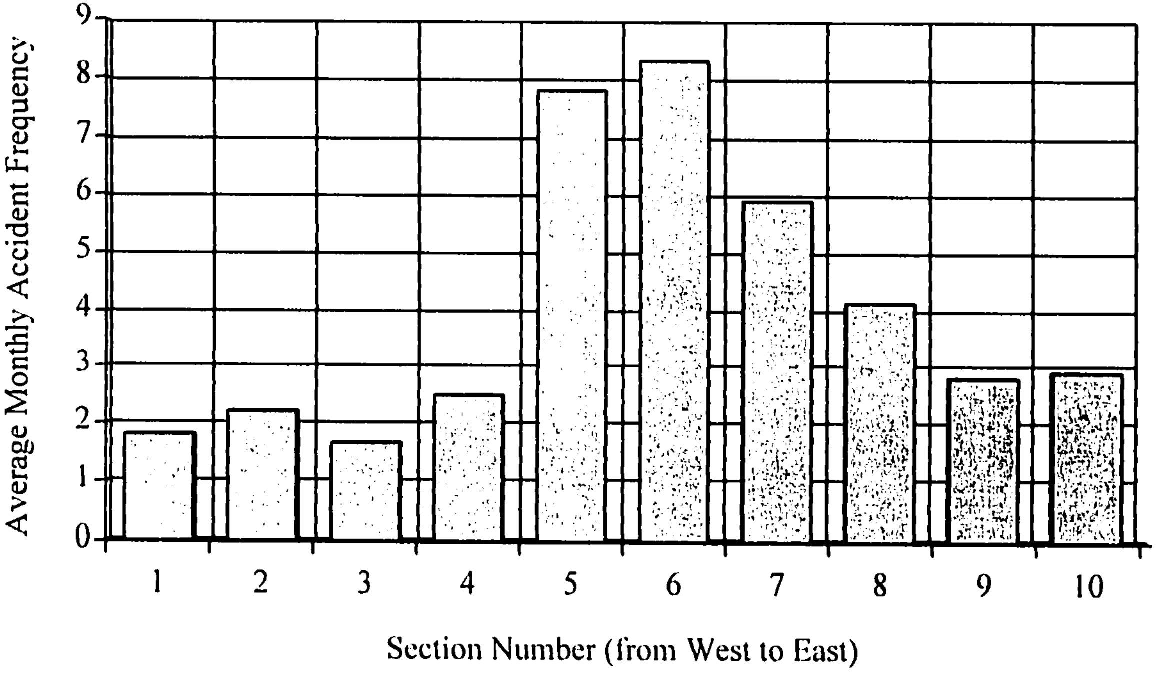

Figure 2. Average Monthly Accident Frequencies on the 10, 6.1 Kilometer Sections (Sections Numbered Sequentially from West to East in the Study Area).

Table 2 shows that the majority of independent variables specified in this model positively affect accident frequency indicating a likelihood in increase in frequency with increasing variable values. The number of curves variable provides some insight into potential geometric hazards. The number of curves with design speeds between 96.5 kph and 128.7 kph appears to have a greater effect (0.117) than those designed under 96.5 kph (0.046). The higher coefficient value for higher design-speed curves is likely capturing the tendency of drivers to slow down for curves with low design speeds due to a combination of the visual effect of the curve and speed reduction signs usually found in those locations.

Grade appears to have a strong positive effect on accident frequency. although in a stepwise. as opposed to a continuous. manner. In comparison to those sections with grades less than 2 percent. those with maximum grades exceeding 2 percent will experience a significant increase in accident frequency. Intuitively. this captures the effect of speed differentials that play a significant role in accident occurrences. although. to some extent. the presence of climbing lanes offsets the detrimental impact of grades especially those impacts caused by slow-moving heavy vehicles. In the present context. however. a geometric variable accounting for climbing lane effects was not found to be significant because there is little variation in this variable across sections. A review of the data showed that any vertical grade reasonably long (longer than 2 kilometers) and exceeding 2 percent had a climbing lane.

Maximum rainfall played a significant. positive role in accident occurrences. Employed as an indicator variable. it captures not only the effect of intensity of rainfall

Table 2. Negative Binomial Estimation Results (Total Section Accident Frequency).

Variable	Estimated Coefficient	t-statistic
Number of horizontal curves designed between 96.5 kph and 128.7 kph	0.117	2.437
Number of horizontal curves designed below 96.5 kph	0.046	2.205
Maximum grade in section indicator (1 if greater than 2%, 0 otherwise)	0.133	2.748
Maximum rainfall indicator (1 if greater than 2.54 centimeters on any given day in the month, 0 otherwise)	0.209	1.401
Number of rainy days in the month	0.018	1.975
Rainfall-Curve interaction indicator (1 if maximum rainfall greater than 2.54 centimeters on any given day in the month and at least one horizontal curve has a design speed less than 96.5 kph, 0 otherwise)	0.184	1.239
Maximum daily snowfall in the month	0.033	2.231
Snowfall-Grade interaction indicator (1 if maximum snowfall greater than 5.1 centimeters on any given day in the month and grade greater than 2%, 0 otherwise)	0.291	1.930
Snowfall-Curve interaction indicator (1 if maximum snowfall greater than 5.1 centimeters on any given day in the month and at least one horizontal curve has a design speed less than 96.5 kph, 0 otherwise)	0.387	2.137
Section location indicator (1 if section number is 5, 6, 7 or 8, 0 otherwise)	-0.466	-1.812
Year of occurrence indicator (1 if 1988, 0 otherwise)	0.273	2.330
Year of occurrence indicator (1 if 1990, 0 otherwise)	-0.167	-1.410
α (dispersion coefficient)	0.418	8.463
Number of observations	464	
Log-likelihood at zero	-2193.39	
Log-likelihood at convergence	-970.93	
ρ^2	0.56	

and potential hydroplaning of vehicles but also may be capturing the effects of exposure and pavement condition. For example, the pavement surface is likely to remain wet or icy during the night or early morning when daily rainfall exceeds 2.54 centimeters.

The number of rainy days played a significant, positive role in accident occurrences. This variable appears to capture exposure effects such as exposure to wet pavements and lower visibility effects. More interestingly, given the fact that the Seattle area generally experiences intermittent rainfall throughout the year, drivers may be inclined to pay less attention to the risk of an accident during rainy weather. The number of rainy days variable could possibly be playing a surrogate role for increased accident risk arising from driver complacency.

Maximum daily snowfall intuitively captures the positive effect that snow plays in accident occurrences. Maximum snowfall exceeding 5.08 centimeters, employed as an indicator variable, appears to account for traction and lane-marking-related problems caused by increasing snow depth on the pavement. In combination with grades, as evidenced by the interaction term, it positively impacts accident frequency. This illustrates the dangerous combination of traction, lane-markings and speed differentials. In addition, it also suggests that the effect of climbing lanes could likely be annulled by the obliteration of lane markings on snow-covered pavements. In the presence of underdesigned horizontal curves, the snowfall variable portrays a stronger effect than the grade interaction by virtue of its higher coefficient.

The section location indicator variable shows that the middle portion of the study corridor (sections 5, 6, 7, and 8, which include the summit and the immediate area surrounding it) is associated with lower accident rates with all other factors held constant[6]. This is likely the result of changes in driver behavior, with drivers becoming more cautious as they gain elevation and approach/depart from the Snoqualmie Pass Summit.

The year 1988 was found to positively affect accident frequencies. Although normal precipitation levels were observed during this year, the positive coefficient value captures unobserved effects such as unusually cold ambient temperatures resulting in ice-covered pavements and construction-related effects such as lane closures[7].

The year 1990, in a similar manner, was specified as an indicator variable. The negative effect of this variable seems to account for some decrease in traffic volumes as well as extra caution used by drivers in the presence of adverse driving conditions created by abnormally high levels of precipitation that occurred during the year.

Finally an examination of ρ^2 (0.52) for the model indicates a good statistical fit, while the dispersion coefficient, α, was estimated to be significantly different from zero (t

---

[6] Note that this does not imply that these sections of the study area have lower overall accident rates (see Figure 2). It only indicates that these sections have lower than expected accident rates when the accumulated effects of geometrics and weather have been taken into account.

[7] It should be noted that since I-90 is a captive corridor with few alternate routes to/from Eastern Washington, construction activities did not cause significant decreases in traffic volumes in the study corridor .

= 8.463) indicating overdispersion of data, a phenomenon that can not be handled by a Poisson distribution[8].

Other issues worthy of note in the estimation context pertain to the impact of weather-related variables on pertinent variables such as traffic volume, temperature and daylight time. The high level of significance of the weather-related variables coupled with their interaction with geometric variables suggests that they capture seasonal trends in traffic volume, temperature and daylight time as well[9]. The significance of weather-related variables and their use as surrogates for traffic volumes is corroborated in previous research (Jones, Janssen and Mannering 1991).

It should also be noted that we tried to include a variety of other interactions between two variables and among three or more variables in our model (e.g., rainfall exceeding 2.54 centimeters on any given day in the month, at least one horizontal curve

---

[8] It is interesting to note that several variables that were found to be significant in a Poisson specification of our model turned out to be insignificant under the negative binomial assumption. This occurred because the Poisson specification underestimated coefficient variances due to the inherent overdispersion of data. Variables found significant in the Poisson but insignificant in the negative binomial included average daily rainfall for the month, number of snowy days in the month, average snowfall in the month and curve radii.

[9] The absence of traffic volume, temperature, and other variables in the model raises the possibility of a model specification error (i.e., an omitted variables bias). To test for this we used a series of month indicator variables (e.g., January, February, etc.) and time of year variables (e.g., winter, summer, spring, and autumn). These variables are highly correlated with traffic volumes (and their seasonal variation), temperature variations, and other possible omitted variables. These indicator variables were all statistically insignificant, suggesting the possible omitted variables bias is not playing a significant role in our model.

with less than 96.5 kph design speed, and a grade greater than 2 percent). However, all such variables produced statistically insignificant coefficients and were thus excluded from our final specification.

Elasticities of independent variables were estimated to determine the impact of those variables on accident frequency. Elasticities can be roughly interpreted as the percentage change in the average frequency of accidents λ_{ij} due to a one percent change in the independent variable. Elasticity of accident frequency λ_{ij}, with respect to x_{ijk} (the kth independent variable for section i in month j) is defined as,

$$E_{x_{ijk}}^{\lambda_{ij}} = \frac{\partial\lambda_{ij}}{\lambda_{ij}} \cdot \frac{x_{ijk}}{\partial x_{ijk}} \qquad (1.5.1)$$

Using equation 3, equation 7 gives,

$$E_{x_{ijk}}^{\lambda_{ij}} = \beta x_{ijk} \qquad (1.5.2)$$

where β is the coefficient corresponding to covariate x_{ijk}.

With equation 8, elasticities of λ_{ij} for each section observation were computed and sample averages were then estimated. Note that the elasticities of indicator variables are not meaningful, so only the elasticities of continuous variables are presented in Table 3.

Table 3 provides some interesting insights. For example, a 1 percent increase in the number of rainy days in a month causes a 0.26 percent increase in accident frequencies. Similarly, a 1 percent increase in the maximum daily snowfall in a month results in a 0.10 percent increase in accident frequencies. This suggests that, at least for these two variables, accident likelihoods may be more sensitive to rain than snow.

However, these are not the only snow/rain variables in the model (i.e., indicator variables are not included in Table 3) and, as will be shown, indicator variables that show an interaction between climatic conditions and roadway geometrics have a large impact on accident frequencies.

Table 3. Accident Frequency Elasticity Estimates.

Elasticity with respect to:	Value
Number of rainy days in the month	0.2624
Maximum daily snowfall in the month	0.1012
Number of horizontal curves designed between 96.5 kph and 128.7 kph	0.1346
Number of horizontal curves designed below 96.5 kph	0.0968

Finally, it is also important to point out that all variables shown in Table 3 are inelastic (elasticity less than unity). This suggests that, while the effect of these variables on accident frequencies is statistically significant, they may be nearing thresholds where accident frequencies have relatively low sensitivity to any changes in the explanatory variables.

To gather some understanding of the relative importance of the indicator variables included in the model, a numerical computation can be performed to provide an idea of the relative effect of indicator variables on average accident frequency. This is accomplished by using a ratio of coefficients. For example, the average accident frequency λ_{ij} for section i in month j can be said to increase 14.0% ($e^{0.133}/e^{0}$), if the

Table 4. Percentage Change in Accident Frequencies Due To Indicator Variables

Variable	Percentage Change in mean accident frequency, λ_{ij}
Maximum grade in section indicator (1 if greater than 2%, 0 otherwise)	14.2
Maximum rainfall indicator (1 if greater than 2.54 centimeters on any given day in the month, 0 otherwise)	23.2 to 48.1[a]
Rainfall-Curve interaction indicator (1 if maximum rainfall greater than 2.54 centimeters on any given day in the month and at least one horizontal curve has a design speed less than 96.5 kph, 0 otherwise)	20.2
Snowfall-Grade interaction indicator (1 if maximum snowfall greater than 5.1 centimeters on any given day in the month and grade greater than 2%, 0 otherwise)	33.8
Snowfall-Curve interaction indicator (1 if maximum snowfall greater than 5.1 centimeters on any given day in the month and at least one horizontal curve has a design speed less than 96.5 kph, 0 otherwise)	47.3
Section location indicator (1 if section number is 5, 6, 7 or 8, 0 otherwise)	-32.3
Year of occurrence indicator (1 if 1988, 0 otherwise)	31.4
Year of occurrence indicator (1 if 1990, 0 otherwise)	-15.4

[a] It is assumed that the change in one indicator variable will not be accompanied by a simultaneous change in any other variable with the exception of the interaction variables. For example, a change in the maximum monthly rainfall variable (to greater than 2.54 cm) does not affect the location dummy. However, by virtue of the monthly rainfall's interaction with horizontal curves, maximum rainfall could have an additive effect because it could influence two variables. This explains the percentage range shown for maximum rainfall.

maximum grade on the section is raised to exceed 2%, assuming the error terms are independent of x_{ij} and remain unchanged. Table 4 presents the change in the average accident frequency caused by threshold changes in the indicator variables.

This table shows that snowfall-horizontal curve and snowfall-grade indicators have a large effect on accident frequencies (47.3% and 33.8% respectively). Rainfall indicators also strongly impact accident frequencies. These findings underscore the importance of accounting for weather/geometric interactions when assessing accident likelihoods.

1.5.1: ACCIDENT FREQUENCY MODELS BY TYPE OF ACCIDENT

In addition to modeling overall accident frequency on highway sections (i.e., as demonstrated above) separate regressions of specific accident types will also provide valuable information. Separate regression models have the potential for providing greater explanatory power relative to a single, overall frequency model because separate models allow coefficient estimates to vary by the type of accident. Intuitively, such variation is seems reasonable. For example, we would expect a steep grade to have a different effect on the likelihood of an overturn accident than it would on a rear-end accident.

To evaluate the impacts of geometrics and weather on specific accident types, models were estimated for accidents classified as, sideswipes, rear-end, parked vehicles, fixed object, overturns, and same direction (all others). Estimation results for these models are presented in tables 5 through 10. All models were negative binomial

Table 5. Negative Binomial Estimation Results Monthly Section "Sideswipe" Accident Frequency.

Variable	Estimated Coefficient	t-statistic
Constant	-2.772	-4.011
Number of horizontal curves designed below 96.5 kph	0.102	1.977
Lowest horizontal curve radius in section (meters)	0.01027	1.104
Number of rainy days in the month	-0.019	-1.132
Maximum rainfall indicator (1 if greater than 2.54 centimeters on any given day in the month, 0 otherwise)	0.959	3.910
Number of snowy days in the month	0.029	1.290
Snowfall-Grade interaction indicator (1 if maximum snowfall greater than 5.1 centimeters on any given day in the month and grade greater than 2%, 0 otherwise)	0.930	3.869
Year of occurrence indicator (1 if 1988, 0 otherwise)	0.483	2.013
α (dispersion coefficient)	0.396	1.420
Number of observations	464	
Log-likelihood at zero	-498.78	
Log-likelihood at convergence	-298.14	
ρ^2	0.40	

Table 6. Negative Binomial Estimation Results Monthly Section "Rear-End" Accident Frequency

Variable	Estimated Coefficient	t-statistic
Constant	-4.368	-5.346
Number of horizontal curves designed below 96.5 kph	0.080	1.679
Maximum grade in section	0.310	2.211
Maximum grade in section indicator (1 if greater than 2%, 0 otherwise)	1.271	1.732
Maximum rainfall on any given day in the month	-0.381	-1.902
Number of rainy days in the month	-0.048	-1.741
Average daily rainfall in any given month	0.149	2.324
Rainfall-Curve interaction indicator (1 if maximum rainfall greater than 2.54 centimeters on any given day in the month and at least one horizontal curve has a design speed less than 96.5 kph, 0 otherwise)	0.983	3.309
Maximum daily snowfall in the month (1 if maximum snowfall greater than on 5.1 centimeters on any given day in the month)	3.468	3.215
Snowfall-Grade interaction indicator (1 if maximum snowfall greater than 5.1 centimeters on any given day in the month and grade greater than 2%, 0 otherwise)	-1.964	-2.382
Snowfall-Curve interaction indicator (1 if maximum snowfall greater than 5.1 centimeters on any given day in the month and at least one horizontal curve has a design speed less than 96.5 kph, 0 otherwise)	-1.707	-2.262
Section location indicator (1 if section number is 5, 6, 7 or 8, 0 otherwise)	0.815	2.408
Year of occurrence indicator (1 if 1988, 0 otherwise)	0.747	2.797
Year of occurrence indicator (1 if 1989, 0 otherwise)	0.762	2.908
α (dispersion coefficient)	0.910	3.023
Number of observations	464	
Log-likelihood at zero	-544.59	
Log-likelihood at convergence	-310.84	
ρ^2	0.43	

Table 7. Negative Binomial Estimation Results Monthly Section "Parked Vehicle" Accident Frequency.

