The Dissertation Committee for Jianming Ma certifies that this is the approved version of the following dissertation:

BAYESIAN MULTIVARIATE POISSON-LOGNORMAL REGRESSION FOR CRASH PREDICTION ON RURAL TWO-LANE HIGHWAYS

Committee:

____
Kara Kockelman, Supervisor

____
Paul Damien

____
Robert B. Gilbert

____
Randy B. Machemehl

____
Daniel A. Powers

BAYESIAN MULTIVARIATE POISSON-LOGNORMAL REGRESSION FOR CRASH PREDICTION ON RURAL TWO-LANE HIGHWAYS

by

Jianming Ma, B.S.; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

the University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May, 2006

To my wife for always giving me happiness and love, and to my parents and my siblings for their support and faith.

Acknowledgements

The author would like to thank Dr. Kara Kockelman for her constant guidance and support in the development of this dissertation and his graduate studies in the University of Texas at Austin. Thanks for being an excellent mentor and for providing him with an example of how to become a respectable educator and researcher. The author would also like to thank the participation and advice of members of the preliminary and final dissertation committees, Drs. Paul Damien, Robert B. Gilbert, Randy B. Machemehl, Shaw-Pin Miaou, and Daniel A. Powers. The author must acknowledge his indebtedness to Dr. Damien, whose statistical expertise guided him. He is also grateful to Dr. Powers whose door was always open for many discussions. The author also has to thank Dr. Machemehl for his help, guidance and understanding throughout the author's career in the University of Texas at Austin.

The author would like to thank the current and former members of the Transportation Program of the Civil, Environmental, and Architectural Engineering in the University of Texas at Austin for their support. The author would also like to show his special gratitude to Ms. Annette Perrone for her never failing willingness to assist and encourage. Special thanks also to Runhua Guo, Jared Heiner, Feng Hong, Tian Huang, Rong Luo, Feng Wang, Xiaokun Wang, Zhong Wang, Yong Zhao, Bin Zhou, and other fellow students for their friendship and kind help.

Finally, the author would like to thank his family; especially his parents and his siblings for their unconditional love and support. The author would also like to express his deep appreciation and gratitude to his wonderful wife, Zheng Li, for her compassionate love and encouragement.

BAYESIAN MULTIVARIATE POISSON-LOGNORMAL REGRESSION FOR CRASH PREDICTION ON RURAL TWO-LANE HIGHWAYS

Publication No.______________

Jianming Ma, Ph.D.

The University of Texas at Austin, 2006

Supervisor: Kara Kockelman

Roadway safety is a major concern for the general public and public agencies. Roadway crashes claim many lives and cause substantial economic losses each year. The situation is of particular interest on rural two-lane roadways, which experience significantly higher fatality rates than urban roads. There have been numerous efforts devoted to investigating crash occurrence as related to roadway design features, environmental and traffic conditions. However, most of the research has relied on univariate count models; that is, traffic crash counts at different levels of severity are estimated separately. The widely used univariate count data models ignore the following issues: (1) interdependence may exist between crash counts at different levels of severity for a specific segment of

roadway, and (2) road geometric design features, road use, and environmental conditions may have distinct effects on crashes of different severity.

The objective of this research is to model correlated traffic crash counts simultaneously at different levels of severity using multivariate Poisson-lognormal (MVPLN) models. The MVPLN specification allows for a more general correlation structure as well as overdispersion. This approach addresses some questions that are difficult to answer by estimating them separately. With recent advancements in crash modeling and Bayesian statistics, the parameter estimation is done within the Bayesian paradigm, using a Gibbs Sampler and the Metropolis-Hastings (M-H) algorithms.

As an illustration, the MVPLN specification is empirically applied to investigate crash frequency by severity using crashes that occurred on Washington State rural two-lane highways in the Puget Sound region in 2002. Thanks to MCMC simulation techniques, the marginal posterior distributions of all parameters of interest were obtained. The estimation results from the MVPLN approach did show statistically significant correlations between crash counts at different levels of injury severity. The non-zero diagonal elements suggested an existence of overdispersion crash counts at all levels of severity. The results lend themselves to several recommendations for highway safety treatments and design policies. For example, wide lanes and shoulders are key for reducing crash frequencies, as are longer vertical curves. Moreover, using a cost-benefit approach and assumptions about travel speed changes, model results suggest that time savings from raising speed limits 10 mi/h (from 50 to 60 mi/h) may not be worth the added crash cost.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

NOTATIONS

The following is the order in which the notation is used in this dissertation.

y : A vector of dependent variables

X : A matrix of explanatory variables

θ : A vector of unknown parameters

$\pi\left(\theta \middle| X\right) = \pi\left(\theta\right)$: The prior density of the random parameters θ (X is not relevant to θ)

$\pi\left(y \middle| \theta, X\right)$: The likelihood of y given a particular set of values for θ and X

$\pi\left(y \middle| X\right)$: The marginal likelihood of y given X

$\pi\left(\theta \middle| y, X\right)$: The posterior density of θ given y and X

λ : The mean and variance of a Poisson distribution

$\Theta^{(m)}$: The values that the state of the chain takes at time (or step) m

$q_{\Theta^{(m)}}\left(\Theta^{(m)} \middle| \Theta^{(m-1)}\right)$: The transition kernel for the state of the chain at time (or step) m given time (or step) $m-1$

$p\left(\theta^{(m-1)}, \theta^{(m)}\right) = P\left(\Theta^{(m)} = \theta^{(m)} \middle| \Theta^{(m-1)} = \theta^{(m-1)}\right)$: The (one-step) transition probabilities for the state of the chain at time (or step) m given time (or step) $m-1$

$p^{(\tau)}\left(\theta^{(m-\tau)}, \theta^{(m)}\right) = P\left(\Theta^{(m)} = \theta^{(m)} \middle| \Theta^{(m-\tau)} = \theta^{(m-\tau)}\right)$: The τ -step transition probabilities for the state of the chain at time (or step) m given time (or step) $m-\tau$

$p_{ij} = p\left(\theta_i, \theta_j\right)$: The (one-step) transition probabilities for the chain at state θ_j given θ_i