Variable	Estimated Coefficient	t-statistic
Constant	-3.290	-5.523
Number of horizontal curves designed below 96.5 kph	-0.167	-1.741
Maximum rainfall indicator (1 if greater than 2.54 centimeters on any given day in the month, 0 otherwise)	0.906	2.274
Maximum daily snowfall (1 if greater than 5.1 centimeters on any given day in the month, 0 otherwise)	2.500	3.705
Snowfall-Grade interaction indicator (1 if maximum snowfall greater than 5.1 centimeters on any given day in the month and grade greater than 2%, 0 otherwise)	-1.381	-2.294
Year of occurrence indicator (1 if 1988, 0 otherwise)	0.629	1.385
Year of occurrence indicator (1 if 1989, 0 otherwise)	1.273	3.179
Spring/Summer month indicator (1 if April, May, June, July or August, 0 otherwise)	-0.819	-1.607
α (dispersion coefficient)	1.505	2.375
Number of observations	464	
Log-likelihood at zero	-487.54	
Log-likelihood at convergence	-177.51	
ρ^2	0.64	

Table 8. Negative Binomial Estimation Results Monthly Section "Fixed Object" Accident Frequency.

Variable	Estimated Coefficient	t-statistic
Constant	-2.156	-5.515
Number of horizontal curves designed between 96.5 kph and 128.7 kph	0.154	1.957
Number of horizontal curves designed below 96.5 kph	-0.130	-2.737
Number of horizontal curves in section	0.285	5.032
Maximum rainfall indicator (1 if greater than 2.54 centimeters on any given day in the month, 0 otherwise)	0.423	1.932
Number of rainy days in the month	0.023	1.821
Rainfall-Curve interaction indicator (1 if maximum rainfall greater than 2.54 centimeters on any given day in the month and at least one horizontal curve has a design speed less than 96.5 kph, 0 otherwise)	-0.507	-2.144
Maximum snowfall indicator (1 if greater than 5.1 centimeters on any given day in the month, 0 otherwise)	0.654	3.198
Number of snowy days in the month	0.050	2.604
Section location indicator (1 if section number is 1, 2, 3 or 4, 0 otherwise)	-1.431	-4.349
Section location indicator (1 if section number is 5, 6, 7 or 8, 0 otherwise)	-1.017	-3.257
Year of occurrence indicator (1 if 1988, 0 otherwise)	0.283	2.206
Spring/Summer month indicator (1 if April, May, June, July or August, 0 otherwise)	-0.294	-2.081
α (dispersion coefficient)	0.282	3.078
Number of observations	464	
Log-likelihood at zero	-845.42	
Log-likelihood at convergence	-610.79	
ρ^2	0.28	

Table 9. Poisson Estimation Results Monthly Section "Overturn" Accident Frequency.

Variable	Estimated Coefficient	t-statistic
Constant	-3.288	-7.961
Average spacing of horizontal curves in section (meters)	0.00784	4.636
Lowest horizontal curve radius in section (meters)	-0.00461	-2.857
Maximum rainfall indicator (1 if greater than 2.54 centimeters on any given day in the month, 0 otherwise)	0.692	3.030
Rainfall-Curve interaction indicator (1 if maximum rainfall greater than 2.54 centimeters on any given day in the month and at least one horizontal curve has a design speed between 96.5 kph and 128.7 kilometers per hour, 0 otherwise)	-0.727	-2.952
Number of snowy days in the month	0.039	2.264
Snowfall-Curve interaction indicator (1 if maximum snowfall greater than 5.1 centimeters on any given day in the month and at least one horizontal curve has a design speed between 96.5 kph and 128.7 kilometers per hour, 0 otherwise)	0.970	4.771
Section location indicator (1 if section number is 1, 2, 3 or 4, 0 otherwise)	2.260	4.119
Year of occurrence indicator (1 if 1988, 0 otherwise)	0.465	2.868
Number of observations	464	
Log-likelihood at zero	-509.17	
Log-likelihood at convergence	-368.75	
ρ^2	0.28	

Table 10. Negative Binomial Estimation Results Monthly Section "Same Direction (All Others)" Accident Frequency.

Variable	Estimated Coefficient	t-statistic
Constant	-4.007	-7.819
Number of horizontal curves designed between 96.5 kph and 128.7 kph	0.471	3.381
Maximum grade in section	0.344	2.939
Rainfall-Curve interaction indicator (1 if maximum rainfall greater than 2.54 centimeters on any given day in the month and at least one horizontal curve has a design speed less than 96.5 kph, 0 otherwise)	0.787	3.857
Maximum daily snowfall (1 if greater than 5.1 centimeters on any given day in the month)	2.923	7.128
Snowfall-Grade interaction indicator (1 if maximum snowfall greater than 5.1 centimeters on any given day in the month and grade greater than 2%, 0 otherwise)	-0.901	-2.218
Snowfall-Curve interaction indicator (1 if maximum snowfall greater than 5.1 centimeters on any given day in the month and at least one horizontal curve has a design speed between 96.5 kph and 128.7 kph, 0 otherwise)	-1.232	-3.338
Spring/Summer month indicator (1 if April, May, June, July or August, 0 otherwise)	-0.805	-2.881
Year of occurrence indicator (1 if 1988, 0 otherwise)	0.577	2.253
Year of occurrence indicator (1 if 1989, 0 otherwise)	0.412	1.647
α (dispersion coefficient)	0.562	2.524
Number of observations	464	
Log-likelihood at zero	-540.84	
Log-likelihood at convergence	-307.04	
ρ^2	0.43	

regressions except the overturn accident frequency model which was a Poisson regression (i.e., statistically the overturn data were not overdispersed).

Interpretations of the results shown in Tables 5 through 10 are presented below. These interpretations are presented by type of variables.

Variable: Number of horizontal curves designed below 96.5 kph

Finding: Tendency to increase sideswipe and rear-end collisions but decrease parked vehicle as well as fixed object collisions

This finding suggests that drivers tend to slow down on underdesigned curves with design speeds less than 96.5 kph because of signing and visual perception and thus avoid severe accidents such as fixed object collisions. However, this possible speed reduction does not appear to be enough to avoid sideswipe or rear-end accidents caused by lane violations or speed differentials due to braking on the curve. In addition, it must be noted that parked-vehicle collisions tend to decrease because drivers are likely to avoid parking on the shoulder (for purposes such as chaining) on a tight curve because they perceive greater safety hazards at such locations.

Variable: Number of horizontal curves designed between 96.5 kph and 128.7 kph

Finding: Tendency to increase same direction (all others) and fixed object collisions

This finding suggests that curves underdesigned between 96.5 kph and 128.7 kph do not create the visual impact on drivers to decrease speeds. The result is an increase in both lane violations (resulting in vehicular collisions) and vehicles running off the roadway and colliding with fixed objects. From a severity viewpoint, fixed object collisions are more likely to result in serious injuries than vehicular collisions in the same direction. Consideration should then be given to upgrading marginally underdesigned curves (96.5 kph to 128.7 kph) if fixed object collisions show increasing trends at certain locations.

Variable: Number of horizontal curves in section

Finding: Tendency to increase same direction (all others) and fixed object collisions

This finding suggests two separate phenomena. Vehicular collisions in the same direction tend to increase on sections as the number of horizontal curves increase because speeds on curves do not decrease enough to avoid lane violations. The fact that fixed object collisions tend to increase with the total number of curves in a section indicates the increased likelihood of fixed objects, such as guardrails, being present on sections with more curves. The presence of such objects prevents a more severe type of accident, such as a vehicle overturn, from occurring. The caveat stemming from this finding is that it is preferable to design longer but fewer horizontal curves where the terrain makes construction of straight sections impossible.

Variable: Average spacing of horizontal curves in section

Finding: Tendency to increase overturn collisions

This finding uncovers an effect of roadway geometrics that was not distinguishable in the overall accident frequency model. The very significant t-statistic (4.636) indicates the significant effect that spacing of horizontal curves in a section has on driver speeds. Intuitively, if curves are spaced farther apart, vehicular speeds are likely to climb as a result of lower caution being exhibited by drivers. Consequently, there is a greater risk of an overturn if curves are spaced farther apart in a section. Careful attention should be paid to the application of corrective measures in this regard. The obvious interpretation is to decrease the spacing of curves to decrease the frequency of overturn accidents; however, it would appear counterintuitive to physically locate curves nearer as a countermeasure. The surrogate action is to place more advance warning signs in sections with longer curve spacing. By placing more advance warning signs and strategically locating them, the spacing of curves in the driver's mind is subliminally altered.

Variable: Lowest horizontal curve radius in section

Finding: Tendency to increase sideswipe collisions and decrease overturn collisions

The lowest horizontal radius suggests the type of terrain the section is located in. Sections with higher minimum radii could lull the driver into lane violations that result in sideswipes. However, low radii curves are usually associated with winding sections of highway which decrease the likelihood of gathering sufficient speed for an overturn accident.

Variable: Maximum grade in section

Finding: Tendency to increase rear-end and same direction (all others) collisions

This finding suggests several processes stemming from the presence of grades in a section. Between any two sections, the section with the steeper maximum upgrade will experience a greater number of rear-end and other same direction accidents. In addition, rear-end accidents will increase substantially if the maximum grade exceeds 2% in that section. Both effects are explained by speed differentials occurring due to the impact of grades. The impact of grades is reversed in the presence of downgrades. Between any two sections, the section with the steeper maximum downgrade will experience fewer rear-end and other same direction accidents, presumably from lower speed differentials. Much of the effect of the higher braking distance on downgrades appears to be offset by the visual impact of brake lights warning drivers of the potential slowing of vehicles ahead. In contrast, drivers are unlikely to use brakes on an upgrade which eliminates a critical warning sign of speed reductions.

Variable: Maximum rainfall on any given day in the month

Finding: Tendency to increase sideswipe, parked vehicle, fixed object and overturn collisions and decrease rear-end collisions

This result reflects some interesting phenomena affecting driver behavior and the driving task. Accidents resulting from the loss of steering control, such as lane violations and running-off-the-roadway, are expected to increase in occurrence with increases in maximum daily rainfall. Maximum rainfall indicates the intensity of rainfall and how that results in water puddles forming in wheel ruts in the pavement. The presence of such puddles contributes to vehicle hydroplaning and also excessive lateral drag resulting in lane violations and off-roadway accidents. On the other hand, as intensity of rainfall increases visibility decreases and drivers maintain greater headways paying more attention to the driving task. Much of this attention is focused on the vehicle ahead and quite likely much less is paid to the area of peripheral vision. This overcompensation on vehicle headways reduces rear-end accident risks but increases other accident types.

Variable: Average daily rainfall in the month

Finding: Tendency to increase rear-end collisions

This likely is an outgrowth of a seasonal effect that is descriptive of pavement condition. As opposed to maximum rainfall on any given day in the month, this variable captures the loss of traction due to wet pavements. An increase in average daily rainfall is indicative of a more prolonged wet-month weather effect. Drivers are less likely to pay attention to prolonged effects as opposed to short-term effects such as thunderstorms. In addition, as mentioned in the discussion of the overall model, the Seattle area receives rainfall for a large portion of the year on an intermittent basis. Drivers in the region may be less likely to acknowledge the hazards of wet pavements.

Variable: Number of rainy days in the month

Finding: Tendency to decrease sideswipe and rear-end collisions and increase fixed object collisions

It is speculated that the findings relating to this variable can be partially attributed to drivers reducing their speed in the presence of other vehicles during rainy periods (so much so that some types of vehicle collisions actually decrease). The positive coefficient for fixed object collisions suggests that this possibility of cautious driving behavior during rainy conditions does not transfer to all driving situations

Variable: Maximum snowfall on any given day in the month

Finding: Tendency to increase rear-end, same direction(all others), parked vehicle, and fixed objects collisions

This finding illustrates a number of consequences associated with intensity of snowfall. Loss of traction, visibility, and obliteration of lane markings act individually or in combination to increase the likelihood of accident types such as rear-end and other collisions in the same direction as well as collisions with parked vehicles and fixed objects.

Variable: Number of snowy days in the month

Finding: Tendency to increase sideswipe, fixed object and overturn collisions

This variable could be capturing a number of effects. For example, the loss of traction, obliteration of lane markings, and seasonal trends in weather and temperature could all be reflected in the significance of this variable.

Variable: Snowfall-Grade Interaction

Finding: Tendency to decrease sideswipe, rear-end, other collisions in the same direction and parked vehicle collisions

The coefficients of the interaction between snowfall and grade on rear-end accidents, other vehicular collisions in the same direction and parked vehicle accidents (as shown in tables 6, 7 and 8) appear counterintuitive. However, a closer examination of the estimation results shown in tables 6, 7 and 8 indicates that the net effect of snowfall and grade is to increase the frequency of these accident types, a conclusion also drawn from the positive coefficient of the snowfall-grade interaction variable for sideswipe collisions. In order to illustrate the net positive effect of snowfall-grade interaction on rear-end collisions, it is observed in Table 6 that the coefficient of the maximum daily snowfall variable is 3.468 which implies that when the daily maximum in any given month exceeds 5.1 centimeters rear-end accident frequencies are expected to increase 32-fold ($e^{3.468}$). In the presence of a significant interaction with grade on the section, (i.e., when maximum grade in the section exceeds 2%) this 32-fold compounding effect is tempered to 4.5-fold ($e^{(3.468-1.964)}$) by the negative coefficient (-1.964) of the interaction between snowfall and grade. This is a very intuitive occurrence indicating that the compounding effect of snowfall is not as severe when grades exceed 2%, quite possibly due to the presence of climbing lanes on upgrades and driver caution on downgrades. Examination of the snowfall-grade coefficient in other relevant accident types indicates a similar pattern. Corrective action then appears to be the construction of climbing lanes in areas where snowfall intensity is severe (exceeding 5.1 centimeters a day) and grades exceed 2%.

Variable: Snowfall-Curve Interaction

Finding: Tendency to increase rear-end, other collisions in the same direction and overturn collisions

This finding is similar to those described previously for the interaction between snowfall and grade. The net effect of snowfall on curves is tempered in the presence of underdesigned curves (< 128.7 kph), presumably due to warning signs and driver caution. Corrective action to mitigate the impact of snowfall appears to be the installation of warning signs in advance of underdesigned curves advising drivers of poor traction and slower speeds.

Variable: Rainfall-Curve Interaction

Finding: Tendency to increase rear-end and other collisions in the same direction, and decrease fixed object and overturn collisions

It is likely that this variable is capturing complex interactions among roadway and geometric conditions and driver behavior. To be able to speculate further on the nature of these findings, additional data on other roadway types (e.g., non-freeways) is necessary. This would allow us to isolate the effect of rainfall-curve interactions by providing greater variance in the data.

Variable: Spring/Summer month indicator

Finding: Tendency to decrease same direction-all others, parked vehicle and fixed object collisions

This indicates primarily the effects of seasonal trends such as daylight duration, ambient temperature. These are important determinants of accidents such as vehicular collisions in the same direction and collisions with parked vehicles and fixed objects. The finding illustrates the impact of pavement conditions, such as black ice, as well as visibility. It should be noted that the "Spring/Summer" indicator was found to significantly decrease accident frequencies in spite of increased exposure due to the higher traffic volumes typically observed during the spring and summer months.

Variable: Year of occurrence indicator

Finding: Tendency to increase all accident types

This finding indicates that some unobserved effects (e.g., ice accumulation on the pavement, and within-day temperature variations) were more severe during the subject year than usual, thus tending to increase the likelihood of an accident.

Variable: Section location indicator

Finding: Tendency to increase rear-end and overturn collisions, but decrease fixed object collisions

Section location indicators capture unobserved factors attributable to specific locations within the corridor. Such unobserved factors could include visual distractions and other attributes of the highway section that are difficult to quantify.

In summary, note that the coefficient estimates presented in these tables show that there are significant differences in the magnitudes of the coefficient estimates (and in some cases the signs of the coefficient estimates) among different accident types. The results of these separate accident frequency models can be used in the same way as the overall accident frequency model. That is, to evaluate the effectiveness of highway design improvements and ITS systems in reducing specific types of accidents.

1.6: IMPLICATIONS OF FINDINGS

The proposed model accounts for plausible and intuitive geometric and weather-related factors that influence accident frequencies. Specifically, the model offers insight into the combined effect of weather and geometric elements through interaction variables. The employment of indicator-type interaction variables allows designers to determine thresholds of geometric variables, such as maximum grade, beyond which their interaction begins to significantly affect accident frequencies.

The findings of this chapter have significant implications for highway design standards. Current standards establish geometric design criteria on the basis of pavement-tire interactions on wet pavements. Our findings show that, in order to reduce accident likelihoods in areas that frequently experience adverse weather, the basis of establishing design criteria should be expanded beyond wet-pavements. Specifically, great effort should be expended to avoid steep grades and horizontal curves with low design speeds in areas with adverse weather. Intuitively, this seems obvious, but our model provides a method of quantifying the impacts of these geometric characteristics. For example, for our study area, eliminating all horizontal curves with a design speed less than 96.5 kph on a roadway section that experiences at least 5.1 cm of snowfall, one or more days in a month, can reduce the monthly accident frequency by 47.3% (see Table 4). Although our model results are site-specific, a more global application of our approach could serve as a basis for a cost-benefit analysis that could guide geometric design policy more effectively than the current wet-pavement approach[10].

---

[10] A more global application would be a negative binomial accident frequency model estimated with roadway sections that had widely varying geometric, traffic flow, and access control characteristics (as opposed to the comparatively homogeneous sections used in this study). Such an application would likely find that variables such as lane widths, shoulder widths, peak hour traffic volumes, daily traffic volumes, percentage of trucks, type of road indicators (i.e., urban freeway, rural arterial, etc.), and type of interchange/intersection indicators, play a role in the frequency of accident occurrence.

CHAPTER 2: STATISTICAL ANALYSIS OF ACCIDENT SEVERITY ON RURAL FREEWAYS

2.1: RESEARCH OBJECTIVE

In addressing the possible safety-related impacts of Intelligent Transportation Systems (ITS), understanding of processes affecting both the frequency and severity of accidents is necessary. Current methodological approaches are limited in their scope of statistical inference of factors affecting frequency and severity. The previous chapter sheds light on methodologies for examining the frequency portion of rural freeway corridor safety under ITS. This chapter addresses the need for a comprehensive mathematical formulation of accident severity processes.

2.2: PREVIOUS RESEARCH

Previous research on accident severity has been diverse and provided important methodological and behavioral insights. Several accident-severity studies conducted have examined particular severity types such as fatalities (Shibata and Fukuda 1994) or concentrated on crashes involving certain vehicle types such as trucks (Golob, Recker and Leonard 1987; Alassar 1988). Other studies have concentrated on enforcement issues and their impact on fatal vehicular crashes related to alcohol and seat belt use (see for example Evans 1986; 1990; Holubowycz, Kloeden and Mclean 1994). Such studies have placed a heavy emphasis on the impact of human factors in determining accident severity.

However, other elements, such as highway design and environmental conditions, while not receiving the extensive attention given to human factors, have also been recognized as important determinants of accident severity (see Mercer 1986, Massie et al. 1993). Overall, past research has provided important insights into the range of factors that influence accident severity.

From a methodological standpoint, a variety of approaches have been employed to study accident severity. Using logistic regression techniques, Jones and Whitfield (1988) modeled severity risk as a function of anthropometric measures, car mass, age of driver and restraint system use. Logistic regression was also employed in a study of driver fatalities to model the probability of fatalities conditioned on the occurrence of an accident (Lui et al. 1988). However, the study used a limited number of variables such as driver age, gender, impact points, vehicle crash severity, restraint system use and car mass. Other important aggravating factors such as inclement weather, location of accident (for example, whether the accident occurred on a curve, or off the road) were omitted. A more comprehensive analysis of accident severity using multinomial logit approaches is available in Shankar and Mannering 1996. Other studies have employed multivariate time-series approaches to successfully develop predictive models of accident severity (Lassarre 1986). Evans (1990) employed a double-pair comparison approach to examine how occupant characteristics affect fatality risk. Still other methodologies such as headway-based severity analysis (Glimm and Fenton 1980), bivariate probit analysis (Hutchinson 1986) and discriminant analysis (Shao 1987) have been used. The latter methodologies, especially the probit and discriminant analyses, allow the researcher to model severity in terms of thresholds. These threshold approaches are consistent with the general categorization of accident severity as being either property damage only, possible injury, evident injury, or disabling injury/fatality.

The present study attempts to extend the empirical and methodological contributions of previous work by developing a predictive model of accident severity that can be used to evaluate the safety-related impacts of ITS and other safety-related countermeasures. In so doing, we will address highway design and environmental issues, along with human factors, in a multivariate context using a nested-multinomial logit approach. The empirical focus of our work will be the rural section of interstate 90 in Washington State, which as discussed in the previous chapters will have an operational ITS in August 1996. This chapter proposes to study past accident severities on this highway in an effort to establish a basis from which the safety effectiveness of the forthcoming ITS can be evaluated. This work follows our previous effort on accident frequencies (Shankar, Mannering, and Barfield, 1995). In combination with models of accident frequencies, the severity models presented in this chapter will enable us to provide a complete assessment of the possible safety impacts of the forthcoming Interstate-90 ITS.

2.3: METHODOLOGY

We begin by developing a conditional model of accident severity (i.e., conditioned on the fact that an accident has occurred).[1] Severity of an accident is

---

[1] For a statistical model of the likelihood of an accident occurring, the reader is referred to our earlier work on accident frequencies (Shankar, Mannering, and Barfield 1995).

specified to be one of four discrete categories: 1) property damage only, 2) possible injury, 3) evident injury, and 4) disabling injury or fatality.[2] Given these four discrete categories, a statistical model that can be used to determine the probability of an accident having a specific severity level can be derived. We start the derivation with the following probability statement,

$$P_n(i) = P(S_{in} \geq S_{In}) \ \forall \ I \neq i \qquad (2.3.1)$$

where $P_n(i)$ is the probability that accident n is severity i, P denotes probability and S_{in} is a function of covariates that determine the likelihood of accident n being severity i (I is the set of possible severities). To estimate this probability, a function defining the severity likelihoods must be specified. We use a linear form such that,

$$S_{in} = \beta_i \mathbf{X}_n + \varepsilon_{in} \qquad (2.3.2)$$

where $\mathbf{X}_n$ is a vector of measurable characteristics that determine the severity (e.g., driver age, driver gender, highway design attributes, prevailing weather conditions, vehicle type, use of seat belts, and so on), β_i is a vector of estimable coefficients, and ε_{in} is an error

---

[2] The determination of this severity is made by the officer at the scene of the accident and reported on the Washington State accident report forms. Also note that accidents are classified based on the most severe consequence of the accident. For example, an accident resulting in both injury and death will be classified as a fatality accident. In addition, it must be noted that total number of classified accidents reported is less than or equal to the number of individual severities since, for example, an injury accident may result in more than one person being injured.

term that accounts for unobserved factors influencing accident severity. The term $\beta_i\mathbf{X}_n$ in this equation is the observable component of severity determination because the vector $\mathbf{X}_n$ contains measurable variables (e.g., highway design attributes at the location of accident n), and ε_{in} is the unobserved portion. Given equations 2.3.1 and 2.3.2, the following can be written,

$$P_n(i) = P(\beta_i\mathbf{X}_n + \varepsilon_{in} \geq \beta_I\mathbf{X}_n + \varepsilon_{In}) \ \forall\, I \neq i \qquad (2.3.3)$$

or,

$$P_n(i) = P(\beta_i\mathbf{X}_n - \beta_I\mathbf{X}_n \geq \varepsilon_{In} - \varepsilon_{in}) \ \forall\, I \neq i \qquad (2.3.4)$$

With equation 2.3.4, an estimable severity model can be derived by assuming a distributional form for the error term. A natural choice would be to assume that this error term is normally distributed. Such an assumption results in a probit model. However, probit models are computationally difficult to estimate (see Ben-Akiva and Lerman 1985). A more common approach for models of this type is to assume that ε_{in}'s are generalized extreme value (GEV) distributed.[3] The GEV assumption produces a closed

---

[3] Discriminant analysis is another alternative to the approach that we have selected to model accident severity (Shao 1987). However, several studies have shown (see for example Press and Wilson 1978), that logit-based modeling approaches (which include the GEV approach) are superior to discriminant analysis for classification primarily because of the violation of the assumption of normality of disturbances in discriminant analysis. Presence of non-normal variables such as qualitative variables (dummy variables) in classification studies causes such a violation. In the present study, several qualitative variables, as will be shown, play significant roles in the determination of accident severity.

form model that can be readily estimated using standard maximum likelihood methods. It can be shown (McFadden 1981) that the GEV assumption produces the simple multinomial logit model.

$$P_n(i) = \exp[\beta_i \mathbf{X}_n] \Big/ \sum_I \exp[\beta_I \mathbf{X}_n] \qquad (2.3.5)$$

where all variables are as previously defined and the vector β_i is estimable by standard maximum likelihood methods.

Unfortunately, the simple multinomial logit model presented in equation 2.3.5 can lead to serious specification problems because this particular form requires us to assume that the unobserved terms (ε_{in}'s) are independent from one severity type to another. This is not likely to be the case because some of the severity types are likely to share unobserved terms and thus be correlated. For example, property damage only and possible injury accidents may share unobservables such as internal injury or effects associated with lower-severity accidents. In the presence of shared unobservables, the logit formulation will erroneously estimate the coefficient vector and severity probabilities. To circumvent this problem, a more generalized form of the severity probabilities can be derived from the GEV distribution. This is referred to as a nested logit model and has the following form (see McFadden 1981).

$$P_n(i) = \exp[\beta_i \mathbf{X}_n + \Theta_i L_{in}] \Big/ \sum_I \exp[\beta_i \mathbf{X}_n + \Theta_i L_{In}] \qquad (2.3.6)$$

$$P_n(j|i) = \exp[\beta_{ji} \mathbf{X}_n] \Big/ \sum_J \exp[\beta_{Ji} \mathbf{X}_n] \qquad (2.3.7)$$

$$L_{in} = \ln\left[\sum_{J} \exp\left(\beta_{Ji} X_n\right)\right] \tag{2.3.8}$$

where $P_n(i)$ is the unconditional probability of accident n having severity i (e.g., evident injury), $P_n(j|i)$ is the probability of accident n having severity j conditioned on the severity being in severity category i (e.g., the probability of having property damage only or possible injury given that there was no evident injury), J is the conditional set of severity categories (conditioned on i) and I is the unconditional set of severity categories. L_{in} is the inclusive value (log sum) which is interpreted as the expected value of the attributes that determine severity probabilities in severity category i. Θ_i is an estimable coefficient which must have a value between zero and one to be consistent with the model derivation (see McFadden, 1981).

The structure of the nested logit model eliminates the adverse consequences of shared unobservables because logit models determine probabilities using the difference in functions defining severity (i.e., the S_{in}'s in equation 2.3.2). Thus when a logit nest contains only those severity levels that share unobserved effects, the unobserved effects will cancel in the differencing and thereby preserve the assumption of independence needed to derive the model. We will discuss estimation concerns relating to this model and show its suitability for analyzing accident severities in the model estimation section of this chapter. For further information on the derivation and application of nested logit models the reader is referred to Ben-Akiva and Lerman (1985), Train (1986) and Mannering and Winston (1985, 1991, 1995).

2.4: EMPIRICAL SETTING

Individual accident, weather, geometric, pavement surface and vehicle data used in the accident frequency analysis in chapter 1 and driver-related data from individual accidents served as the basis for the analysis. To reiterate the characteristics of the database, roadway geometric data based on 10 equal 6.1 kilometer sections (see chapter 1; Shankar, Mannering, and Barfield 1995) were used. Individual accident data included information on primary identified causes, most severe consequence of the accident, time of day of accident, accident location with respect to the traveled way (on or off the roadway, whether the accident occurred on a curve or straight section or a grade, roadway illumination information, types of roadside objects involved in collision, and accident type). Weather data included whether or not the accident occurred during rainy, snowy, or foggy conditions. The geometric data included (for the section of highway in which the accident occurred) radii of horizontal curves, vertical grades, number of horizontal and vertical curves per kilometer, percentage length of horizontal curves. Pavement surface data included information on whether the accident occurred on icy, snowy, wet or dry pavement. Vehicle data included information on number and type of vehicles, restraint system[4] used by driver and occupants at the time of the accident, ejection status

---

[4] Although this information is subject to bias based on when the reporting officer arrives at the scene, uncertainty about restraint system use significantly diminishes in the case of injury-related accidents in which subjects are incapacitated to the extent of being unable to remove their restraint systems. In the case of property damage and possible injury accidents, the significance of restraint system use is minimal. In this context, it must be noted that uncertainty about restraint system use generally results in information on restraint system use being coded "restraint system use not known".

of occupants (i.e., whether or not occupants have been ejected from the vehicle) and number of occupants in each vehicle. Driver-related data included information on driver sobriety at the time of accident, and driver ages and gender.

A total of 1,505 individual vehicular accidents[5] were used in this study, with 1,020 of those accidents resulting in property damage only.[6] Out of the remaining 485 accidents, 10 were fatality collisions, 63 evident injury and 208 disabling injury collisions. Table 11 provides additional information on the distribution of severity by important variables such as daytime/night, sobriety, accident location (horizontal curve as opposed to a straight section), and number of vehicles involved in the collision. As shown in the table, factors such as drunk driving appear to play a role in higher-severity accidents judging by the relative frequencies in the severity categories. Examining the contemporaneous

---

[5] As mentioned in chapter 1, a total of 2,225 individual accidents were reported for the 65-month period between January 1988 and May 1993. However, weather data corresponding to 720 accidents was not available because of equipment failure or faulty operation. Also, it is important to note that our data were obtained only from reported accidents. It is likely that many accidents (particularly those that are minor in severity) may go unreported. This means that our accident sample is not a random sample of all accidents. Fortunately this will have a minimal impact on model estimation results. In fact, all coefficients will be correctly estimated with the exception of the constant terms. If the number and severity of unreported accidents were known, the three constant terms reported in this chapter could be adjusted by a simple calculation and no additional estimation would be necessary (see Ben-Akiva and Lerman (1985) for details on such stratified-sample adjustments).

[6] As mentioned previously, accident classification is based on the most severe consequence of the accident.

impact of geometric, temporal and environmental variables on accident severity in a multi-variate context is the major objective. However, to gain insights into the human factors dimension of accidents, this research will also examine how the effects of risk-taking behavior such as drunk driving are aggravated when interactions that compound their influence come into play, (for example interaction of curves with sobriety). Such interactions if significant will offer roadway engineers insights in to how driving environments can be designed to minimize the impacts of risk-taking behavior.

Table 11. Accident severity distribution by key variables.

	Severity Frequency				
Accident Conditioning Variable	Property Damage	Possible Injury	Evident Injury	Disabling Injury	Fatality
Daylight (excluding dawn & dusk)	609	135	31	126	6
Night	353	53	27	64	3
Drunk-Driving	31	1	2	9	2
Sober Driving	989	203	61	199	8
Horizontal Curve	410	76	25	88	8
Straight Section	610	128	38	120	2
Single-vehicle collision	587	99	44	128	5
Two-vehicle Collision	377	91	16	67	4
Multi-vehicle Collision (greater than two vehicles)	56	14	3	13	1

2.5: MODEL ESTIMATION

To estimate the nested logit model specified in equations 2.3.6, 2.3.7, and 2.3.8, we use a sequential estimation procedure. In this procedure, the lower conditional level of the nest (equation 2.3.7) is estimated as a simple multinomial logit (MNL) model using standard maximum likelihood methods and the estimated coefficients are used to compute the inclusive value of that level (i.e., L_{in} in equation 2.3.8). The next step involves estimating the higher level nest treating it as a simple MNL form but conditioning it on the estimated coefficients of the lower nest. This is done by introducing the computed value of L_{in} for the lower nest as an explanatory variable. All possible nested structures (which examine possible correlation among the unobserved effects of various severity levels) were considered. Statistically, as measured by likelihood ratio tests, the structure shown in Figure 3 proved to be the correct model form.[7] This nesting indicates that the property damage only and possible injury severity levels shared unobserved terms that would have caused a serious model specification error had a simple multinomial logit model been estimated (as shown in equation 2.3.5).

Maximum likelihood estimation results are presented in Tables 12 and 13. Table 12 presents the estimation of the lower nest (property damage only and possible injury)[8]

---

[7] We also tested this specification for possible correlation among unobservables using the specification test developed by Small and Hsiao (1985). The tests showed that this specification does not have statistically significant specification error.

[8] Possible injury accidents (which may seem a somewhat vague category) are determined at the scene by Washington State troopers using well-defined, uniformly taught

and Table 13 shows the estimation of the overall model of accident severity (upper nest). The inclusive value coefficient of 0.4153 with its t-statistic of 2.6391 suggests that shared unobservables significantly present between property damage only and possible injury alternatives.[9] Both models resulted in good statistical fits,[10] with the lower level of the nest showing a ρ^2 of 0.39 and the overall model a ρ^2 of 0.52.

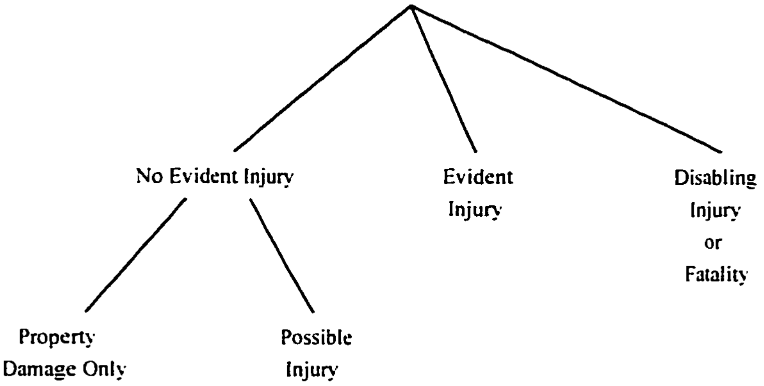

Figure 3. Nested Structure of Accident Severities.

---

identification procedures. Our testing of various model structures suggests that this is a unique severity category and must be considered separately (i.e., even though the accident will eventually be classified as an injury or property damage only accident).

[9] If no correlation between these unobserved terms was present, the coefficient value would not be significantly different from one. When the coefficient value of the inclusive value term is equal to one, the nested logit structure reduces to the simple multinomial structure as shown in equation 5.

[10] ρ^2 is defined as $1-[L(\beta)-L(0)]$ where $L(\beta)$ is log-likelihood at convergence and $L(0)$ is initial log-likelihood when all parameters are set to zero. A modified form of ρ^2 is the

Table 12. Estimation of property damage and possible injury probabilities conditioned on the occurrence of a non-injury accident.

Variable	Estimated Coefficient	t-statistic
Constant (specific to property damage alone)	3.1950	5.19
Overturn accident indicator (1 if accident type was "single-vehicle overturn", 0 otherwise; specific to possible injury)	1.3993	3.49
Rear-end accident indicator 1 (1 if accident type was "rear-end" accident and occurred on wet pavement, 0 otherwise; specific to possible injury)	0.4351	1.00
Rear-end accident indicator 2 (1 if accident type was "rear-end" accident and involved exactly two vehicles, 0 otherwise; specific to possible injury)	1.3415	5.00
Percentage of horizontal curve length per kilometer of roadway (specific to possible injury)	0.0141	1.37
Number of horizontal curves per kilometer of roadway (specific to possible injury)	0.4931	1.94
Illumination indicator (1 if surroundings were dark with no street lights present, 0 otherwise; specific to property damage)	0.3271	1.72
Sideswipe accident indicator (1 if accident type was "sideswipe" involving more than two vehicles, 0 otherwise; specific to possible injury)	1.2686	1.74

---

adjusted ρ^2 that takes into account the number of parameters included and is given by $1-[(L(\beta)-K)/L(0)]$, where K is the number of parameters. The adjusted ρ^2 for the two models were determined to be 0.38 and 0.50 respectively.

Table 12. (continued). Estimation of property damage and possible injury probabilities conditioned on the occurrence of a non-injury accident.