$P = \left(p_{ij}\right)$: The one-step transition probability matrix

$\pi\left(\theta_i \middle| \theta_{-i}\right)$: The full conditional posterior density of θ_i given values of other parameters $\theta_{-i}\left(=\theta_j, j \neq i, j \in \{1,2,\ldots,K\}\right)$

$\alpha\left(\theta^{(m-1)},\theta^{(m)}\right)$: The acceptance rate function

$q\left(\theta^{(\tau_1)},\theta^{(\tau_2)}\right)$: A Markov chain transition kernel

$I\{\bullet\}$: An indicator function

Ω_i: The variance covariance matrix of crash counts by severity on segment i

y_{is}: The number of crash counts on segment i at severity level s

x_{is}: A vector of independent variables for segment i at severity level s

β_s: A vector of unknown parameters at severity level s

$\boldsymbol{\varepsilon} = \left(\vec{\boldsymbol{\varepsilon}}_1', \vec{\boldsymbol{\varepsilon}}_2', \ldots, \vec{\boldsymbol{\varepsilon}}_n'\right)'$: The severity-level-specific unobserved heterogeneity across roadway segments

$\lambda_{is} = \exp\left(x_{is}'\beta_s + \varepsilon_{is}\right)$: The mean of Poisson distribution on segment i at severity level s

Σ: The variance covariance matrix of severity-level-specific error terms

$\xi_{is} = \exp\left(x_{is}'\beta_s\right)$

$$\Lambda_i = diag\left(\vec{\boldsymbol{\lambda}}_i\right) = \begin{bmatrix} \lambda_{i1} & 0 & \cdots & 0 \\ 0 & \lambda_{i2} & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & \lambda_{iS} \end{bmatrix}$$

$\phi_S(0,\Sigma)$: The S-variate normal density with zero mean and covariance matrix Σ

$f_W\left(\nu_\Sigma, V_\Sigma\right)$: The Wishart distribution with ν_Σ degrees of freedom and scale matrix V_Σ

C_i: The constant terms which do not involve the parameters ($\vec{\boldsymbol{\varepsilon}}_i$) to be estimated

$\pi^{p}\left(\vec{\boldsymbol{\varepsilon}}_i \middle| \vec{\mathbf{y}}_i, \beta, \Sigma\right)$: The posterior density after separating C_i

$f_T\left(\vec{\boldsymbol{\theta}} \middle| \hat{\boldsymbol{\theta}}, V_\theta, \nu_\theta\right)$: The multivariate t distribution with mean $\hat{\boldsymbol{\theta}}$, variance covariance matrix V_θ, and ν_θ degrees of freedom

$\vec{\mathbf{g}}_{\varepsilon_i}$: A vector of gradient for $\vec{\boldsymbol{\varepsilon}}_i$

H_{ε_i} : A Hessian matrix for $\vec{\boldsymbol{\varepsilon}}_i$

C_{-s} : The constant terms which do not involve the parameters (β_s) to be estimated

$\vec{\mathbf{g}}_{\beta_s}$: A vector of gradient for β_s

H_{β_s} : A Hessian matrix for β_s

CHAPTER 1 INTRODUCTION

1.1 Traffic Crashes – An Overview

Roadway safety has been a major concern for the general public and various government agencies for many decades. In 2002, the World Health Organization (WHO 2002a) estimated the death toll to exceed 1.2 million. In that same year, traffic crashes ranked tenth among top killers in terms of years of life lost world wide (WHO 2002b). In 2004, 42,636 people lost their lives in U.S. motor-vehicle crashes (NHTSA 2005a). In the U.S., traffic crashes cause more loss of human life (as measured in human-years) than almost any other cause – falling behind only cancer and heart disease (NHTSA 2005b).

Traffic injuries greatly exceed fatalities (the ratio is about 69:1), with about 2,889,000 occurring in 2003 in the U.S. alone (NHTSA 2005c). Traffic crashes also cause property damage to vehicles as well as public and private facilities. The present cost of U.S. crashes is estimated to be $230.6 billion annually; per capita this figure is over $800 per year (Blincoe et al. 2002). These costs do not include the cost of delays imposed on other travelers, which may be similar in magnitude. Schrank and Lomax (2002) estimated that approximately half (52-58%) of all traffic delays are due to non-recurring events, such as crashes. In urban areas, they estimated the non-recurrent cost to be $250 per year per capita. Thus, while vehicle and roadway design are improving, and growing congestion may reduce impact speeds, crashes are becoming more critical in many ways, particularly in societies that continue to motorize.

1.2 Fatal Crashes on Rural Roadways

As of 2003, rural roadways accounted for 76.3% of all publicly maintained lane mileage in the U.S. (USDOT 2004). Fatal crashes on these roadways comprise 57.4% of the U.S. total in 2004 (NHTSA 2005a). However, only 37.5% of vehicle miles traveled (VMT) takes place on these roadways.

Rural two-lane roadways are the largest single class of roads by mileage in the U.S. As shown in Table 1, crashes on rural two-lane roadways accounted for over 50% of all fatal U.S. crashes in 2004.

Table 1 Distribution of Fatal Crashes by Number of Travel Lanes and Functional Class in the U.S. in 2004 (NHTSA 2005a)

	Number of Travel Lanes									**Total**
	Blank	1	2	3	4	5	6	7+	Unknown	
Rural	10	148	19,523	543	1,545	54	39	9	94	21,965
Urban	22	382	9,259	2,149	2,955	408	461	116	282	16,034
Unknown	3	14	159	11	44	7	5	2	9	254
Total	35	544	28,941	2,703	4,544	469	505	127	385	38,253

Table 2 Distribution of Fatalities by Number of Travel Lanes and Functional Class in the U.S. in 2004 (NHTSA 2005a)

	Number of Travel Lanes									**Total**
	Blank	1	2	3	4	5	6	7+	Unknown	
Rural	12	177	22,118	643	1,787	63	48	10	117	24,975
Urban	24	402	10,044	2,320	3,215	437	516	128	295	17,381
Unknown	3	14	182	12	46	7	5	2	9	280
Total	39	593	32,344	2,975	5,048	507	569	140	421	42,636

1.3 Crashes on Washington State Rural Two-Lane Roadways

There are about 20,000 state-maintained centerline miles of public roadways in the State of Washington, over 67% of which are rural roads. The nation's Fatality Analysis Reporting System (FARS) crash data indicate that 63% of fatal

Washington crashes occur on rural roadways and over 37% of these are single-vehicle run-off-road events. The number of rural two-lane roadway fatalities in Washington in 2004 was 325, accounting for over half of that state's fatalities.

In 2004, 563 persons died on Washington State roadways – resulting in 1.01 fatalities per 100 million VMT, or 9.55 fatalities per 100,000 in population. NHTSA (2004) estimates annual Washington State crash costs to be $5.31 billion, a considerable loss for the state and its citizens.

Table 3 Distribution of Fatal Crashes by Number of Travel Lanes and Functional Class in the State of Washington in 2004 (NHTSA 2005a)

	Number of Travel Lanes						**Total**
	1	2	3	4	5	6	
Rural	6	286	15	11	0	1	319
Urban	4	107	17	47	4	3	182
Unknown	0	4	0	0	0	0	4
Total	10	397	32	58	4	4	505

Table 4 Distribution of Fatalities by Number of Travel Lanes and Functional Class in the State of Washington in 2004 (NHTSA 2005a)

	Number of Travel Lanes						**Total**
	1	2	3	4	5	6	
Rural	6	325	20	12	0	1	364
Urban	4	115	20	48	5	3	195
Unknown	0	4	0	0	0	0	4
Total	10	444	40	60	5	4	563

1.4 Factors Affecting Crash Occurrence and Their Significance

There have been numerous research efforts devoted to relating crash occurrence, environmental conditions, roadway design features, traffic conditions, vehicle characteristics, and driver attributes, as well as traffic laws and enforcement levels. (See, e.g., Solomon 1964; West and Dunn 1971; Burritt 1976; Dart 1977;

Baum et al. 1989, 1991; Brown et al. 1990; Chang and Paniati 1990; Garbacz 1990; Loeb 1995, 2001; Shankar et al. 1997; Abdel-Aty et al. 1998; Milton and Mannering 1998; Wong and Wu 1998; Chu 1999; Davis 2000; Garber and Ehrhart 2000; Newstead et al. 2001; Bédard et al. 2002; Khattak et al. 2002; Golob and Recker 2002; Furness 2003; Shankar et al. 2003; Prinsloo and Goudanas 2003; Banihashemi and Dimaiuta 2005; Kweon and Kockelman 2004, 2005; and Ma and Kockelman 2006.)

Raised speed limits are expected to result in more crash fatalities and injuries (e.g., Fowles and Loeb 1989; Levy and Asch 1989; and Snyder 1989). Based on the physical laws, a vehicle's kinetic energy is proportional to the square of its speed. That is to say, for a twofold increase in speed, kinetic energy will increase by a factor of four. Thus, speed plays an important role in predicting crash severity. However, several researchers have found that raised speed limits are associated with a reduction in crash fatalities and injuries (Lave and Elias 1994, 1997; and Houston 1999).

Usually, the roadway geometric design focuses on function, capacity, economy, and aesthetics while little attention is placed on traffic safety. According to Lamm et al. (1991), over 50 percent of highway fatalities occur on rural two-lane roads, and half of these fatalities occur on road segments with curves. Human factors are usually identified as a major cause of crashes; it is difficult to control and design for driver's psychological and physical conditions. Highway engineers cannot influence drunk driving and seat-belt usage. However, good geometric design should help to control traffic operating speeds and to reduce crashes.

Rural roadways are typically poorly lit or not lit at all. Insufficient light conditions may result in more crashes in the evening or nighttime.

In addition, hospital density and emergency response can affect the probability of a victim surviving serious injuries. Hospital density is the number of hospitals per square mile. Increased proximity to medical care facilities generally raises the likelihood of surviving a crash or suffering less serious injuries. Therefore, a high hospital density is expected to enhance highway safety, all else constant. According to the National Highway Traffic Safety Administration (NHTSA), emergency response time (i.e. the time of notification until the arrival at the scene) decreases as the wireless subscribership of a region increases. In rural areas, the EMS notification time is about 6.8 minutes in contrast to the urban EMS notification time of 3.7 minutes (NHTSA 2005d). Such extra minutes can be critical in resuscitating and protecting victims. Additionally, in some cases, regardless of whether the road is rural or urban, some states lack the necessary information to make decisions on potential highway safety solutions (USGAO 2004).

The findings related to speed limits, design, environmental factors, and traffic conditions are useful for roadway designers, law-enforcement agencies and policy makers because these stakeholders are in a position to affect a variety of factors, thereby reducing crash losses. For instance, policy makers and law-enforcement agencies might want to predict the effects (including monetary benefits and costs) of increasing speed limits, given design features and traffic conditions. Certainly, traffic engineers and roadway designers can better appreciate the impacts of their management and design choices on safety.

Moreover, insufficient funds and resources also have been hampering the rural highway safety development efforts because of rural highways' huge mileage, Although, many sates are permitted to use their funds for public road safety improvements, the usage of funds is restricted to the development of certain rural highway systems only. On the other hand, as local authorities are responsible for the maintenance of most of the rural highways, they might not be able to invest large amounts of money in improving these highways. Moreover, investing large amounts of resources on rural roads might be questionable due to concerns related to cost effectiveness, as these highways account for lower traffic volume as compared to urban highways (USGAO 2004).

1.5 Limitations of Existing Crash Occurrence Models

One way of addressing the highway safety issues related to rural highways is to reduce crashes by implementing applicable countermeasures. The other way is to reduce occupant injury severity, and this is especially important because of increased injury severities on rural highways. However, these two methods can be applied only if the relevant factors contributing towards the occurrence and increased severity of crashes are known.

There has been considerable research into the reasons behind crash occurrence (e.g., Hauer 1986; Persaud 1990; Hauer 1997; Abdel-Aty and Radwan 2000; Kweon and Kockelman 2005; and Lord et al. 2005). Crash records are commonly collected for both frequencies and severities on relatively homogenous road segments, and, usually, separate analyses are performed to model counts at different levels of severity.