Variable	Estimated Coefficient	t-statistic
Same-direction accident indicator (1 if accident type was "same-direction" involving more than two vehicles, 0 otherwise; specific to possible injury)	1.0640	2.07
Fixed object accident indicator (1 if accident type was "fixed object", 0 otherwise; specific to possible injury)	1.0597	3.14
Icy pavement indicator (1 if accident occurred on icy pavement and involved only one vehicle, 0 otherwise; specific to property damage alone)	0.5323	2.11
Single-vehicle collision indicator (1 if accident involved one vehicle, 0 otherwise; specific to property damage)	0.6490	1.91
Number of observations	1224	
Log-likelihood at zero	-848.41	
Log-likelihood at convergence	-518.40	
ρ^2	0.39	

Table 13. Estimation of overall nested logit model of accident severity probabilities.

Variable	Estimated Coefficient	t-statistic
Constant (specific to evident injury)	-2.8468	-3.53
Constant (specific to disabling injury/fatality)	-2.4882	-4.78
Angle accident type indicator (1 if accident type is "angle", 0 otherwise; specific to evident injury and disabling injury/fatality)	1.5813	1.97
Overturn accident type indicator (1 if accident type is "overturn", 0 otherwise; specific to disabling injury/fatality)	0.5192	2.24
Speeding indicator 1 (1 if "exceeding posted speed" was primary cause, 0 otherwise; specific to evident injury)	0.9640	1.72
Speeding indicator 2 (1 if "exceeding reasonable safe speed for conditions" was primary cause, 0 otherwise; specific to evident injury)	-0.8855	-2.57
Speeding indicator 3 (1 if "exceeding reasonable safe speed for conditions" was primary cause, 0 otherwise; specific to disabling injury/fatality)	-0.3160	-1.69
Restraint system use indicator (1 if a restraint system was not in use by at least one driver involved in collision, 0 otherwise; specific to evident injury and disabling injury/fatality)	0.6376	2.72
Occupant ejection indicator (1 if any occupant was partially or totally ejected, 0 otherwise; specific to evident injury)	2.0070	3.78
Gender of driver (1 if all drivers involved in collision were male, 0 otherwise; specific to disabling injury/fatality)	1.0008	2.12

Table 13. (continued). Estimation of overall nested logit model of accident severity probabilities.

Variable	Estimated Coefficient	t-statistic
Percentage of horizontal curves per kilometer of roadway (specific to evident injury and disabling injury/fatality)	0.0302	3.39
Number of horizontal curves per kilometer of roadway (specific to no evident injury and disabling injury/fatality)	0.7204	1.93
Curve-sobriety interaction (1 if accident occurred on a horizontal curve and at least one driver involved was identified as "had been drinking and alcohol-impaired," 0 otherwise; specific to disabling injury/fatality)	1.2755	2.31
Snow-covered pavement indicator 1 (1 if accident occurred on snow-covered pavement, 0 otherwise; specific to evident injury)	-0.9450	-2.51
Snow-covered pavement indicator 2 (1 if accident occurred on snow-covered pavement, 0 otherwise; specific to disabling injury/fatality)	-0.5310	-2.86
Vehicle-mass difference indicator (1 if accident involved collision of a single truck and a single passenger car, 0 otherwise; specific to evident injury and disabling injury/fatality)	0.5214	1.83
Accident location indicator 1 (1 if accident occurred off the road, 0 otherwise; specific to evident injury)	1.2054	3.81
Accident location indicator 2 (1 if accident occurred off the road, 0 otherwise; specific to disabling injury/fatality)	0.5118	2.54
Age-sobriety interaction (1 if all drivers involved in accident were older than 55 years and at least one driver involved was identified as "had been drinking and alcohol-impaired," 0 otherwise; specific to evident injury)	1.6541	1.34

Table 13. (continued). Estimation of overall nested logit model of accident severity probabilities.

Variable	Estimated Coefficient	t-statistic
Night-time-pavement interaction (1 if accident occurred at night and on icy pavement, 0 otherwise; specific to evident injury and disabling injury/fatality)	0.2475	1.00
Fixed-object-horizontal curve interaction (1 if accident type was "fixed-object" and occurred on a horizontal curve, 0 otherwise; specific to disabling injury/fatality)	0.4580	1.99
Fixed-object-icy pavement interaction (1 if accident type was "fixed object" and occurred on icy pavement, 0 otherwise; specific to no evident injury)	0.5606	2.21
Inclusive value of property damage and possible injury (L_{in}, specific to no evident injury)	0.4153	2.64
Number of observations	1505	
Log-likelihood at zero	-1653.4	
Log-likelihood at convergence	-802.3	
ρ^2	0.52	

Turning first to the coefficient estimates of the lower nest (i.e., property damage only and possible injury conditioned on the accident having no evident injuries) we find that all variable coefficients included in the specification are statistically significant and have plausible signs. The implications of each of the coefficient estimates is discussed below.

Variable: Overturn accident indicator

Finding: Greater probability of possible injury relative to property damage only

The "single-vehicle" overturn accident indicator's positive coefficient indicates a greater likelihood of possible injury than property damage only. This shows that single-vehicle accidents with no evident injury tend to be more severe in nature.

Variable: Wet-pavement rear-end accident indicator

Finding: Greater probability of possible injury relative to property damage only

This variable captures the effect of rear-end accidents occurring in rainy weather. Such weather conditions make vehicles in front more difficult to see and increase the distance required to stop. It may also be argued that inclement weather may lower driver speeds and reduce risk of possible injury to a statistically insignificant level. However. intermittent and light rainfall. in spite of making the pavement wet and slippery. may not be dense enough to significantly lower driver speeds. The rear-end accident indicator may be capturing the effect of higher-than-expected vehicle speeds at the time of impact.

Variable: Two-vehicle rear-end accident indicator

Finding: Greater probability of possible injury relative to property damage only

While the previous finding reflects rear-end accidents in general. this variable captures the effect of two-vehicle collisions only. This coefficient is highly significant. statistically. indicating that injury. though not evident such as disabling. may be internalized to a greater extent than previously thought in such collisions. It is speculated that one important factor relating to the high significance of this variable could be the dissipation of kinetic energy and momentum per vehicle. The lower the number of vehicles involved. the greater the impact on each vehicle. thus increasing the likelihood of internal injuries, such as whiplash, which would be coded at the scene of the accident as a possible injury . The significantly higher coefficient (1.415 versus 0.4351 for the previous variable) corroborates the effect of inclement weather on driver speeds.

Variable: Percentage of horizontal curve length per kilometer of roadway

Finding: Greater probability of possible injury relative to property damage only

This variable captures the effect of terrain on the severity of an accident with no evident injuries. The high proportion of horizontal curves was found to increase the likelihood of a possible injury accident.

Variable: Number of horizontal curves per kilometer of roadway

Finding: Greater probability of possible injury relative to property damage only

This variable further confirms the finding offered by the curve-length variable. A greater number of curves on a particular section of roadway, although in some cases a speeding-deterrent, will affect steering control and reduce sight distance and thus be more likely to result in a possible injury collision.

Variable: Illumination indicator

Finding: Night-time conditions with no street lights present increase the probability of property damage only

This variable is likely an artifact of roadway design practices. Since the most dangerous portion of the road are the likely to be illuminated, we would expect a positive correlation between the absence of illumination and the likelihood of a property damage only accident.

Variable: Sideswipe accident indicator

Finding: Greater probability of possible injury relative to property damage only in multi-vehicle accidents

This variable (sideswipes involving more than two vehicles) primarily captures the exposure to possible injury. If the number of vehicles involved in a sideswipe accident exceeds two, the exposure increases in terms of number of occupants involved in the accident. Thus the greater likelihood of a possible injury. This variable may also be capturing the level of severity generally associated with this type of accident.

Variable: Same-direction accident indicator

Finding: Greater probability of possible injury relative to property damage only in multi-vehicle accidents

This variable (same direction accidents involving more than two vehicles) further illustrates the exposure, in terms of the number of occupants likely to be involved in the accident, that was also attributed to the sideswipe accident indicator. The finding is consistent with previous findings on the relationship of possible injury to increased exposure.

Variable: Fixed-object accident indicator

Finding: Greater probability of possible injury relative to property damage only

This variable is consistent with intuition which suggests that given that an accident resulted in no evident injuries, there is a greater probability of suffering possible injury from collisions with fixed objects. It must be noted that this applies only to accidents resulting no evident injuries.

Variable: Icy pavement indicator

Finding: Greater probability of property damage only relative to possible injury

This finding suggests that for single-vehicle accidents that occur on icy pavements, property damage will occur with greater probability than possible injury. This finding is consistent with previous conclusions on exposure in terms of number of vehicles involved and illustrates the effect of icy conditions. While icy pavement conditions hinder braking and steering control, they also tend to lower vehicle speeds. This effect reduces the risk of possible injury and limits the severity of an accident to property damage only.

Variable: Single-vehicle collision indicator

Finding: Greater probability of property damage only relative to possible injury

This finding corroborates earlier observations that fewer involved-vehicles increase the likelihood of property damage only. It also provides an important severity measure for accidents involving only one vehicle.

We now turn our attention to the estimation results of the overall model as presented in Table 13. The interpretation of coefficient estimates is provided below.

Variable: Angle accident type indicator

Finding: Greater probability of evident injury or disabling injury/fatality than no evident injury[11]

---

[11] No evident injury accidents include property damage only and possible injury where possible injury is typically a minor injury that is not evident at the scene of the accident.

In a freeway corridor, angle accidents can occur when a leading vehicle is turned sideways positioning it at angle to the flow of following traffic, thereby making severe collisions more likely. Angle accident indicators may also be acting as surrogates for factors such as black ice which are not strictly observed due to weather data limitations.

Variable: Overturn accident indicator

Finding: Greater probability of evident injury or disabling injury/fatality

This finding corroborates the finding documented for "single-vehicle" overturn effects in the lower level model. After correcting for single-vehicle effects which are incorporated in the no evident injury category, we observe that overturns result in a greater probability of evident injury or disabling injury/fatality.

Variable: Speeding indicator 1 (exceeding posted speed limit)

Finding: Greater probability of evident injury relative to no evident injury or disabling injury/fatality

This finding isolates the effect of speeding over posted speed limits on accident severity. Current knowledge and intuition suggest that speeding is a primary cause in severe accidents such as those resulting in disabling injury/fatality. However, there are associated factors such as number of curves in a section, sobriety and age which confound the effects of speed. Controlling for these factors may uncover specific effects of speed in isolation. In the present model, we control for all such factors (as discussed below) and isolate the effects of speed.

Variable: Speeding indicators 2 and 3 (exceeding safe speed for prevailing conditions)

Finding: Greater probability of no evident injury relative to evident injury or disabling injury/fatality

This finding illustrates an important distinction in the effects of high and low speeds in that it examines the impact of low speeds on severity. By examining speed-related effects in accidents where exceeding the posted limit was the primary cause. we essentially restrict the population of accidents related to speed to above the speed limit (104 kilometers per hour). The variable under discussion examines the effect of speeds over the range of possible speeds below the speed limit.[12] As mentioned previously, several factors interact in association with speed and aggravate its effect. For speeds under the posted limit but exceeding reasonable speeds for prevailing conditions, aggravating factors typically include weather-related variables such as pavement surface conditions, age, and grade or curve-related factors. As discussed in a later section, we control for these factors and isolate the effect of exceeding safe speeds for prevailing conditions. Isolating the effect of safe speeds indicates that at lower speeds, it is more likely that the accident will have no evident injury. This finding[13] illustrates the

---

[12] It must be noted that once speeds exceed the posted limit, speeding indicator 1 overrides speeding indicator 2 as the primary cause from a reporting perspective. Hence, speeding indicators 1 and 2 split the accident population into "above speed limit" and "below speed limit" sub-populations. This segmentation provides unique insights into the impacts of speeds because the effects of these two speed categories are quite different.

[13] The parameters for safe speed were specified initially for the evident injury and disabling injury/fatality alternatives simultaneously. By so doing, we constrain the β's to be the same for both alternatives. We removed this constraint and specified the β's separately for the alternatives. Relaxing the constraint allowed us to conclude the impact of safe speed with respect with evident injury was statistically different from that associated with disabling injury/fatality.

importance of a more comprehensive model specification for providing better insights into underlying processes.

Variable: Restraint system use indicator

Finding: Greater probability of evident injury or disabling injury/fatality relative to no evident injury if at least one driver did not use a restraint system at the time of the accident

This findings is in agreement with other studies (Evans 1986). An interesting observation was that separating the restraint system used by driver and passengers did not yield significantly different coefficients for passengers.[14]

Variable: Occupant ejection indicator

Finding: Greater probability of evident injury relative to no evident injury or disabling injury/fatality

This finding indicates that after controlling for factors such as overturn collisions or run-off-the road accidents. ejection of the occupant (partial or total) will result in a greater likelihood of evident bodily injury as opposed to death or disabling injury. This variable accounts. along with the restraint system use indicator. for factors such as structural integrity of the vehicle and door failures on impact.

---

[14] The statistically insignificant parameter for passenger restraint possibly indicates very high collinearity between driver and passenger restraint system use. In addition. when accident types such as rear-ends. angle and sideswipes are explicitly accounted for in the specification. rear-seat passenger injury is largely accounted for.

Variable: Gender of driver

Finding: Greater probability of disabling injury/fatality relative to no evident injury or evident injury if the accident involved all male drivers

This finding suggests that male drivers may be inherently greater risk takers and that risk is compounded by the exposure factor in multi-vehicle collisions when all drivers are male.

Variable: Percentage of curve length per kilometer of roadway

Finding: Greater probability of evident injury or disabling injury/fatality than no evident injury

This finding is consistent with the earlier finding on the same variable in the no evident injury model (as shown in Table 11). The finding implies that curve-length percentage increases the likelihood of an injury on a roadway section by possibly affecting the driving task and driver behavior.

Variable: Number of horizontal curves per kilometer of roadway

Finding: Greater probability of no evident injury or disabling injury/fatality relative to evident injury

This variable illustrates that evident injury is a less likely consequence as the number of curves per kilometer increases. This may be because some drivers' natural reaction is to slow-down when faced with many curves in close proximity, thus decreasing the likelihood of injury accidents.

Variable: Curve-sobriety interaction

Finding: Greater probability of disabling injury/fatality relative to no evident injury or evident injury

This variable captures the aggravating impact of curves on drunk driving. From a design perspective, this is an important finding because it presents opportunities for highway engineers to mitigate circumstances that aggravate drunk driving effects. A drunk driver's lack of control is particularly critical on horizontal curves resulting in lane violations and ensuing multi-vehicle collisions or severe run-off-the road impacts.

Variable: Snow-covered pavement indicators[15]

Finding: Greater probability of no evident injury relative to evident injury or disabling injury

This finding indicates the impact of seasonal[16] as well as location-specific effects on accident severity. The presence of snow on the pavement at the time of the accident may indicate a general caution observed by drivers. If an accident were to still occur, the greater caution exercised by drivers helps mitigate the severity of an accident by reducing the effect of aggravating factors such as speed. On the other hand, presence of snow may also capture the higher observed frequency of "parked vehicle" accidents (i.e., disabled

---

[15] As Table 3 shows, the β's for the evident injury and disabling injury/fatality categories were estimated unconstrained (i.e. separate coefficients for each severity category) and found to give statistically superior results relative to the constrained case (as measured by a likelihood ratio test).

[16] Seasonal effects capture, in addition to direct weather effects, factors such as traffic volume. Reduced traffic volumes during the months of November through March reduces the likelihood of multi-vehicle accidents. Indirectly, this accounts for exposure.

vehicles or those vehicles parked to put chains on) which tend to be property damage only. Lane obliteration may cause lane violations and ensuing collisions such as sideswipe and same direction accidents which were observed to be milder in severity.

Variable: Vehicle-mass difference indicator

Finding: Greater probability of evident injury or disabling injury/fatality relative to no evident injury

This variable captures the effect of truck-passenger car collisions on accident severity in two-vehicle accidents. By isolating two-vehicle collisions, we truly capture vehicle-mass difference effects, as opposed to a combination of vehicle-mass and exposure-related effects that would be present in multivehicle collisions involving more than two vehicles.

Variable: Accident location indicators

Finding: Greater probability of evident injury or disabling injury/fatality relative to no evident if the accident occurred off the road.

This variable captures the impacts of off the road accidents due to roadside features such as ditches and embankments. Such features tend to cause an injury. The findings indicate that the likelihood of evident injury is significantly greater than disabling injury/fatality in off-the-road collisions. Accounting explicitly in the specification for specific accident types such as overturns, which typically occur in off-the-road collisions, allows us to isolate the impact of the off-the-road coefficient for disabling injury/fatality severities.[17]

---

[17] Off-the-road coefficients for evident injury and disabling injury/fatality were estimated separately. By doing so, the parameter for disabling injury/fatality was determined to

Variable: Age-sobriety interaction

Finding: Greater likelihood of evident injury relative to disabling injury/fatality or no evident injury

This variable provides insight into an important two-way interaction that has not been investigated prior to this study. Age and sobriety have long been identified to play separate but significant roles in accident occurrences and severities. Several studies (Jonah 1986; Mayhew et al. 1986) have shown that older drivers are less prone to risk taking than younger drivers. Coupled with this, the risk of crash involvement of older drivers is also reduced due to greater driving experience. In addition, it has also been noted that alcohol-related impairment in driving is greater among older drivers. Given that an accident occurs, the combination of these factors results in injury accidents that are not as severe as disabling/fatality collisions. Being less likely to take risk and having greater driving experience seems to offset the greater impairment in driving that alcohol causes in older drivers, at least in terms of severity. However, the effect that such factors have on overall accident frequency is an open question that is not addressed in this study.

Variable: Night-time-pavement condition interaction

Finding: Greater likelihood of evident injury or disabling injury/fatality relative to no evident injury

This variable models the effect of night-time conditions and icy pavements on accident severities. In the event an accident occurred under such conditions, the positive

---

be significantly lower than that for evident injury indicating that run-off-the-road by themselves are more likely to cause evident injury than disabling injury/fatality. It is in the presence of collisions such as vehicle overturns that the likelihood of disabling injury/fatality is enhanced.

coefficient of this variable with respect to evident injury and disabling injury/fatality indicates the influence of temperature-related and seasonal factors on driving. The importance of this interaction term stems from the compounding effect that night-time conditions have on driver behavior under icy conditions. It may be argued that the propensity of accidents occurring in icy weather could be lower during night because of increased caution among drivers; however, given that an accident occurs, the severity is likely to be high. This higher severity may be impart caused by slower driver-reaction times which tend to be significantly slower at night.[18]

Variable: Fixed-object-horizontal curve interaction

Finding: Greater probability of disabling injury/fatality relative to evident injury or no evident injury

This interaction term accounts for the impact of roadside features on accident severities on horizontal curves. The finding underscores the importance of roadside design on horizontal curves.