There are several drawbacks to such methods. First, a separate analysis may result in a substantial decrease in efficiency for parameter estimation. For a particular roadway segment, unobserved factors (e.g., pavement quality, adjacent land uses, sight distance, and grade) may simultaneously influence traffic crash counts at different levels of severity. These unobserved factors create correlations that can improve prediction and estimation efficiency. Secondly, separate analyses of severity can be conducted only on roadway segments where crashes have occurred – i.e., segments with at least one crash. In practice, road segments with at least one crash may form a very biased sample of segments of interest. Moreover, an estimate of the marginal probability of observing a crash of a certain severity, conditional on first observing a crash, usually is not the same as the unconditional marginal probability of observing a crash at the same severity level. In practice, the latter should prove more useful to policy makers, transportation engineers, and others.

1.6 Models of Multivariate Count Data

To address the drawbacks in the separate, univariate analyses of correlated count data, some efforts have turned to multivariate specifications. Arbous and Kerrich (1951) examined crash proneness by assuming the number of crashes in several disjointed time intervals to be distributed as a mixture of independent Poisson distributions with a common gamma error term. This is known as the multivariate Poisson-gamma (MVPG) mixture model. Tsionas (1999, 2001) and Karlis (2003) adopted a special case of the multivariate Poisson (MVP) model with a single common additive term. Kockelman (2001) applied a multivariate negative binomial (MVNB) specification to analyze time- and budget-constrained activity demand. In her MVNB model, the same gamma error term was assumed for all demands generated by a single household. Though they recognize some correlations between counts, these model specifications are rather restrictive for real applications since they assume that all pairs of counts share the same correlation (due to the same multiplicative gamma term or additive Poisson term).

Aichison and Ho (1989) discussed a multivariate lognormal (MVLN) mixture of independent Poisson distributions. The MVLN distribution's rich covariance structure can be retained, while preserving non-negativity of crash rates and counts.

The multivariate Poisson-lognormal (MVPLN) distribution can offers a more flexible correlated count data process.

1.7 Limitations of Frequentist Approaches in Crash Occurrence Modeling

The MVP specifications discussed above are relatively hard to implement because of difficulties in parameter estimation using frequentist approaches. Karlis (2003) developed an Expectation Maximization (EM) algorithm for estimating his special case of MVP regression models. However, this algorithm only finds the parameter point estimates and does not provide full information on their distribution. Moreover, in dealing with the multidimensional integrals, frequentist approaches appear to be too complicated for most practitioners to understand. In very recent years, Bayesian methods have found applications in the traffic crash analysis (see, e.g., MacNab 2003; Miaou and Song 2005; and Ma and Kockelman 2006).

1.8 Advantages of Bayesian Techniques in Crash Occurrence Modeling

1.8.1 Hierarchical Modeling via Mixtures

Hierarchical modeling via mixtures of distributions (e.g., a Poisson rate following a gamma distribution) provides a great degree of flexibility in data analysis. For example, Qin et al. (2003) employed scale mixture methods to model skewed and heavy-tailed data while relaxing the assumption of homoscedasticity.

1.8.2 *Parameter Estimation via Markov Chain Monte Carlo Simulation (MCMC)*

As mentioned, parameter estimation in multivariate count data models can be complicated using frequentist approaches. In a Bayesian approach, the unknown parameters are assumed to be random variables. Data augmentation techniques can estimate both parameters of interest and nontrivial variables, such as latent variables (e.g., total number of crash counts under the situation of underreporting). Therefore, knowing the kernel density of all unknown variables is enough to estimate the parameters of interest (which form some of the unknown variables). In a frequentist approach, the likelihood function has to be transformed into unconditional distributions by integrating out the non-trivial variables. Sometimes, the multidimensional integrals can be very complicated and hard to implement. The MCMC simulation techniques, such as Gibbs sampling and M-H algorithms, can be very powerful in estimating such complicated models.

1.8.3 *Incorporate Prior Beliefs with Data*

In data analysis, one may have some sense of the parameters of interest before looking at the data. Bayesian statistics provides a coherent way to incorporate prior beliefs about parameter values via Bayes' theorem. In contrast, frequentist approaches have no way to incorporate such useful information into the modeling process. Prior beliefs can be very useful to the problem solution, particularly when only a small data sample is available.

1.8.4 *Inference*

In simple terms, frequentists assume that there is an unknown but fixed set of parameters. Bayesians assume that unknown parameters are by their nature random variables and then use probability theory to quantify this uncertainty. Frequentists consider all possible data sets generated by the unknown fixed parameters, while Bayesians take data as given, and consider all possible values of unknown parameters. For example, one may wish to estimate a population mean. Frequentists believe that the population mean is fixed, but unknown and unknowable. They estimate the population mean using the sample mean. Given the distribution of the sample mean, the resulting 95% confidence interval is interpreted as follows: 95% of such intervals will contain the population mean, assuming each interval is obtained from a different random sample. However, Bayesians take the data as given, and assume that the population mean follows a posterior distribution obtained by combining both the sample mean and their prior beliefs about the population mean. Therefore, in a Bayesian paradigm, the population mean is estimated as a distribution. The credible interval in Bayesian statistics can be interpreted as a so-called interval that contains the true population mean with 95% probability.

1.9 Study Objective

The objective of this research is to model traffic crash frequency by severity simultaneously with multivariate Poisson-lognormal models. This approach addresses some questions that are difficult to answer by estimating them separately. The parameter estimation will be done within the Bayesian paradigm. A Gibbs Sampler and M-H algorithms are constructed to assist the Bayesian statistical inference. The methodology is empirically applied to investigate crash frequency by severity using crashes which occurred on Washington State rural two-lane highways in the Puget Sound region. The results lend themselves to recommendations for highway safety treatments and design policies.

1.10 Organization

The rest of the dissertation is organized as follows. Chapter 2 conducts an extensive review of early research on crash occurrence modeling. Chapter 3 introduces the basics of Bayesian statistics. This chapter also covers the two most widely used Markov chain simulation methods: the Gibbs sampler and the M-H algorithms. Chapter 4 starts from UVP regression models and then moves to MVP models. In addition, this chapter also establishes two MCMC simulation procedures to estimate the unknown parameters in the MVPLN model. Chapter 5 illustrates the merits of MVPLN models to investigate the correlated crash data on Washington State rural two-lane highways in the Puget Sound region. A cost-benefit analysis of raising speed limits is implemented based on the results from the MVPLN model. Chapter 6 discusses the limitations and future extensions of MVPLN model and offers conclusions.

1.11 Summary

Roadway safety is a major concern for the general public and public agencies. Roadway crashes claim many lives and cause substantial economic losses each year. The situation is of particular interest on rural two-lane roadways, which experience significantly higher fatality rates than urban roads. With recent advancements in crash modeling and Bayesian statistics, this dissertation will develop, calibrate, and apply multivariate count data models.

CHAPTER 2 LITERATURE REVIEW

There has been extensive research performed over the past decades to investigate traffic crash occurrence as related to many influencing factors such as roadway geometrics, environmental and traffic conditions, highway user attributes, and vehicular- and crash-related characteristics. Various statistical approaches have been utilized in the process of developing an association between crash occurrence and those factors. Research on the frequency of traffic crashes can be classified into two major streams: univariate count data models and multivariate count data models.

2.1 Univariate Count Data Models

In the first stream of research, the frequency of traffic crashes by severity is usually separately modeled using univariate count data models[1]. Poisson and negative binomial (NB) regression models based on counting processes have found extensive use in studies of crash frequency over the last 20 years. For example, in traffic safety analysis, Poisson and the related negative binomial regression models[2] (where the latter are deduced from the combination of a Poisson process with a gamma distribution of rates to allow for unobserved heterogeneity and allow for "overdispersion" in the conditional distributions)

---

[1] Some research modeled crash rates as a continuous number and applied linear regression models to investigate the crash occurrence (see, e.g. Shah 1968; and Mulinazzi and Michael 1969). It is a straightforward method that models crash rates for a relatively homogeneous segment. Although linear regression models are relatively easy in estimating and interpreting the results, research has suggested that linear regression models have many drawbacks, such as lack of distributional properties to describe random, sporadic, crash count events. The linear regression models for crash counts can, thus, result in inconsistent parameter estimates.

[2] The Poisson model has been criticized because of its implicit assumption that the variance of the dependent variable equals its mean. Many extensions of the Poisson model that relax this assumption have been proposed by many econometricians in 1980s. The negative binomial model is one kind of such extensions of the Poisson model by relaxing the assumption of equi-dispersion.

have been the primary device used for investigating the associations between crash frequency and several factors such as traffic volume, access density, posted speed limit and number of lanes. The following sections describe previous research efforts.

2.1.1 Poisson Regression Models

In 1992, Miaou et al. employed a Poisson regression model to investigate the relationship between truck crashes and roadway geometric design variables such as annual average daily traffic per lane, horizontal curvature, vertical grade, and shoulder width. The data set they used was collected from rural interstate highways from 1985 to 1987 through the Highway Safety Information System (HSIS), operated by the University of North Carolina Highway Safety Research Center (HSRC) and LENDIS Corporation, and administrated by the Federal Highway Administration (FHWA). Besides quantifying the relationships between truck crashes and the key variables listed above, they also suggested that better truck exposure data and additional important variables could improve the goodness-of-fit of the Poisson specification.

Miaou and Lum (1993) applied the Poisson regression model to evaluate the effects of highway geometric design on truck crash occurrence. They also estimated and quantified the uncertainties of the expected reductions in truck crashes due to improvements of highway geometric design. Truck crash data from 1985 to 1989 on Utah rural interstate highways was obtained through the HSIS to illustrate the use of the proposed methods.

Kumara and Chin (2005) adopted a Poisson specification to describe the reported crash counts and a probit model to represent the reporting mechanism to investigate the crash occurrence at three-legged signalized intersections. After taking into account the underreporting, they found that some geometric, traffic and traffic control factors significantly affect the crash occurrence; for example, left-turn volumes and shorter sight distance may increase the crash occurrence at the three-legged signalized intersections.

There has been a considerable interest in Poisson regression models that capture the non-negativity and discrete characteristics in traffic crash data. (See, e.g., Abdel-Aty and Pemmanaboina 2005; Khattak et al. 2004; Daniel and Chien 2004; Hallmark and Muller 2004; Oh et al. 2004; Whitfield and Whitfield 2004; Chipman et al. 2003; Daniel et al. 2002; Khattak and Knapp 2001; Quddus et al. 2001; Ivan et al. 2000; Abdel-Aty and Radwan 1999; Ivan et al. 1999; Scuffham and Langley 1997; and Laberge-Nadeau et al. 1996.)

2.1.2 Negative Binomial Regression Models

A well-known limitation of the Poisson distribution is its equi-dispersion; i.e., its mean equals its variance. The possibility of overdispersion (its variance is greater than its mean) is always a concern in modeling traffic crashes and may produce incorrect inference. Negative binomial distribution has been widely used as an alternative to relax the equi-dispersion constraints imposed by the Poisson model.

Johansson (1996) examined the effects of speed limit reduction on the crashes which occurred on Swedish motorways. He applied the extensions of the Poisson and negative binomial regression model to address the serial correlation as well as overdispersion. His results suggested that lowering speed limit was predicted to reduce the number of crashes involving minor injuries and vehicle damage.

Vogt and Bared (1998) investigated the crash occurrence for rural two-lane roadway segments as well as intersections in the states of Minnesota and Washington. They employed a negative binomial specification and its extension to control for many important variables such as traffic, horizontal and vertical alignments, lane and shoulder widths, roadside hazard rating, channelization, and the number of driveways. Their results indicated that segment crash occurrence primarily depends on roadway features, while intersection-related crashes are mainly determined by traffic.

Vogt (1999) investigated the crashes occurring on rural roadway intersections in the states of California and Michigan for the years 1993-1995. The crash data was obtained through the HSIS containing the type of intersection, traffic volume, roadside hazard rating, the number of driveways, channelization, intersection angles, and speed limits. He employed the negative binomial regression models to address the overdispersion in the crash count data.