Variable: Fixed-object-icy pavement interaction

Finding: Greater probability of no evident relative to evident injury or disabling injury/fatality

---

[18] The element of surprise and emergency response are important factors affecting driver reaction times. Several studies (Triggs and Harris 1982; Olson 1989; Hooper and McGee 1983; Taoka 1982) have evaluated driver reaction times under varying conditions and for different age groups, and concluded that night-time reaction times could be significantly higher than daytime values. The significance of the night-time-pavement interaction term presents a surrogate factor for reaction time, and illustrates the importance of potentially challenging highway geometrics.

This variable further corroborates, as mentioned previously, the impact of speed on the severity of fixed-object collisions. Icy weather acts as a deterrent to speeding, and as a result the consequence of fixed-object collisions are likely to be less severe. Again, this finding does not relate to the frequency of such collisions which could be expected to be higher under such conditions.

In addition to examining the impact of key variables on accident severity, elasticities of important design variables were also examined. Elasticity is the measure of the percentage change in the probability of a specific severity level for a unit percentage change in an independent variable. It is generally computed as a point-measure for continuous variables.[19] An elasticity greater than unity in absolute value indicates that the dependent variable is elastic with respect to the subject independent variable. The elasticities of overall accident severity probability with respect to curve-length percentage and number of horizontal curves per kilometer of roadway were computed to be -0.2704 and -0.9017 respectively. Intuitively this says that a 1 percent increase in the percentage of horizontal curve length per kilometer will result in a 0.2704 percent increase in the likelihood of an accident being evident injury or disabling injury/fatality. Also, a 1 percent increase in the number of horizontal curves per kilometer will result in a 0.9017 percent increase in the probability of the accident resulting in no evident injury or disabling injury/fatality. While both elasticities are less than 1, the elasticity computation

---

[19] Elasticities over a larger range of independent variable values are misleading when computed using this formula. In addition, elasticities for indicator variables which have binary values of 0 or 1 are meaningless.

provides interesting insight into the comparative importance of these two variables in determining accident severity.

2.6: IMPLICATIONS OF FINDINGS

By developing a probabilistic model that contains several important variables representing geometric, weather, and human factors we have shown that ambiguity and bias stemming from confounding effects in a partially specified model can be eliminated. In addition, this research provides suggestive results by its use of variables such as curve-sobriety interaction and curve-pavement surface interaction. Specifically, it suggests that ITS may be an effective means of compensating for adverse design, human factors, and weather conditions. A well designed ITS could significantly improve the driving task in the presence of adverse factors such as alcohol, inclement weather, and complex roadway geometrics. A significant shift in the distribution of accident severities toward milder accidents in combination with lower accident frequencies (Shankar, Mannering, and Barfield 1995) will provide a basis for ITS evaluation. It is also demonstrated through this inquiry that with appropriate distributional assumptions, intuitively appealing severity structures and associated underlying processes can be unraveled. Potential bias stemming from lack of appropriate distributional assumptions is thus avoided. The lack of adequate sample observations in the fatality category limits the structural possibilities that could arise otherwise. Disabling and fatal injuries may share unobservables related to time response effects, especially in rural areas. Large first-aid response times in rural areas can lead to the progression of a disabling injury to fatal consequences. It must therefore be remembered that while the structure presented in this inquiry appears promising, the often-time limitation on fatality samples may pose limitations on structural alternatives.

CHAPTER 3: MODELING THE ENDOGENEITY OF LANE-MEAN SPEED AND LANE-SPEED DEVIATION: A STRUCTURAL EQUATIONS APPROACH

3.1: RESEARCH OBJECTIVE

The previous chapters examined the likelihoods of accidents and provided insights into how design speeds affect those likelihoods. From a probabilistic standpoint, it was inferred that depending on the geometric characteristics, the impact of design speed varies. While intuition suggests that this variation affects traffic flow dynamics and the interaction between vehicles, formulation of explicit relationships remained. Effective mathematical formulation of such relationships providing insights in to the cross-sectional dynamics of traffic flow will allow roadway engineers to examine ways (such as ITS) to optimize roadway capacity while minimizing unsafe vehicle interactions.

Prior speed-flow relationship studies have focused on single-regime or multi-regime functional relationships that were generally univariate or bivariate in nature. Linkages between speed and flow were generally studied over different traffic density ranges. Engineering intuition suggests that such approaches offer only a limited understanding of the underlying processes governing speed-flow relationships. Particularly in the context of intelligent transportation systems (ITS) where the use of technological components will likely result in fundamental shifts of assumed speed-flow relationships. In the presence of ITS, it is important that the causality underlying the processes affecting traffic speed-flow relationships and consequently safety be uncovered, because systemic affects associated with such technologies are potentially wide-ranging and often simultaneous. This chapter attempts to macroscopically address endogeneity issues related to lane-mean traffic speeds and lane-speed deviations.

3.2: PREVIOUS RESEARCH

Prior theories and empirical validations have established speed-flow relationships that are unidirectional and regime-based (see for example, Greenshields 1935, Edie 1961, May and Keller 1968). Suggestions on structural modeling (i.e., a simultaneous equations approach), with its potential to provide an improved understanding of the interrelationships among the contemporaneous influences of lane-mean speeds, lane-speed deviations, environmental conditions, geometric elements, vehicle-types, and temporal and seasonal factors, have been conceptual for the most part. Instead, significant effort has been focused on the use of independent ordinary or non-linear least squares estimation (see for example Easa and May 1980). Use of independent regression equations that separately estimate speed and flow-related parameters without accounting for the contemporaneous correlation of the disturbances will cause the respective estimated parameters to be biased and inconsistent (Greene 1993). Apart from the specification aspects mentioned above relating to the causal modeling of traffic speed and flow, little evidence is available on modeling frameworks that simultaneously incorporate the influence of environmental, geometric, temporal and traffic-flow factors. Some efforts in this area have focused on the impact of weather (Ibrahim and Hall 1994) and geometrics (Iwasaki 1991), while others have focused on the temporal variations in traffic flow (see for example Brilon and Ponzlet 1996).

The attempt of this research is to combine the need for a complete model that is comprehensive in factors identified in previous research with the need for an estimation framework that is structural in nature. It should be noted here that the focus of the research is on the structural relationship between lane-mean speeds (i.e., time-mean speeds) and related lane-speed deviations and the traffic characteristics, environmental conditions, and temporal and seasonal factors. As such, the investigation will focus on the contemporaneous inter-relationships at a given location in a given time period.

3.3: METHODOLOGY

Our intent is to develop a model of mean speeds and speed deviations (measured over some time interval) for each lane of a multi-lane roadway. Turning first to lane-mean speeds, from a structural point of view, it is important to note that the mean speed in each lane will not only be a function of traffic characteristics in the lane, but also a function of the mean speeds in the adjacent lanes. This suggests an equation system in which lane-mean speeds are determined simultaneously across the roadway's lanes. In a similar fashion, speed deviations in each lane will be dependent on speed deviations in adjacent lanes. Lane-speed deviations will also be a function of the lane's mean speed and the mean speeds in adjacent lanes. Because of this interrelationship, lane-speed deviations must also be determined in a simultaneous equation system with mean speeds entering the equation system in a recursive fashion.

The structural equation system for lane-mean speeds and lane-speed deviations can be written as follows: For lane-mean speeds, over some time interval, the equation system is,

$$\begin{array}{l} u_1 = \alpha_1 + \beta_1 X_1 + \lambda_1 Z_1 + \theta_1 \overline{u_1} + \varepsilon_1 \\ u_2 = \alpha_2 + \beta_2 X_2 + \lambda_2 Z_2 + \theta_2 \overline{u_2} + \varepsilon_2 \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ u_n = \alpha_n + \beta_n X_n + \lambda_n Z_n + \theta_n \overline{u_n} + \varepsilon_n \end{array} \qquad (3.3.1)$$

where u_n is the mean speed in lane n, X_n is a vector of exogenous variables influencing the mean speed in lane n, Z_n is a vector of endogenous variables influencing the mean speed in lane n (i.e., traffic flow characteristics that may be influenced by lane-mean speeds such as proportion of total roadway traffic in the lane), $\overline{u_n}$ is a vector of mean speeds in lanes adjacent to lane n, α_n, β_n, λ_n, and θ_n are vectors of estimable coefficients, ε_n is a disturbance term. Similarly, lane-speed deviations, over some time interval, can be written as,[1]

$$\begin{array}{l} \sigma_1 = \rho_1 + \eta_1 V_1 + \tau_1 Y_1 + \gamma_1 \overline{u_1} + \omega_1 \overline{\sigma_1} + \nu_1 \\ \sigma_2 = \rho_2 + \eta_2 V_2 + \tau_2 Y_2 + \gamma_2 \overline{u_2} + \omega_2 \overline{\sigma_2} + \nu_2 \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \sigma_n = \rho_n + \eta_n V_n + \tau_n Y_n + \gamma_n \overline{u_n} + \omega_n \overline{\sigma_n} + \nu_n \end{array} \qquad (3.3.2)$$

where σ_n is the standard deviation of speed in lane n, V_n is a vector of exogenous variables influencing the standard deviation of speed in lane n, Y_n is a vector of endogenous variables influencing the standard deviation of speed in lane n (i.e., traffic

---

[1] Note that our equations model lane-speed deviations as dependent variables which are functions of lane-mean speeds. The reverse relationships between lane-mean speeds and lane-speed deviations is not specified. This is there is no basis for assuming lane-mean speeds are influenced by speed deviations. This was borne out during some preliminary estimation runs that found lane-speed deviations to be statistically insignificant when included in Equation 1.

flow characteristics that may be influenced by lane-speed deviations such as proportion of total roadway traffic in the lane), $\overline{u_n}$ is a vector of mean speeds in lane n and in other lanes, $\overline{\sigma_n}$ is a vector of the standard deviation of speeds in lanes adjacent to lane n, ρ_n, η_n, τ_n, γ_n, and ω_n are vectors of estimable coefficients, ν_n is a disturbance term.

To estimate equations 3.3.1 and 3.3.2, three-stage least squares (3SLS) is appropriate. This approach allows for simultaneous estimation of coefficients using information from the equation system. By so doing, it ensures that coefficient estimates are generally more efficient (asymptotically) than alternative simultaneous-equation estimation approaches such as the indirect least-squares (ILS), two-stage least squares (2SLS), and limited- information maximum likelihood (LIML).[2] An alternative estimation approach is full-information maximum likelihood (FIML), but because the asymptotic variance-covariance matrices of FIML and 3SLS can be shown to be equal, the choice of 3SLS is acceptable. The 3SLS estimation procedure is conducted by first getting two-stage least squares (2SLS) estimates of the equation system which are calculated using instruments (endogenous variables regressed against all exogenous variables). The 2SLS estimates are then used to estimate the equation system's disturbances which are subsequently used to estimate the contemporaneous variance-covariance matrix of disturbances. Finally, generalized least-squares (GLS) is applied to estimate model coefficients using the estimated contemporaneous variance-covariance

---

[2] The 3SLS procedure is more efficient than single-equation methods such as ILS, 2SLS, and LIML, when the variance-covariance matrix is not diagonal. This will be the case when there is contemporaneous correlation among disturbance (i.e., the unobserved factors affecting mean speed in one lane are correlated with those unobserved factors that affect mean speed in other lanes). If these unobserved factors are not correlated (i.e., the case of a diagonal variance-covariance matrix), it can be readily shown that 3SLS reduces to 2SLS.

matrix of disturbances as a basis. See Greene (1993) for a complete description of the procedure.

3.4: EMPIRICAL SETTING

To model lane-mean speeds and lane-speed deviations at this location, data were collected using magnetic loop detectors. Interstate 90, in the study corridor as mentioned previously, is a three-lane divided freeway in each direction with the eastbound alignment on a 1.5 percent upgrade and the westbound alignment on a 2.5 percent downgrade. Eastbound and westbound traffic data were collected by lane. Data on spot speeds by lane, vehicle classification by lane, were gathered in the fall of 1994 and the winter, spring and summer months of 1995. Speed data were collected in speed bins of 10 miles per hour, aggregated over one hour.[3] Classification of vehicle types was based on four wheelbase classes of up to 26, 26 to 39, 39 to 65, and 65 to 114 feet. Lane-by-lane data were collected for spot speeds and vehicle classifications in both eastbound and westbound directions. Table 14 shows computed lane-mean speeds and lane-speed deviations by lane using one hour time periods.

---

[3] Aggregation of speed data over one hour is likely to mask some underlying variation in the speed distribution; however, the level of detail that is afforded at micro-speed data such as 5-second or 20-second data is not likely to significantly alter the structure of the cause-effect relationship between speed and speed deviation. Any additional insight into the cause-effect relationship could stem from the stochasticity of peak hour flows. As will be demonstrated later, the stochasticity of peak hour flows and its impact on speed-speed deviation relationships will be captured adequately by indicator variables acting as surrogates for peak hour phenomena thus eliminating potential omitted variable biases. The authors do acknowledge that micro-speed data does provide insight into merge and weave phenomena and shock-wave-related incremental impacts on traffic flow continuums, but point out that the use of such data is different, namely to investigate "resulting conditions" stemming from inconsistencies in traffic flow.

Table 14. Summary of lane-mean speeds and lane-speed deviations.

		Hourly Grouped Speeds					
		Grouped Lane-Mean Speed (miles per hour)			Grouped Lane-Speed Deviation (miles per hour)		
Direction	Location	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Eastbound	Right Lane	70.193	31.250	76.760	7.164	4.440	16.150
	Middle Lane	75.612	32.580	79.820	5.548	3.780	13.860
	Left Lane	78.012	34.880	90.000	4.858	0.000	21.680
Westbound	Right Lane	72.986	40.470	79.430	7.000	4.580	15.640
	Middle Lane	76.441	43.570	81.940	5.756	3.000	14.210
	Left Lane	78.830	40.000	86.670	5.310	0.000	28.720

3.5: MODEL ESTIMATION

Tables 15 and 16 show the results of the 3SLS estimation of grouped lane-mean speeds at the study location. Tables 17 and 18 show the results of the 3SLS estimation of grouped lane-speed deviations. For estimation purposes, the logarithm of the lane-mean speed was used as the dependent variable in the lane-mean speed model system. As seen in the tables, exogenous variables significantly determining lane-mean speed and lane-speed deviation include time-of-day, time-of-week, and seasonal, indicators. Vehicle mix and the distribution of traffic across the lanes were also found to be significant determinants of lane-mean speed and lane-speed deviation.[4] All estimated coefficients were found to be of plausible sign. For the eastbound direction, the system R^2 for the lane-mean speed model was 0.8629 and 0.3288 for the lane-speed deviation model. For the westbound direction, system R^2 was 0.9232 and 0.3087 for the lane-mean speed and lane-speed deviation models, respectively. The interpretation of the estimation results is provided below. were found to be of plausible sign. For the eastbound direction, the system R^2 for the lane-mean speed model was 0.8629 and 0.3288 for the lane-speed deviation model. For the westbound direction, system R^2 was 0.9232 and 0.3087 for the lane-mean speed and lane-speed deviation models, respectively. The interpretation of the estimation results is provided in the following section.

---

[4] For estimation purposes these variables were instrumented (see description in Tables 2-5) because of possible endogeneity. This is because changes in lane-mean speeds and/or lane-speed deviations can affect the distribution of traffic flow over the lanes. Thus changing values in the dependent variable could change values in the independent variable, which is violation of least-squares assumptions. Not correcting for this will result in biased and inconsistent coefficient estimates.

Table 15. Three-stage least squares estimation of grouped lane-mean speeds for eastbound I-90.

Variable*	Estimated Coefficient	t-statistic
Equation 1: Logarithm of Right-Lane Mean Speed (Dependent Variable)		
Constant	-0.1106	-2.6503
Lane traffic flow indicator (1 if traffic flow in right lane is less than 75 vehicles per hour, 0 otherwise)	0.0021	2.7372
Truck percentage in right lane	-0.0292	-15.2267
High truck flow in right lane (1 if hourly truck flow is greater than 100 vehicles per hour, 0 otherwise)	0.0030	6.8196
Relative truck flow indicator 1 (1 if truck percentage in right lane exceeds 60% and total traffic flow in right lane is less than 50 vehicles per hour, 0 otherwise)	0.0047	5.8343
Relative truck flow indicator 2 (1 if truck percentage in right lane is less than or equal to 20% and total traffic flow in right lane exceeds 200 vehicles per hour, 0 otherwise)	0.0034	5.3791
Logarithm of middle-lane mean speed	1.0107	104.7271
Time-of-day indicator 1 (1 if hour of observation is between midnight and 6:00 AM, 0 otherwise)	-0.0030	-3.3621
Seasonal indicator 1 (1 if it is winter, 0 otherwise)	-0.0021	-3.5543
Seasonal indicator 2 (1 if it is spring, 0 otherwise)	-0.0010	-2.7644
Time-of-week indicator (1 if it is weekend, 0 otherwise)	0.0104	8.7886
Time-of-day indicator 2 (1 if it is PM peak hour, 0 otherwise)	0.0018	3.7352
Time-of-day indicator 3 (1 if it is AM peak hour, 0 otherwise)	-0.0014	-2.7911
Number of observations	2233	
R-squared	0.9072	
Corrected R-squared	0.9067	

* Variables in italics are instrumented because of possible endogeneity. This is done by regressing the variable against exogenous variables and using the regression-predicted values for the 3SLS estimation. Variables in bold are endogenous and part of the simultaneous equation estimation. Finally, trucks are defined as vehicles with wheelbases exceeding 65 feet.

Table 15 (continued). Three-stage least squares estimation of grouped lane-mean speeds for eastbound I-90.