Miaou (2001) employed the negative binomial regression model to estimate vehicle roadside encroachment rates for rural two-lane undivided roads. His data was obtained from the seven states cross-section data base of the FHWA. His model controlled for AADT, lane width, horizontal curvature, and vertical grade.

Zegeer et al. (2001) employed Poisson and negative binomial regression models to examine the effects of the presence of marked crosswalks on pedestrian crashes. The data sets used involved five years of pedestrian crashes at 1000 marked crosswalks and 1000 unmarked comparison sites, all of which are unsignalized, uncontrolled intersections. Besides the presence of marked crosswalks, they controlled for traffic volume, pedestrian exposure, the number of lanes, the type of median, speed limits, and other site variables. Their study showed that raised medians could significantly reduce pedestrian crash rates on multilane roads, and older pedestrians tended to have higher crash rates.

2.1.3 Zero-Inflated Count Data Models

Poisson regression models provide a standard framework for the analysis of crash count data. However, count data are often overdispersed relative to the Poisson distribution. One frequent manifestation of overdispersion is that the incidence of zero counts is greater than expected for the Poisson distribution and this is of interest because zero counts frequently have special status. For example, a segment of roadway may have no crashes during the study period, either because it is so well designed that no crash would occur, or simply because no crash has yet occurred on it by chance. This is the distinction between structural zeros, which are inevitable, and sampling zeros, which occur by chance. In recent years, there has been considerable interest in models for count data that allow for excess zeros, such as zero-inflated Poisson (ZIP), and zero-inflated negative binomial (ZINB) models. Zero-inflated (or zero-altered) count data models[3] have been applied to capture the apparent excess zeros, which commonly exist in crash data sets, such as segment-based crash data.

As an extension of a standard Poisson regression model, zero-inflated Poisson regression models have gained considerable recognition in analyzing crashes. The zero-inflated Poisson regression models allow one to distinguish safe roads with a zero probability of crash occurrence from unsafe roads which have recorded no crash during the study period due to randomness.

Miaou (1994) investigated the relationship between truck crash occurrence and geometric design of road segments using Poisson regression, ZIP regression, as well as negative binomial regression models. He examined the performance of

---

[3] Zero inflated count models basically assume two sources for zero outcomes. One source is because a segment of roadway happens to have zero crashes due to randomness. Another source is that that segment of roadway will never have crashes regardless of the characteristics of the roadway. This assumption is apparently not appropriate for modeling crash occurrence since no roadways will never have crashes given a long period of time.

these models from the following five perspectives: (1) estimated coefficients, (2) overall goodness-of-fit, (3) estimated relative frequency of truck crash occurrence across road segments, (4) analyzed sensitivity to the short road segments, and (5) predicted total number of truck crashes.

Shankar et al. (1997) applied a zero-altered counting process to analyze road traffic crashes. In the zero-altered count process, they tried to distinguish roadway segments that are truly safe from those that are not, but happen to have zero crashes during the period of investigation. They estimated the ZIP as well as ZINB regression models using crashes on principal arterials in western Washington for the years 1992 to 1993.

Garber and Wu (2001) applied the Poisson, negative binomial, ZIP, and ZINB regression models using traffic data obtained from the Smart Travel Lab at the University of Virginia. The controlled variables include volume, speed, occupancy, curvature, exposure, and standard deviation of speed at the time of crash occurrence. Their results indicated that negative binomial and ZIP are the preferred modeling methods in their case.

Lee and Mannering (2002) applied ZIP and ZINB regression models to examine single-vehicle run-off-roadway crashes occurring on a section of highway in Washington State. Their focus was on the effects of roadside features on crash occurrence. Their results showed the following treatments which can be implemented to reduce run-off-roadway crashes: 1) avoiding cut side slopes, 2) decreasing the distance from outside shoulder edge to guardrail, 3) decreasing the number of isolated trees along roadway sections, and 4) increasing the distance from outside shoulder edge to utility poles.

Kumara and Chin (2003) employed the zero-inflated negative binomial (ZINB) specification to address the problem of excess zero crash counts. The ZINB model allows one to distinguish safe intersections from those at which no crash happened due to randomness.

Rodriguez et al. (2003) examined the effects of truck-driver wages and working conditions on crash occurrence. They adopted a zero-inflated Poisson regression model to quantify the impacts of human capital and occupational factors such as pay, job tenure, and percentage of miles driven during winter. Their results indicated that higher pay rates and pay increases are associated with lower crash rates.

Shankar et al. (2003) estimated the NB and ZIP models using data on reported crashes involving pedestrians and motorized traffic in Washington State. Their results based on the data suggested that zero-inflated count models are promising approaches to explore pedestrian-related crashes.

Qin et al. (2004) analyzed the relationship between crashes and exposure measures using crashes occurring on two-lane highway segments in Michigan. They used ZIP specification to model crash counts as a function of AADT, segment length, speed limit and roadway width. They found that the relationship between crash counts and AADT is nonlinear and is different from the relationship between crash counts and segment length.

Lord et al. (2005) provided some guidance in selecting appropriate count data models to predict crashes in terms of statistical fit and theoretic foundation. For example, they would suggest using zero-inflated models for datasets with a preponderance of zeros.

There also has been considerable interest in models that allow for excessive zeros, such as zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) regression approaches. (See, e.g., Liu et al. 2005.)

2.1.4 Panel Count Data Models

To address the heterogeneity issue across individuals (e.g., a segment of roadway, an intersection), some recent studies have analyzed panel data sets using panel count data models such as random effect negative binomial (RENB) and fixed effect negative binomial (FENB) regression models.

Chin and Quddus (2003) proposed the use of random effect negative binomial (RENB) models to account for unobserved heterogeneity and serial correlation in the crash data. They conducted an empirical study by examining the relationship between crash occurrence and the geometric, traffic and control characteristics of signalized intersections in Singapore.

Kumara et al. (2003) utilized the random-effect negative binomial model to examine crashes occurring at signalized intersections. Their results, based on the Singapore data, showed the following factors resulting in more crashes: 1) uncontrolled left-turn, 2) insufficient sight distance, 3) a large number of signal phases, 4) permissive right-turn phases, 5) horizontal curves, and 6) total and left-turn volumes.

Noland (2003) analyzed the effects of road infrastructure improvements on traffic fatalities and injuries using a fixed-effects negative binomial regression model to account for heterogeneity in the data. Noland's negative binomial models accounted for various variables at the aggregate level, total lane miles, the proportion of lane miles in different categories of roadways, average number of lanes for each road category, lane widths, as well as demographics. Their results suggested the following factors contributing to overall reductions in fatalities: 1) increased seat-belt use, 2) reduced alcohol consumption, and 3) increases in medical technology.

Kumara and Chin (2004) estimated a fixed-effect negative binomial (FENB) model to examine the effect of socioeconomic and infrastructure factors on fatal traffic crashes in Asian Pacific countries. The data set was collected from 41 countries for the years of 1980 to 1994. Their FENB model results showed the following factors resulting in more fatal crashes: 1) size of road network, 2) gross national product per capita, 3) population, and 4) the number of registered vehicles. Additionally, they found that fatal crashes decrease with time for all countries.

Kweon and Kockelman (2004) employed the fixed-effects and random-effects Poisson and negative binomial regression models to examine the impacts of the repeal of the National Maximum Speed Limit (NMSL) of 55 mi/h law using crashes occurring on Washington State highways for the years of 1993 to 1996. They relied on segment-based panel data for thousands of roadway segments averaging just 0.1 mile in length. They predicted crash counts based on a number of design variables (such as degree of curvature and vertical curve length); however, their data set contained only five interstate highways and did not consider speed choices.

Kweon and Kockelman (2005) examined the safety effects of speed limit changes using panel count data models as well as non-panel count data models. Based on the two information criteria (AIC and BIC), the RENB model proved the most effective for modeling fatality, fatal crash, injury, injury crash, property-damage-only (PDO) crash, and total crash occurrence.

2.1.5 Other Count Data Models

Fridstrøm et al. (1995) investigated the road traffic crashes using a generalized Poisson regression model. Based on the analysis of a four-country, segmented data base, they concluded that randomness and exposure played a dominant role in predicting crash occurrence.

Unlike previous research, Ulfarsson and Shankar (2003) adopted a negative multinomial specification to analyze the panel median crossover crash data and to explore the section-specific serial correlation across time. Their results showed that the negative multinomial regression model is significantly better than the negative binomial and random-effects negative binomial regression models in terms of fit. Besides, the negative multinomial specification provides more intuitive results.

Pernia et al. (2004) employed the Poisson, negative binomial and lognormal regression models to investigate the two-way left-turn lane (TWLTL) median treatment on crash occurrence. They used traffic crashes in Florida for the years of 1996 to 1998 as the data sample which also included some variables such as access density, posted speed limits, and number of lanes.

There have been many other efforts devoted in modeling crash occurrence using some extensions of count data models. (See, e.g., Djauhari 2002; Koorey 2005; and Miaou et al. 2005.)

2.1.6 Limitations of Univariate Approaches

The methods discussed above more or less ignored the following issues: (1) interdependence may exist between crashes of different levels of severity for a specific segment of roadway, and (2) road geometric design features, road use, and environmental conditions may have distinct effects on crashes of different severity.

Specifically, there are at least three drawbacks making a separate analysis inappropriate for addressing this problem. First, a separate analysis may result in a substantial decrease in efficiency for parameter estimation[4] since the relationship between crash severity and frequency is ignored in this setting. Second, a separate analysis of severity can only be conducted on road segments where crashes occurred, i.e., the segments with at least one crash. In practice, the road segments with at least one crash may consist of a biased sample of the roadway of interest. Moreover, an estimate of the marginal probability of observing a certain severity level conditional on observing a crash usually are not the same as the unconditional marginal probability of observing a crash at the same severity level. For example, there are 1000 crashes (e.g., 600 property-damage-only (PDO) crashes, 350 injury crashes, and 50 fatal crashes) in a city, which has a population of 1 million. For an arbitrary person living in the city, the probability of becoming involved in a fatal crash is 50/1,000,000=0.5%. If we know this person was involved in a crash, the probability of that crash being fatal is 50/1000=5%. Obviously, the marginal probability, 5%, is far larger than the unconditional probability of 0.5%. In practice, the latter is more useful to policy makers, transportation engineers, etc.

2.2 Multivariate Count Data Models

Multivariate count data models, which can address the above three issues, are the focus of the second stream of research. In the second stream of research, the

[4] When the interdependence exists among crash counts by severity on a segment, the estimates of standard errors of the regression coefficients will be underestimated because each observation actually contributes less information than it is assumed to contribute under the assumption of independence. The consequence is that the t-ratios will be inflated and incorrect statistical inferences will be made.

frequency of traffic crashes by severity is simultaneously modeled using multivariate count data models such as multivariate Poisson, multivariate zero-inflated Poisson, and Poisson-lognormal regression models. Ladron de Guevara and Washington (2004) investigated the simultaneity of fatality and injury crash outcomes. Bijleveld (2005) estimated the variance-covariance structure between the number of crashes and the number of injuries.

2.3 Bayesian Approaches

The multivariate Poisson specification is challenging to implement because of difficulties in the estimation of parameters. Karlis (2003) developed an EM algorithm for estimating multivariate Poisson regression models.

In more recent years, Bayesian methods have found applications in the traffic crash analysis. Christiansen et al. (1992) developed a hierarchical Poisson model for investigating traffic crashes. MacNab (2003) investigated traffic crash and injury surveillance data using a Bayesian hierarchical model. Miaou and Song (2005) employed the Bayesian methodologies in ranking the sites for engineering safety improvements. Liu et al. (2005) used a hierarchical Bayesian framework to estimate ZIP regression models to develop safety performance functions (SPF) for two-lane highways.

However, to our knowledge, none of the studies applied Bayesian methods to estimate multivariate Poisson models in traffic safety analysis. The Bayesian methods give an approximation to the whole posterior distribution of parameters of interest, as opposed to the EM algorithm which only gives the mode of parameters of interest.