Variable*	Estimated Coefficient	t-statistic
Equation 2: Logarithm of Middle-Lane Mean Speed (Dependent Variable)		
Constant	0.3628	12.0474
Logarithm of right-lane mean speed	0.4257	59.2642
Logarithm of left-lane mean speed	0.4960	80.9855
Hourly traffic flow in middle lane	-0.000014	-10.2548
Lane use distribution between middle lane and right lane (ratio of flows in middle lane to right lane)	-0.0010	-4.2564
Time-of-day indicator 4 (1 if it is night-time, 0 otherwise)	-0.0030	-3.5560
Time-of-week indicator (1 if it is weekend, 0 otherwise)	-0.0072	-13.2349
Number of observations	2233	
R-squared	0.9022	
Corrected R-squared	0.9019	
Equation 3: Logarithm of Left-Lane Mean Speed (Dependent Variable)		
Constant	-0.6949	-12.0579
Truck percentage in left lane	0.0057	2.1422
Lane distribution between left lane and middle lane (ratio of flows in middle lane to right lane)	0.0050	3.1882
Logarithm of middle-lane mean speed	1.1671	87.7616
Time-of-day indicator 4 (1 if it is night-time, 0 otherwise)	0.0050	3.0050
Number of observations	2233	
R-squared	0.7961	
Corrected R-squared	0.7958	
System R-squared	0.8629	

* Variables in italics are instrumented because of possible endogeneity. This is done by regressing the variable against exogenous variables and using the regression-predicted values for the 3SLS estimation. Variables in bold are endogenous and part of the simultaneous equation estimation. Trucks are defined as vehicles with wheelbases exceeding 65 feet.

Table 16. Three-stage least squares estimation of grouped lane-mean speeds for westbound I-90.

Variable*	Estimated Coefficient	t-statistic
Equation 1: Logarithm of Right-Lane Mean Speed (Dependent Variable)		
Constant	-0.4308	-14.3947
Truck percentage in right lane	-0.0144	-9.2982
High truck flow in right lane (1 if hourly truck flow is greater than 100 vehicles per hour, 0 otherwise)	0.0017	2.7649
Logarithm of middle-lane speed	1.0895	157.8630
Seasonal indicator 1 (1 if it is winter, 0 otherwise)	0.0012	2.5684
Time-of-week indicator 1 (1 if it is weekend, 0 otherwise)	0.0046	4.8303
Time-of-day indicator 3 (1 if it is AM peak hour, 0 otherwise)	-0.0013	-2.5191
Number of observations	2230	
R-squared	0.9472	
Corrected R-squared	0.9470	
Equation 2: Logarithm of Middle-Lane Mean Speed (Dependent Variable)		
Constant	0.1919	7.7056
Logarithm of right-lane mean speed	0.4539	62.1349
Logarithm of left-lane mean speed	0.5047	66.5657
Hourly traffic flow in middle lane	-0.000015	-9.5357
Lane use distribution between middle lane and right lane (ratio of flows in middle lane to right lane)	-0.0012	-4.4492
Time-of-day indicator 4 (1 if it is night-time, 0 otherwise)	-0.0036	-5.1740
Time-of-week indicator (1 if it is weekend, 0 otherwise)	-0.0030	-6.6528
Number of observations	2230	
R-squared	0.9454	
Corrected R-squared	0.9452	

* Variables in italics are instrumented because of possible endogeneity. This is done by regressing the variable against exogenous variables and using the regression-predicted values for the 3SLS estimation. Variables in bold are endogenous and part of the simultaneous equation estimation. Trucks are defined as vehicles with wheelbases exceeding 65 feet.

Table 16 (continued). Three-stage least squares estimation of grouped lane-mean speeds for westbound I-90.

Variable*	Estimated Coefficient	t-statistic
Equation 3: Logarithm of Left-Lane Mean Speed (Dependent Variable)		
Constant	-0.1134	-2.6376
Hourly traffic flow in left lane	0.000035	6.3908
Lane distribution between left lane and middle lane (ratio of flows in middle lane to right lane)	0.0040	2.6966
Logarithm of middle-lane mean speed	1.0321	104.1540
Time-of-day indicator 4 (1 if it is night-time, 0 otherwise)	0.0067	4.6959
Number of observations	2230	
R-squared	0.8797	
Corrected R-squared	0.8795	
System R-squared	0.9232	

* Variables in italics are instrumented because of possible endogeneity. This is done by regressing the variable against exogenous variables and using the regression-predicted values for the 3SLS estimation. Variables in bold are endogenous and part of the simultaneous equation estimation. Trucks are defined as vehicles with wheelbases exceeding 65 feet.

Table 17. Three-stage least squares estimation of grouped lane-speed deviations for eastbound I-90.

Variable*	Estimated Coefficient	t-statistic
Equation 1: Right-Lane Speed Deviation (Dependent Variable)		
Constant	34.5707	7.6989
Speed Deviation in middle lane	0.1996	2.5356
*Logarithm of right-lane mean speed***	3.6272	2.6788
*Logarithm of middle-lane mean speed***	-10.1006	-10.0808
Time-of-day indicator 1 (1 if hour of observation is between midnight and 6:00 AM, 0 otherwise)	0.2384	3.4834
Time-of-day indicator 2 (1 if it is PM peak hour, 0 otherwise)	-0.0917	-1.4435
Seasonal indicator 1 (1 if it is winter, 0 otherwise)	-0.2234	-4.3246
Time-of-week indicator (1 if it is weekend, 0 otherwise)	-0.2969	-5.8715
Truck-to-passenger car flow ratio	-0.1238	-7.8162
Number of observations	2233	
R-squared	0.2959	
Corrected R-squared	0.2934	

* Variables in italics are instrumented because of possible endogeneity. This is done by regressing the variable against exogenous variables and using the regression-predicted values for the 3SLS estimation. Variables in bold are endogenous and part of the simultaneous equation estimation. Trucks are defined as vehicles with wheelbases exceeding 65 feet.

** Lane-mean speeds are instrumented variables in the speed deviation system. Predicted values from the lane-mean speed system were used in this 3SLS estimation.

(Continued)

Table 17 (continued). Three-stage least squares estimation of grouped lane-speed deviations for eastbound I-90.

Variable*	Estimated Coefficient	t-statistic
Equation 2: Middle-Lane Speed Deviation (Dependent Variable)		
Constant	34.6818	10.6756
Speed Deviation in right lane	-0.0516	-1.3422
Speed Deviation in left lane	0.3791	13.6739
*Logarithm of right-lane mean speed***	-31.0753	-11.7211
*Logarithm of middle-lane mean speed***	11.2859	8.9724
*Logarithm of left-lane mean speed***	12.0551	5.0897
Time-of-week indicator (1 if it is weekend, 0 otherwise)	0.4592	8.0088
Seasonal indicator 1 (1 if it is winter, 0 otherwise)	-0.1081	-2.3134
Time-of-day indicator 1 (1 if hour of observation is between midnight and 6:00 AM, 0 otherwise)	0.2430	6.1934
Time-of-day indicator 2 (1 if it is PM peak hour, 0 otherwise)	-0.1829	-3.5588
Number of observations	2233	
R-squared	0.3598	
Corrected R-squared	0.3572	

* Variables in italics are instrumented because of possible endogeneity. This is done by regressing the variable against exogenous variables and using the regression-predicted values for the 3SLS estimation. Variables in bold are endogenous and part of the simultaneous equation estimation. Trucks are defined as vehicles with wheelbases exceeding 65 feet.

** Lane-mean speeds are instrumented variables in the speed deviation system. Predicted values from the lane-mean speed system were used in this 3SLS estimation.

(Continued)

Table 17 (continued). Three-stage least squares estimation of grouped lane-speed deviations for eastbound I-90.

Variable*	Estimated Coefficient	t-statistic
Equation 3: Left-Lane Speed Deviation (Dependent Variable)		
Constant	22.9298	3.5773
Speed Deviation in middle lane	1.0753	10.3436
*Logarithm of middle-lane mean speed***	-29.5472	-13.8764
*Logarithm of left-lane mean speed***	24.0178	11.6909
Passenger car percentage	-1.0800	-2.4323
Seasonal indicator 1 (1 if it is winter, 0 otherwise)	0.6884	6.5079
Time-of-day indicator 2 (1 if it is PM peak hour, 0 otherwise)	0.4557	3.3125
Number of observations	2233	
R-squared	0.3285	
Corrected R-squared	0.3267	
System R-squared	0.3288	

* Variables in italics are instrumented because of possible endogeneity. This is done by regressing the variable against exogenous variables and using the regression-predicted values for the 3SLS estimation. Variables in bold are endogenous and part of the simultaneous equation estimation. Finally, trucks are defined as vehicles with wheelbases exceeding 65 feet.

** Lane-mean speeds are instrumented variables in the speed deviation system. Predicted values from the lane-mean speed system were used in this 3SLS estimation.

Table 18. Three-stage least squares estimation of grouped lane-speed deviations for westbound I-90.

Variable*	Estimated Coefficient	t-statistic
Equation 1: Right-Lane Speed Deviation (Dependent Variable)		
Constant	-0.5669	-0.3505
Speed Deviation in middle lane	0.8431	37.1615
*Logarithm of right-lane mean speed***	9.7616	9.7441
*Logarithm of middle-lane mean speed***	-9.0162	-8.7505
Time-of-day indicator 1 (1 if hour of observation is between midnight and 6:00 AM, 0 otherwise)	-0.1023	-2.0570
Time-of-week indicator (1 if it is weekend, 0 otherwise)	-0.1552	-3.8241
Number of observations	2230	
R-squared	0.3965	
Corrected R-squared	0.3951	

* Variables in italics are instrumented because of possible endogeneity. This is done by regressing the variable against exogenous variables and using the regression-predicted values for the 3SLS estimation. Variables in bold are endogenous and part of the simultaneous equation estimation. Finally, trucks are defined as vehicles with wheelbases exceeding 65 feet.

** Lane-mean speeds are instrumented variables in the speed deviation system. Predicted values from the lane-mean speed system were used in this 3SLS estimation.

(Continued)

Table 18 (continued). Three-stage least squares estimation of lane-by-lane grouped lane-speed deviations for westbound I-90.

Variable*	Estimated Coefficient	t-statistic
Equation 2: Middle-Lane Speed Deviation (Dependent Variable)		
Constant	2.1069	08975
Speed Deviation in right lane	1.1797	31.0057
Speed Deviation in left lane	-0.0332	-1.6305
*Logarithm of right- lane mean speed***	-12.3373	-3.6331
*Logarithm of middle-lane mean speed***	9.5426	7.1049
*Logarithm of left-lane mean speed***	1.6071	0.5003
Time-of-week indicator 1 (1 if it is weekend, 0 otherwise)	0.1856	3.5658
Seasonal indicator 1 (1 if it is winter, 0 otherwise)	0.0380	1.5280
Time-of-day indicator 1 (1 if hour of observation is between midnight and 6:00 AM, 0 otherwise)	0.1429	2.4480
Number of observations	2230	
R-squared	0.3460	
Corrected R-squared	0.3436	

* Variables in italics are instrumented because of possible endogeneity. This is done by regressing the variable against exogenous variables and using the regression-predicted values for the 3SLS estimation. Variables in bold are endogenous and part of the simultaneous equation estimation. Finally, trucks are defined as vehicles with wheelbases exceeding 65 feet.

** Lane-mean speeds are instrumented variables in the speed deviation system. Predicted values from the lane-mean speed system were used in this 3SLS estimation.

(Continued)

Table 18 (continued). Three-stage least squares estimation of grouped lane-speed deviations for westbound I-90.

Variable*	Estimated Coefficient	t-statistic
Equation 3: Left-Lane Speed Deviation (Dependent Variable)		
Constant	29.3353	5.8855
Speed Deviation in middle lane	0.3794	5.2335
*Logarithm of middle-lane mean speed***	-33.2354	-15.0579
*Logarithm of left-lane mean speed***	26.9555	11.1272
Seasonal indicator 1 (1 if it is winter, 0 otherwise)	0.4621	4.2952
Number of observations	2230	
R-squared	0.2682	
Corrected R-squared	0.2669	
System R-squared	0.3087	

* Variables in italics are instrumented because of possible endogeneity. This is done by regressing the variable against exogenous variables and using the regression-predicted values for the 3SLS estimation. Variables in bold are endogenous and part of the simultaneous equation estimation. Finally, trucks are defined as vehicles with wheelbases exceeding 65 feet.

** Lane-mean speeds are instrumented variables in the speed deviation system. Predicted values from the lane-mean speed system were used in this 3SLS estimation.

3.5.1: 3SLS estimation of lane-by-lane grouped mean speeds

Equation 1 (right lane)[5]

Variable: Lane traffic-flow indicator (flows less than 75 veh/h)

Finding: Positively affects lane mean speeds in the eastbound direction

This finding is intuitive in that it illustrates driver tendency to drive the allowable safe speed under near free-flow conditions. Under near free-flow conditions, the visual constraints posed by the presence of adjacent vehicles are removed thereby allowing lane mean speeds to increase significantly beyond normal operating speeds (around the speed limit.) The effect appears to be significant in the eastbound direction only and it is likely that the downgrade effect for the westbound direction annuls the significance of low volumes on lane-mean speeds in the right lane.

Variable: Truck percentage in right lane[6]

Finding: Negatively affects lane-mean speeds in both directions

This finding reflects the impact of truck percentage on speed-flow distributions. Under general conditions, with no constraints on flow levels and accounting for the effect of all other factors, increasing truck percentage will tend to decrease lane-mean speeds. However, as will be illustrated in the following discussions, certain truck percentage-flow combinations will create desirable conditions for traffic flow.

---

[5] The lanes are defined as right, middle, and left relative to the direction of travel.

[6] Trucks are defined as vehicles with wheelbases exceeding 65 feet.

Variable: High truck flow in right lane

Finding: Increases right-lane mean speeds in both directions

This finding suggests that when truck flow in the right lane exceeds a threshold of flow, lane-mean speeds will increase as a result of a combination of factors. Truck drivers driving in high truck volumes tend to "draft" taking advantage of the relatively greater uniformity of vehicle type in the lane. This finding is consistent with the truck equivalency factors presented in the U.S. Highway Capacity Manual (Transportation Research Board, 1994).

Variable: Relative truck flow indicators (truck percentage exceeding 60% and total traffic flow less than 50 veh/h or truck percentage less than or equal to 20% and total lane flow exceeding 200 veh/h)

Finding: Increases lane-mean speeds in the eastbound direction

This finding is illustrative of the significance of the impact of vehicle mix on traffic flow distribution. Under low or near free-flow conditions but with a high percentage of trucks, or under higher volume conditions but with a relatively low percentage of trucks, lane-mean speeds are found to increase because of the uniformity of vehicle type. The non-uniform range that consists of flow-mix combinations of 50-150 veh/h and truck percentages of 20% to 60% is likely to cause the most detrimental impact on lane-mean speeds, as evidenced by the general finding on truck percentage. This finding is based on flows observed in the "flat portion" (i.e., the low-flow portion) of the classic speed-flow curve. As congestion increases, it is likely that the effect of vehicle mix by lane might cause a redistribution of lane use by vehicle type. The effect appears to be significant in the eastbound direction only and it is likely that the downgrade effect for the westbound direction annuls the significance of these effects on lane-mean speeds in the right lane.

Variable: Adjacent lane mean speed (middle lane)

Finding: Increasing middle-lane speeds increases right-lane mean speeds in both directions

This variable captures the endogenous lateral cause-effect relationships between adjacent lane speeds.[7] As will be evidenced in subsequent discussions, adjacent lanes tend to positively affect traffic speeds. The underlying process this factor captures is the need to drive faster to merge into adjacent lanes and also the psychological impact faster traffic in the adjacent lane has on drivers.

Variable: Time-of-day indicator (midnight to early morning)

Finding: Negatively impacts right-lane mean speeds

This finding represents selection effects of drivers choosing the right lane for travel in the morning. Drivers who tend to use the right lane under free-flow conditions, as expected in the midnight to early morning hours, usually consist of slower passenger-car drivers or truck drivers. This portion of the population tends to have lower travel speeds.

Variable: Seasonal indicators (winter, spring)

Findings: Tend to decrease right-lane mean speeds in the eastbound direction and

---

[7] Note that only the immediately adjacent lane has a statistically significant impact on lane speeds (i.e., the left-lane speeds were not fount to affect right-lane speeds).

increase right-lane mean speeds in the westbound direction

These variables capture the effect of weather on right-lane operations. Particularly in this area of I-90 where snow and associated inclement conditions occur in winter and early spring, right lanes tend to operate at lower speeds due to vehicle chaining requirements and the deterrence of adverse driving conditions in the eastbound direction. The westbound direction seems to experience anomalous effects, however, but this is likely an artifact of the data, especially the positive effect of winter coupled with no significant effect for Spring. The artifact of the data potentially arises from the location of the speed detectors. The westbound direction detectors are situated in the end portion of the "chain-up" zone for crossing the Cascade mountain range. Therefore speed data collected at the location represent a self-selected sample of vehicles whose speed distributions remain relatively unaffected by "chain-up zone" requirements.

Variable: Time-of-week indicator (weekend)

Finding: Tends to increase right-lane mean speeds in both directions

This variable represents the near free-flow conditions that exist on weekends, in addition to capturing the effect of uniformity of traffic mixes. Truck traffic in weekend periods is minimal and as evidenced before, with greater vehicle type uniformity, right lane mean speeds are expected to increase.

Variable: Time-of-day indicators (PM and AM peak hours)

Finding: Right-lane speeds increase during the PM peak hour in the eastbound direction and decrease during the AM peak hour in the eastbound and westbound directions

This peak hour variable captures the effect of several factors such as commute direction and vehicle mix uniformity. Westbound I-90 carries commuter traffic in the morning peak hour, and little or no commuter traffic occurs in the eastbound direction. In addition, freight movement is greater during the morning peak hour than in the evening peak hour. The combination of these factors leads to greater uniformity of vehicle mix in the evening peak hour and more mixed flow in the morning peak hour. The lack of a significant PM peak hour effect in the westbound direction is likely an artifact of the data. and in generic situations likely will play a significant role in both directions.

Equation 2 (middle lane)

Variable: Adjacent lane mean speeds (left and right lanes)

Finding: Increasing adjacent lane speeds increase middle lane mean speeds in both directions

This variable corroborates the finding on endogenous lateral cause-effect relationships between adjacent lane speeds. The finding on the greater impact of left lane operations further affirms our conclusion that differential lane speeds are critical to the analysis of the overall speed distribution.

Variable: Hourly traffic flow in middle lane

Finding: Tends to negatively impact middle-lane speeds in both directions

This finding is consistent with flow-speed relationships observed in other empirical studies. Given that truck-related factors were not found to significantly affect middle-lane speeds, this finding indicates that as flow in the middle lane (as opposed to the right lane) increases it represents the gradual approach to congestion, and the consequent decrease in speeds.

Variable: Lane use distribution between middle and right lanes (ratio of middle- to right-lane flows)

Finding: Increase in ratio decreases middle-lane speeds in both directions

This finding illustrates the effect of congestion and the declining choice of the middle lane as a passing lane as a result of increasing congestion. As congestion levels are approached, the use of the middle lane changes from a passing lane to a capacity lane. Consequently, driver behavior appropriately reflects a tendency to slow down under increasing flows.

Variable: Time-of-day indicator (night-time)

Finding: Tends to decrease middle-lane speeds in both directions

This finding is consistent with the intuitive expectation that night-time conditions present more challenges to the driving task, and hence cause drivers to slow down.

Variable: Time-of-week indicator (weekend)

Finding: Tends to decrease middle-lane speeds in both directions

As opposed to a positive impact on right-lane speeds, weekend effects tend to decrease middle-lane speeds. Although this finding appears counter-intuitive, when viewed within a free-flow regime context, it appears tenable. During weekends, when near free-flow conditions exist, lane usage is not governed by the need to pass, but by arbitrary choice. Vehicles that use the middle lanes in weekend periods therefore in general are not speeding to pass, as opposed to a weekday situation. As a result, it is not unusual to expect slower moving vehicles in the middle lane.

Equation 3 (left lane)

Variable: Truck percentage in left lane

Finding: Increasing truck percentage increases left-lane speeds in eastbound direction

This finding illustrates the primary effect of a passing lane on cross-sectional flow-speed relationships when from a capacity standpoint. A higher truck percentage in the left lane reflects truck drivers' tendencies to pass slower traffic in order to accelerate up the steeper grade that is immediately upstream of the eastbound direction. The geometric constraints oncoming terrain poses to truck drivers causes this phenomenon. The westbound direction experiences no significant effect due to the significant downgrade that exists.

Variable: Lane use distribution between left and middle lanes (ratio of left to middle lane flows)

Finding: Increase in ratio increases left-lane speeds in both directions

This finding illustrates that as traffic flows in the middle and right lanes approach thresholds where lane speeds have to decrease to maintain safe operations, the use of the left lane as a passing lane increases thereby attracting faster drivers.

Variable: Adjacent lane-mean speeds (middle lane)

Finding: Increasing adjacent lane speeds increase left-lane mean speeds in both directions

This variable corroborates the finding on endogenous lateral cause-effect relationships between adjacent lane speeds. The finding on the isolated impact of middle lane operations is consistent with our findings on the impact of middle lane operations on right-lane mean speeds.

Variable: Time-of-day indicator (night-time)

Finding: Increases left-lane speeds in both directions

This finding appears counter-intuitive, but provides interesting insight into drivers' perception of lane usage by time-of-day. Under night-time conditions, the use of the middle lane as a passing lane declines in favor of the left lane for drivers who tend to drive significantly faster than the average driver. Thus the night-time factor captures aggressive driving behavior and the locational occurrence of such behavior in a cross-sectional context.

3.5.2: 3SLS estimation of lane-by-lane grouped speed deviations

Equation 1 (right lane)

Variable: Speed deviations in middle lane

Finding: Middle-lane deviation positively affects speed deviation in right lane in both directions

This variable captures the lateral lane effects across the roadway. However the impact includes the car-following response effect (not expected in adjacent lane speed effects) due to lane changes that adjacent-lane deviations bring. Greater deviations in the middle lane indicate to drivers in the right lane more opportunities, although intermittent, for lane changing than a lower deviation would. Hence, the car-following driver response in the right lane is simultaneously being influenced by the opportunity for lane changing which causes the sub-conscious effect of higher fluctuation in in-lane speeds.

Variable: Lane mean speeds[8]

Finding: Right-lane mean speeds positively affect right-lane speed deviations while middle-lane mean speeds negatively affect right-lane speed deviations

This finding is intuitive and consistent with the relationships drawn in previous studies between the coefficient of dispersion and mean speeds (see for example May 1990). The negative impact of middle-lane speeds on right lane deviations indicates that drivers tend reduce their deviations as adjacent lane speeds go up in order to make their lane changing operations safer.

Variable: Time-of-day indicators (early morning and PM peak)

Finding: Early morning effects cause an increase in right-lane deviations in the eastbound direction while PM peak hour effects cause a decrease in right-lane speed deviations in the eastbound direction. In the westbound

---

[8] Lane mean speeds are instrumented variables in the speed deviation system. Predicted values from the lane mean speed system were used in this 3SLS estimation.

direction early morning effects cause a decrease in right-lane deviations while PM peak hour effects are insignificant.

The "midnight to early morning" variable, as discussed previously in its effects on lane speeds, captures driver response under near-free-flow conditions. Depending on whether it is an upgrade or a downgrade, driver response in car following is expected to change. In the eastbound direction, where a significant upgrade follows the loop detector locations, deviations tend to increase in the most-used lanes during that time of day, namely, the right and middle lanes. In contrast, the downgrade in the westbound direction collapses the speed distribution and has a downward effect on speed deviations in general.

In the PM peak hour, traffic flow increases to levels that warrant use of the middle and left lanes from a capacity standpoint, and coupled with the greater uniformity in vehicle mix, the net effect on speed deviations in the right lane is a decline. That this effect was not found to be significant in the westbound direction is explained by lack of significant commuter traffic in that direction at the location being considered. In fact, any commute-related effects in the westbound direction is marginally captured by the "early morning" variable which, as mentioned previously, has a negative impact on right-lane speed deviations.

Variable: Seasonal indicator (winter)[9]

Finding: Winter effects tend to decrease speed deviations in the eastbound direction

---

[9] In the absence of microscopic weather data, lateral lane effects are captured by the seasonal indicator. While this does not cause an omitted variable bias, real-time microscopic weather information will provide interesting insights into the impacts of factors such as precipitation versus snow pileup, and rainfall versus pavement drainage on driver behavior.

Winter effects capture the effects of driver behavior under inclement conditions. Although speeds tend to decline under inclement conditions, driver behavior is altered to the extent that significantly more attention is paid to the driving task. Drivers tend to maintain constant headways, and minimize lane changing operations.[10] The net effect of such behavior is an associated decline in right-lane deviations. It is also important to note that the "chain-up" zone occurs upstream of the eastbound direction, causing additional constraints on traffic dispersion. In the westbound direction, due to the fact that the detectors are downstream of the "chain-up" zone, such constraints are minimal. Nevertheless, the finding on the westbound effects of the season variables may merely be artifacts of the data for other reasons.

Variable: Time-of-week indicator (weekend)

Finding: Decreases right-lane speed deviations in both directions

The finding on this variable illustrates selectivity in the driving population that chooses the right lane on weekends. As mentioned previously, perhaps, this class of drivers not only maintain lower speeds but also lower deviations because they are risk averse.

Variable: Truck-to-passenger car ratio

Finding: Decreases right-lane speed deviations in the eastbound direction with no significant effect in the westbound direction

---

[10] Such behavior is more prevalent amongst drivers who choose to use the right and middle lanes. On the contrary, as will be evidenced later in the discussion of weather effects on left-lane speed deviations, the self-selection of riskier drivers in left lanes will likely cause an increase in speed deviations.

Increasing truck-to-passenger car ratio effects reiterate the impact of vehicle-mix uniformity and "truck drafting phenomena" on reduction of speed deviations. In the presence of significant upgrades, there is self-selection of the right lane by heavier traffic. In the presence of a downgrade, as evidenced in the westbound direction, this need is not compelling.

Equation 2 (middle lane)

Variable: Speed deviations in right and left lanes

Finding: Right-lane deviations negatively affect middle-lane speed deviations while left-lane speed deviations have a positive impact in the eastbound direction. In the westbound direction, the effects are opposite

The finding on these variables are consistent with the unobserved effects due to grades as presented in previous discussions.

Variable: Lane-mean speeds

Finding: Right-lane mean speeds have a negative impact on middle-lane speed deviations while middle- and left-lane mean speeds have positive impacts in both directions

Higher right-lane mean speeds indicate that the vehicle-to-vehicle interaction in the traffic flow continuum is smoother with drivers experiencing a decreased need for lane changing. Consequently speed deviation in the middle lane is affected inversely with the lane change need. The positive impact of middle- and left-lane speeds on in-lane deviations appears aberrational and inconsistent with previous findings However, it is likely capturing the flux in driver selectivity in the middle lane. Middle-lane users are arguably the most diverse in terms of their inherent driving natures, and hence may have a fundamental tendency to vary their speeds more. Consequently in situations where in-lane or left-lane speeds increase, drivers may be increasing their speeds in order to change to the left lane or that traffic volumes are quite below capacity.[11]

Variable: Time-of-week indicator (weekend)

Finding: Increase in on middle-lane speed deviations in both directions

This finding corroborates previous inferences given that traffic volumes in the middle lane during weekends are expected to be minimal.

Variable: Seasonal indicator (winter)

Finding: Decrease in middle-lane speed deviation in the eastbound direction while increasing in the westbound direction

This finding is consistent with evidence on winter effects found in previous variables. However, it appears that westbound direction experiences adverse effects (positive) in the passing lanes, the effects most likely being downgrade-related.

Variable: Time-of-day indicator (midnight to early morning)

Finding: Increases middle-lane speed deviations in both directions

---

[11] In the westbound direction, the effect of left-lane speed is statistically insignificant (t-statistic = 0.50). However as mentioned previously, this could be an artifact of the data and it is likely that when grade-related effects are quantified as continuous variables and interactions by lane, such anomalies will diminish. The intent of this paper to present as general a specification that offers comparable insights into the effects of endogenous variables on speed and speed deviation relationships.

This finding indicates the effect of free-flow conditions on driver behavior when the population selection effect (as evidenced in right lane relationships) is absent. Coupled with the fact that the middle lane and left lanes serve as passing lanes under such conditions, causes an increase in speed deviations.

Variable: Time-of-day indicator (PM peak hour)

Finding: Decrease in middle-lane speed deviations in the eastbound direction and insignificant in the westbound direction

This finding indicates the effect of commute-related volume effects on speed deviations and is consistent with earlier inferences on traffic flow variable-related impacts.

Equation 3 (left lane)

Variable: Speed deviation in middle lane

Finding: Positively affects speed deviations in the left lane in both directions

The lateral "friction" effects caused by interaction between the middle and left lanes are captured by this variable. Given that this portion of I-90 is largely in the "flat" portion of the upper part of the speed flow curve, the middle and left lanes serve predominantly as passing lanes, thereby experiencing a self-selected sample of drivers who are more risk-prone and significantly influenced by variations in speed. The positive impact of the variable captures such lateral lane effect dynamics.

Variable: Middle- and left-lane speeds

Finding: Increasing middle-lane speeds decrease left lane deviations while increasing in-lane speeds increase in-lane deviations in both directions

This variable captures similar to the endogenous relationships between the middle and right lanes, the cascading effect of speed variation from the right lane to the left lane. With higher middle lane speeds the need to change to the left lane decreases, thereby minimizing friction in the left lane. On the other hand, when left-lane speeds increase, there is a consequent increase in speed deviation because the self-selection of the left lane to the most risk-prone drivers is greatest.

Variable: Passenger car percentage

Finding: Negatively affects eastbound speed deviations

This finding illustrates locale-specific effects related to grades. The eastbound direction which experiences significant upgrades consequently also experiences a greater distribution of truck traffic across the cross-section. The passenger car variable captures this effect and corroborates the impact of uniformity of vehicle mix on traffic flow dispersion.

Variable: Time-of-day indicator (PM peak hour)

Finding: Increases eastbound speed deviations

This finding is consistent with earlier discussions in that the middle and left lanes serve as passing lanes and commuter lanes in the PM peak hour in the eastbound direction. The positive effect may also be capturing the tendency of drivers in their home-bound commute to take greater risks, evidence that cannot be supported in the westbound direction because of its lack of commute effects.

Variable: Seasonal indicator (winter)

Finding: Increases left-lane speed deviations in both directions

This finding is consistent with those presented for middle-lane speed deviations.

3.6: IMPLICATIONS OF FINDINGS

The findings provide useful insights into the complex interactions between vehicles in traffic streams and how those interactions are affected by temporal, seasonal and traffic vehicle mix factors. Accounting for endogeneity and the forward recursive relationship between speed and speed deviation also gives indications that recursive relationships seem to be an integral part of traffic flow regimes. When the time window is collapsed to micro-levels such as 20-second or 5-second intervals as opposed to 1-hour intervals, it is quite likely that locale-specific recursive relationships accounting for multiplier effects (in the form of pre-determined, exogenous lagged dependent variables) might be uncovered. This would be especially beneficial in investigating "shock wave" phenomena from a behavioral standpoint, and help identify traffic control strategies to counter the adverse effects of such phenomena.

CHAPTER 4: CONCLUSIONS AND DIRECTIONS FOR FURTHER RESEARCH

4.1: CONCLUSIONS

4.1.1: GLOBAL FINDINGS

It has been shown through this research effort that the accident likelihood models for predicting frequency and severity as well as the structural models of traffic flow hold promise in unraveling the complexity of underlying processes. By accounting for interactions in addition to main effects, the potential for expanding the realm of strategies to counter adverse weather effects is enhanced. Fundamental understandings of how existing roadway design benefits can be optimized through ITS implementation serve as cornerstones of this research. In particular, the approaches presented herein can be used to thoroughly evaluate the safety and operational impacts of variable-message/speed-limit signs, in-vehicle units, and other ITS technologies. Such evaluations will serve as a cornerstone to justify future ITS expenditures.

However, the approaches embodied in this dissertation, while being local in model specification, have broader applicability beyond ITS, encompassing critical infrastructure design, programming and investment issues relating to traffic safety and operations at regional and national levels. Agency decision making especially from a traffic safety and operational standpoint, is now afforded increased sophistication and rationale in analyses of benefits and cost-effectiveness of investment strategies.

4.1.2: MODEL-SPECIFIC FINDINGS

In terms of modeling marginal likelihoods of accidents, this study has shown that the negative binomial and Poisson approaches are appropriate. It was found that environmental and geometric factors (main effects and interactions) differed in their effects on overall accident likelihoods compared to specific accident types. This finding is consistent with site-specific accident reports and provides roadway engineers important control factors that are interactive and not limited to the pavement-tire interaction approaches currently being used.

The study provides a framework for estimating accident severity likelihood conditioned on the occurrence of an accident. It was concluded that a nested logit model which accounted for shared unobservables between property damage and possible injury accidents provided the best structural fit for the observed distribution of accident severities. This represents an important step in the methodological evaluation of ITS with respect to accident safety. By developing a probabilistic model that contains several important variables representing geometric, weather, and human factors we have shown that ambiguity and bias stemming from confounding effects in a partially specified model can be eliminated. In addition, this research provides suggestive results by its use of variables such as curve-sobriety interaction and curve-pavement surface interaction. Specifically, it suggests that ITS may be an effective means of compensating for adverse design, human factors, and weather conditions. A well designed ITS could significantly improve the driving task in the presence of adverse factors such as alcohol, inclement weather, and complex roadway geometrics. A significant shift in the distribution of accident severities toward milder accidents in combination with lower accident frequencies (Shankar, Mannering, and Barfield, 1995) will provide a basis for ITS evaluation.

Endogenous relationships within lane speeds and between lane speeds and speed deviations were found to be statistically valid. The westbound and eastbound directions of our study site experienced dissimilar effects related to grade, time-of-day and time-of-week characteristics. On the other hand, the endogenous relationships in large part are similar, with estimated coefficients of like sign, means and standard errors. Our findings show that in-lane speeds are affected only by adjacent-lane speeds and in-lane speed deviations are affected progressively by adjacent lane speed deviations and in addition, in-lane and adjacent-lane speeds. Coupled with findings on the contemporaneous impact of temporal and vehicle-mix factors, such inferences corroborate the need for a comprehensive investigation into lane-mean speed and lane-speed deviation relationships. To be sure, the data we used was limited (i.e., a single site) it that it did not allow us to explore variations in geometric characteristics, functional classifications, and other factors that might vary from site to site. Further insights could be gained from a more diverse data set that encompasses various regions and roadway functional classes.

4.2: DIRECTIONS FOR FURTHER RESEARCH

4.2.1: GLOBAL DIRECTIONS

The promise the dissertation has shown through the WSDOT programming study on I-90 for expanding the continuum of improvement strategies in concert with sensible benefit-cost analyses leads the author to believe that turn-key approaches similar to those embodied in this effort are worthwhile in a national context that is copious in its diversity of environmental and geometric interactions.

The proposed methodology offers significant benefits while suffering from some limitations. The benefits of this research include the development of new methodologies that are comprehensive in model specification and employ appropriate underlying distributions. The methodologies also provide a framework for assessing the incremental impacts of parameters (elasticities), as well as in evaluating new technologies. Likelihood ratio tests and tests of a similar nature applicable to structural models using constrained and unconstrained parameter estimations can be employed to assess the effectiveness of ITS. Depending on how these tests are constructed, inferences on the overall structural shift in parameter distribution with ITS deployment as well as the effects of specific independent variables before and after ITS can be examined. As part of the before-and-after analysis, overall temporal stability should be evaluated as well (see Mannering, Murakami, and Kim, 1994 for an application of such coefficient stability tests). Given that the before-data period is of longer duration, it is quite likely that changes into automobile design and consequent effects on driver behavior or weather changes may be playing a significant role. If non-ITS technologies such as these are found to be significant, model parameters should be re-estimated to provide for temporal stability in the combined before-and-after period. Such an analysis is important because we are not simply testing for differences in before and after accident frequencies, but isolating the true causality of these differences due to ITS by controlling for the complex interaction between geometrics and weather conditions. The finding of statistically significant instability in coefficients could then be attributed to the variable message/speed-limit signing and the in-vehicle signing systems. A more simplistic comparison of before and after data could easily lead to erroneous conclusions. For example, one could conclude that the signing system was ineffective in reducing accidents but slight variations in weather between before and after data could be masking the system's effectiveness. In addition to being able to determine whether or not the proposed signing system was effective, an analysis of changes in coefficient elasticities and the magnitudes of indicator variables will allow us to more precisely isolate the effectiveness of the signing system. For example, we may be able to specifically state that the signing system mitigated the adverse effects of high snowfall on grades greater than 2 percent. Such specificity is needed to make definitive statements regarding ITS technologies.

From a consumer standpoint, these methodologies also provide agencies the framework to assess cost savings due to technological implementation. The limitations of this research are that the database is limited to a rural freeway corridor. Freeway cross sections and other traffic elements such as traffic volumes in this corridor have little or no variation. In addition, weather-related data used in model specifications are more global than local in nature. The corridor is also a single roadway corridor, not encompassing roadways of different functional classifications.

In terms of future research building on the proposed methodologies, an expanded database that includes encompasses different locales using the same technology would address ITS effectiveness issues in greater detail. However, it must be borne in mind that technological implementation that is invariant across locales will not be achieved until the market demand for such systems justifies the returns. In addition, it must also be noted that human response to such technology may not be invariant across systems, a fundamental assumption in the assessment of new systems across populations. Other issues regarding system reliability are worth examining in the long term. Duration models that predict the risk of failure conditioned on the fact that the system has not failed to that point in time could be employed as plausible statistical tools. The deployment of local weather stations may provide the researcher opportunities to dynamically forecast real-time weather conditions and inter-link such information with accident prediction and speed-flow-speed deviation models.

4.2.2: MODEL-SPECIFIC DIRECTIONS

The accident frequency models are limited to rural interstate freeways and as a result suffer from limitations pertaining to functional class and locational nature of the accidents. Intersection frequency models have been in use in limited fashion (Poch and Mannering 1996); the need for regionwide and national-level intersection models remains.

The nested structure of the severity model provides insights into factors affecting conditional likelihoods of accident severity. Conversely, if one were to examine the social welfare of accident savings using joint distributions that model frequencies and severities simultaneously, models of accident severity cost can be constructed to assess system wide strategies that optimize savings in accident cost. Motivations for such structures need not be limited to shared unobservables; rather, examine accident distributions from a variance standpoint. Such approaches can provide interesting insights into safety treatments for low-accident locations versus high-accident locations, and the cascading effects this research can have on regional safety programming agendas can be significant. Additionally, it must be noted that the cross-nest parametric constraints used in this study may or may not hold under joint distributional examination; in fact other parametric constraints of interest may surface.

The findings gathered from the endogenous relationship between lane-mean speeds and lane-speed deviations appear promising for further application of the structural equations methodology to macroscopic traffic-flow modeling. It is quite possible that dynamic effects could be uncovered to a greater extent with more microscopic data by incorporating pre-determined lane-mean speed and lane-speed deviation variables in the specifications. Such a study could have objectives relating to the unraveling of incremental dynamics in traffic flow under smaller time windows and greater seasonal, vehicle-mix constraints. While the present work offers generic insights, understanding the cause-effect relationships between lane-mean speed and lane-speed deviations under such constraints could enrich our knowledge of driver response under specific conditions. Such knowledge will be beneficial to the design and planning of advanced traffic management systems intended for the improvement of traffic flow and safety.

BIBLIOGRAPHY

Alassar, L. Analysis of heavy truck accident severity, *Journal of Advanced Transportation*, 22:77-91;1988.

Bates, G. E.; Neyman, J. Contributions to the theory of accident proneness: I. An optimistic model of the correlation between light and severe accidents, *University of California Public Statistics*, 1:215-254;1952.

Ben-Akiva, M.; Lerman, S.R. Discrete choice analysis: Theory and application to travel demand, *The MIT Press*, Cambridge, Massachusetts; 1985.

Breslow, N. Extra-Poisson variation in log-linear models, *Applied Statistics*, 33:38-44; 1984.

Boughton, C. J. *Accidents and Geometric Design Research Rreport ARR44*, Vermount South, Victoria, Australia, Australian Road Research Board; 1975.

Brilon, W. ; Ponzlet, M. Variability of speed-flow relationships on German autobahns, *Presented at the 75th Annual Meeting of the Transportation Research Board*, Washington, D.C., 1996.

Brodsky, H.; Hakkert, A. S. Risk of a road accident in rainy weather, *Accident Analysis and Prevention*, 20:161-176; 1988.

Boyle, A. J.; Wright, C. C. Accident "migration" after remedial treatment at accident blackspots, *Traffic Engineering and Control*, 25:260-267; 1984.

Cameron, A. C.; Trivedi, P. K. Econometric models based on count data: Comparison and applications of some estimators and tests, *Journal of Applied Econometrics*, 1:29-53; 1986.

Cameron, A. C.; Trivedi, P. K. Regression-based tests for overdispersion in the Poisson model, *Journal of Econometrics*, 46:347-364; 1990.

Easa, S.M. ; May, A.D. Generalized procedure for estimating single- and two-regime traffic flow models, *Transportation Research Record*, 772:24-37; 1980.

Edie, L.C. Following and steady-state theory for non-congested traffic. *Operations Research*, 9:66-76; 1961.

Engel, J. Models for response data showing extra-Poisson variation, *Statistical Neerlandica* 38:159-167; 1984.

Evans, L. Double pair comparison -- a new method to determine how occupant characteristics affect fatality risk in traffic crashes, *Accident Analysis and Prevention*, 18(3):217-228; 1986.

Evans, L. The effectiveness of safety belts in preventing fatalities, *Accident Analysis and Prevention*, 18(3):229-241; 1986.

Evans, L. The fraction of traffic fatalities attributable to alcohol, *Accident Analysis and Prevention*, 22(6):587-602; 1990.

Federal Highway Administration. Synthesis of safety research related to traffic control and roadway elements, *Report TS-82-232*, Washington, D.C.: Federal Highway Administration, Office of Research, Development and Technology; 1982.

Frisbie, G. A. Ehrenberg's negative binomial model applied to grocery store trips. *Journal of Marketing Research*, 17:385-390; 1980.

Glimm, J.; Fenton, R.E. An accident-severity analysis of a uniform-spacing headway policy, *IEEE Transactions of Vehicular Technology*, VT-29(1):96-103; 1980.

Golob, T.F.; Recker, W.W.; Leonard, J.D. An analysis of the severity and incident duration of truck-involved freeway accidents, *Accident Analysis and Prevention*, 19(5):375-395; 1987.

Greenberg, H. An analysis of traffic flow, *Operations Research*, 7:78-85; 1959.

Greene, W. Econometric Analysis, Macmillan Publishing, New York, NY; 1993.

Greenshields, B.D. A study in highway capacity, *Highway Research Board, Proceedings*, 14: 448-477; 1935.

Hammerslag, R.; Roos, J. P.; Kwakernaak, M. Analysis of accidents in traffic situations by means of multiproportional weighted Poisson model, *Transportation Research Record* 847:29-36; 1982.

Hanbali, R. M. Influence of winter road maintenance on traffic accident rates, *Ph.D. Dissertation*, Marquette University; 1992.

Hellerstein, D. M. Using count data models in travel cost analysis with aggregate data, *American Journal of Agricultural Economics*, 73:860-867; 1991.

Holubowycz, O.T.; Kloeden, C.N.; McLean, J.A. Age, sex, and blood alcohol concentration of killed and injured drivers, riders, and passengers, *Accident Analysis and Prevention*, 26(4):483-492; 1994.

Hooper, K.; McGee, H. Driver perception-reaction time: Are revisions to current specification values in order?, *Transportation Research Record*, 904:21-30; 1983.

Hutchinson, T.P. Statistical modelling of injury severity, with special reference to driver and front seat passenger in single-vehicle crashes, *Accident Analysis and Prevention*, 18(2):157-167; 1986.

Ibrahim, A.T. ; Hall, F.L. Effect of adverse weather conditions on speed-flow-occupancy relationships, *Transportation Research Record,* 1457:184-191; 1994 .

Ivey, D.; Griffin, L. I.; Newton, T. H.; Lytten, R. L.; Hankins, K. C. Predicting wet weather accidents, *Accident Analysis and Prevention*, 13:83-99; 1981.

Iwasaki, M. Empirical analysis of congested traffic flow characteristics and free speed affected by geometric factors on an intercity expressway, *Transportation Research Record*, 1320:242-250; 1991.

Johnston, J. Econometric methods, McGraw-Hill, Inc.; 1984.

Jonah, B.S. Accident risk and risk-taking behaviour among young drivers, *Accident Analysis and Prevention*, 18(4):255-271; 1986.

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

Jones, I.S.; Whitfield, R.A. Predicting injury risk with "new car assessment program" crashworthiness ratings, *Accident Analysis and Prevention*, 20(6):411-419; 1988.

Joshua, S. C.; Garber, N. J. Estimating truck accident rate and involvements using linear and Poisson regression models, *Transportation Planning and Technology*, 15:41-58; 1990.

Jovanis, P.; Chang, H. L. Modeling the relationship of accidents to miles traveled, *Transportation Research Record*, 1068:42-51; 1986.

Jovanis, P.; Delleur, J. Exposure-based analysis of motor vehicle accidents, *Accident Analysis and Prevention*, 13:83-99; 1981.

Lassarre, S. The introduction of the variables "traffic volume," "speed" and "belt-wearing" into a predictive model of the severity of accidents, *Accident Analysis and Prevention*, 18(2):129-134; 1986.

Lawless, J. F. Negative binomial and mixed Poisson regression, *The Canadian Journal of Statistics*, 15:209-225; 1987.

Lui, K.J.; McGee, D.; Rhodes, P.; Pollock, D. An application of a conditional logistic regression to study the effects of safety belts, principal impact points, and car weights on drivers' fatalities, *Journal of Safety Research*, 19:197-203; 1988.

Maher, M. J. Accident migration - a statistical explanation?, *Traffic Engineering and Control*, 28:480-483; 1987.

Maher, M. J. A bivariate negative binomial model to explain traffic accident migration, *Accident Analysis and Prevention*, 22:487-498; 1990.

Maher, M. J. A new bivariate negative binomial model for accident frequencies, *Traffic Engineering and Control*, 32:422-425; 1991.

Mannering, F. L. Modeling driver decision-making: A review of methodological alternatives, in *Human Factors in Intelligent Vehicle Highway Systems, Woodrow Barfield and Tom Dingus, eds.*, Lawrence Erlbaum; 1995.

Mannering F. L.; Kilareski, W. Principles of Highway Engineering and Traffic Analysis. John Wiley & Sons, New York, NY; 1990.

Mannering, F. L.; Murakami, E.; Kim, S. G. Models of travelers' activity choice and home-stay duration: analysis of functional form and temporal stability, *Transportation* 21(4):371-392; 1994.

Mannering, F.; Winston, C. Dynamic empirical analysis of household vehicle ownership and utilization, *Rand Journal of Economics*, 16(2): 215-236; 1985.

Mannering, F.; Winston, C. Brand loyalty and the decline of American automobile firms, *Brookings Papers on Economic Activity, Microeconomics*, 67-114; 1991.

Mannering, F.; Winston, C. Automobile air bags in the 1990's: Market failure or market efficiency?, *Journal of Law and Economics*, 38(2):265-279; 1995.

Manton, K. G; Woodbury, M. A; Stallard, E. A variance components approach to categorical data models with heterogeneous cell population: Analysis of spatial gradients in lung cancer mortality rates in North Carolina counties, *Biometrics*, 37:259-269; 1981.

Massie, D. L.; Campbell, K. L.; Blower, D. F.; Development of a collision typology for evaluation of collision avoidance strategies, *Accident Analysis and Prevention*, 25:241-257; 1993.

May, A.D. ; Keller, H.E.M. Evaluation of single- and two-regime traffic flow models, *Fourth International Symposium on the Theory of Traffic Flow Proceedings*, Karlsruhe, Germany; 1968.

May, A.D. Traffic flow fundamentals, Prentice Hall, Englewood Cliffs, NJ; 1990.

Mayhew, D.R.; Donelson, A.C.; Beirness, D.J.; Simpson, H.M. Youth, alcohol and relative risk of crash involvement, *Accident Analysis and Prevention*, 18(4):273-287; 1986.

McCullagh, P; Nelder, J. A. Generalized linear models, Chapman and Hall, London; 1983.

McFadden, D. Econometric model of probabilistic choice. *In Structural analysis of discrete data with econometric applications, Manski and McFadden, eds.*, The MIT Press, Cambridge, Massachusetts; 1981.

McGuigan, D. R. D. Accident "migration" -- or a flight of fancy?, *Traffic Engineering and Control*, 26:229-233; 1985.

Mercer, W.G. Influences of passenger vehicle casualty accident frequency and severity: unemployment, driver gender, driver age, and drinking driving and restraint device use, *Accident Analysis and Prevention*, 19(3):231-236; 1987.

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

Miaou, S. P.; Hu, P. S.; Wright, T.; Davis, S. C.; Rathi, A. K. Development of relationships between truck accidents and highway geometric design: Phase I, *Technical memorandum prepared by the Oak Ridge National Laboratory*, Washington, D.C: Federal Highway Administration; March and November 1991.

Miaou, S. P.; Lum, H. Modeling vehicle accidents and highway geometric design relationships. Accident Analysis and Prevention. 25:689-709; 1993.

Mori, H.; Uematsu, T. Statistical analysis of road accidents as related to traffic environmental factors. *NRIPS Report*, 1:13-29; 1967 (in Japanese).

National Cooperative Highway Research Program Cost and safety effectiveness of highway design elements, NCHRP Report 197, Washington, D.C.: Transportation Research Board; 1978.

Okamoto, H.; Koshi, M. A method to cope with the random errors of observed accident rates in regression analysis, *Accident Analysis and Prevention*, 21:317-332; 1989.

Olson, P.L. Driver perception response time, *SAE Paper No.* 890731:67-77; 1989.

Piegorsch, W. W. Maximum likelihood estimation of the negative binomial dispersion coefficient, *Biometrics*, 46:863-867; 1990.

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

Press, S.J.; Wilson, S. Choosing between logistic regression and discriminant analysis, *Journal of the American Statistical Association*, 73:699-705; 1978.

Sabey, B.; Taylor, H. The known risks we run: The highway. Crowthorne, United Kingdom.: *Supplementary Report SR 567*, Transport and Road Research Laboratory; 1980.

Senn, S. J.; Collie, G. S. Accident blackspots and the bivariate negative binomial, *Traffic Engineering and Control*, 29:168-169; 1988.

Shankar, V.N.; Mannering, F.L.; Barfield, W. Effect of roadway geometrics and environmental conditions on rural accident frequencies, *Accident Analysis and Prevention*, 27(3): 371-389; 1995.

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

Shankar, V.N.; Mannering, F.L. An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity, *Journal of Safety Research*, 27(3):183-194; 1996.

Shao, S.P. Estimating car driver injury severity in car/tractor-trailer collisions, *Accident Analysis and Prevention*, 19(3):207-217; 1987.

Shibata, A.; Fukuda, K. Risk factors of fatality in motor vehicle traffic accidents, *Accident Analysis and Prevention*, 26(3):391-397; 1994.

Small, K.A.; Hsiao, C. Multinomial logit specification tests, *International Economic Review*, 26:619-627; 1985.

Snyder, J. F. Environmental determinants of traffic accidents: An alternate model, *Transportation Research Record*, 486:11-18; 1974.

Taoka, G.T. Statistical evaluation of brake reaction time, *Compendium Of Technical Papers*: 30-36, 52nd Annual Meeting of Institute of Transportation Engineers; 1982.

Train, K. Qualitative choice analysis: theory, econometrics, and application to automobile demand, The MIT Press, Cambridge, Massachusetts; 1986.

Treat, J. R. A study of precrash factors involved in traffic accidents, *Research Review*, 10(6)/11(1): University of Michigan, Highway Safety Research Institute: 1980.

Transportation Research Board. Highway Capacity Manual, *Special Report 209*, National Research Council, Washington, DC; 1994.

Triggs, T.; Harris, W.G. Reaction time of drivers to road stimuli, *Human Factors Report HFR-12*, Monash University, Australia; 1982.

Wong, Y. D.; Nicholson, A. Driver behavior at horizontal curves: risk compensation and the margin of safety, *Accident Analysis and Prevention*, 24(4): 425-436; 1992.

VITA

Venkataraman N. Shankar

University of Washington

1997

Venkataraman Shankar is currently employed as a Research Engineer with the Washington State Department of Transportation in its Olympia office. He is responsible for the statistical and econometric modeling of safety- and infrastructure-related issues statewide. Mr. Shankar currently serves as Principal or Co-Investigator on seven research projects involving infrastructure and accident prediction models for the WSDOT and studies on automobile safety led by the Brookings Institution. He has authored and co-authored nearly twenty peer-reviewed journal articles and conference presentations on traffic safety, operations and transportation planning. Prior to joining the Washington State Department of Transportation, Mr. Shankar served as a Pre-Doctoral Research Associate with the Department of Civil Engineering at the University of Washington from 1993 to 1996 developing statistical and econometric models on TravelAid, a major ITS effort on Snoqualmie Pass sponsored jointly by the National Highway Traffic Safety Administration (NHTSA) and WSDOT.

Prior to his involvement in the statistical and econometric modeling of transportation systems, Mr. Shankar worked in the transportation consulting sector for nearly seven years conducting over one hundred fifty projects encompassing the entire West coast of the United States. Mr. Shankar is a registered professional civil engineer in the state of Washington, and holds a Bachelor of Technology (Honors) degree in Civil Engineering from the Indian Institute of Technology, Madras, India and a Master of Science degree in Civil Engineering from the University of Washington, Seattle.