2.4 Previous Models of Rural Two-lane Roadway Crash Occurrence

Crashes on rural roads accounted for more than 60% of US 42,815 traffic death in 2002 (USGAO 2004), while rural roadways carry only 40% of US traffic each year. There are more than 3.9 million miles of roadway in the US, about 50% of which are in rural areas. The fatality rate on rural roadways was 2.29 per million vehicle miles traveled (MVMT) in 2002, while the fatality rate was 0.97 per MVMT for urban roadways (USGAO 2004).

During the past two decades, the number of drivers and vehicles has risen steadily. The number of traffic fatalities in 2002 was 2.6 percent less than the number in 1982, while the number of fatalities on rural roadways increased slightly. The increase is possibly due to a variety of reasons, including the aging of the roadway surfaces and an evolution in population settlement patterns. Given their relatively low traffic volumes, it is difficult to justify the major construction necessary to redesign or rehabilitate rural roadways. A strong understanding of the benefits of such actions is needed. To this end, numerous efforts have been devoted to investigating crash occurrence as related to other factors, such as average annual daily traffic (AADT), shoulder width, degree of curvature, vertical grades, and speed limits. (See, e.g., Zeeger et al. 1981; Vogt 1998, 1999; Ogden 1997; Fitzpatrick 2002; and Lamm et al. 2002.)

Fitzpatrick et al. (2003) examined the different characteristics between urban and rural crashes in Texas. They found that urban areas had 46% of possible injury (C-level) crashes, compared with 28% in rural areas; urban areas had less than 20% of single-vehicle crashes, compared with 50% of single vehicle crashes in rural areas; a higher percentage of rural crashes happened at night, and were twice as likely to involve striking a fixed object compared to urban crashes.

Carlson et al. (2005) examined daytime high-speed passing maneuvers using surveys on a rural two-lane, two-way highway in Texas. They found that the AASHTO Passing Sight Distance (PSD) model would provide inadequate PSD for speeds overtaken greater than those assumed.

Chen and Cottrell (2005) evaluated the effects of centerline rumble strips (CLRS) on rural two-lane and undivided multilane highways to reduce cross-over-the-centerline (COCL) crashes. A benefit-cost analysis they conducted showed a b/c ratio of at least 7.6 for installing CLRS.

Two-lane rural roads are perilous everywhere. Passing occurs in the opposing lane, turning vehicles affect flow in both directions, access is minimally controlled (with driveways introducing slow-moving vehicles), and head-on collisions are not uncommon. Streff and Kostyniuk (1997) found that most fatal crashes in rural areas of Michigan occurred on two-lane collectors and local roads with speed limits of 55 miles per hour. And Huang et al. (2001) found that rural two-lane roads in North Carolina (from 1993 to 1997) were associated with the highest crash severities.

Horizontal curves also present special dangers. Viner's (1995) analysis of a variety of roadway types across the U.S. noted that the areas of highest risk for run-off-road crashes are on rural two-lane roads in the outer lanes of horizontal curves. Using data from two-lane rural roads in both the U.S. and Germany, Lamm et al. (2002) regressed operating speeds and crash rates on various design features of horizontal curves, including rate of change in curvature, curve length and superelevation, lane and shoulder widths, sight distance, grades, and traffic volume. While they found only curvature change rate to be a statistically significant contributor to crash rates, their sample size was quite limited.

In reality, a great many factors likely are at play (e.g., pavement and weather conditions). Intersections also perform very differently from basic roadway segments. Using data from Minnesota and Washington State, Vogt and Bared's (1998) negative-binomial crash models for rural two-lane roadway segments and intersections controlled for traffic levels, roadside hazard ratings, degree of curve, grades, lane and shoulder widths, channelization and driveway frequency. Their results suggest that basic roadway segment crashes depend mainly on design features, while intersection crashes depended largely on traffic levels.

Vogt's (1999) investigation of rural intersections in four states controlled for a variety of factors, including the number of approach legs, control type (signalized or stop-controlled), the number of approach lanes (four and two), the use of channelization, the angle of intersection, left-turn and truck percentages, and speed limits. His negative binomial model for predicting crash counts suggested that virtually all variables were statistically significant (thanks in part to a very large sample of over 20,000 intersections.

Work zones also tend to be hazardous locations. Venugopal and Tarko's (2000) negative binomial models for crashes in work zones on rural two-lane highways controlled for AADT, work-zone length, project duration, and project cost. They found that crash rates in such zones depend greatly on geometrics, traffic control, and traffic levels.

One-vehicle crashes are distinct from multi-vehicle crashes. Ivan and Wang's (2000) Poisson regression models for rural two-lane crash counts found that daylight conditions, a higher volume-to-capacity ratio, narrow shoulders, extensive passing permission, and the presence of many intersections tended to reduce the number of single-vehicle crashes. For multi-vehicle crashes, the number of intersections also had a negative effect, but daylight conditions and driveway density had positive effects.

Information on adjacent activity types/land uses also aids prediction. Using a logistic regression model to predict crash counts on two-lane undivided rural and suburban roadways in Strafford County, New Hampshire, Ossenbruggen et al. (2001) found that "village" sites were safer than residential and shopping sites. In addition, something as simple as striping can play an important role: Miller's (1992) review of the literature suggested that edgelines on rural two-lane highways reduce crashes by 8 percent. In contrast, something as fundamental as speed limits may not play a real role: Najjar et al.'s (2002) study of rural two-lane Kansas highways found no statistically significant increases in system-wide crash rates after speed limits were raised. However, some Kansas sections did experience crash rate (and fatal crash rate) increases and these mainly were rural two-lane highways.

There is also research on the crash countermeasures and safety improvement methods on rural two-lane roads. Agent et al. (2001) analyzed a sample of 150 fatal crashes that occurred on a rural two-lane road. Based on their analysis, they recommended several crash countermeasures and evaluated the potential effectiveness of these countermeasures. The specific countermeasures were divided into roadway and non-roadway areas (e.g., centerline rumble strips, clear zone improvements, and regular pavement inspection); the non-roadway area was further divided into legislation, enforcement, and education/training. Agent and Pigman's (2001) before-after analysis of 25 locations with added lanes and 24 locations with wider lanes and shoulders estimated crash rates (total, as well as injury and fatal crash rates) to fall by over 50 percent in both cases. Council and Stewart's (2000) cross-sectional models predicted that conversion of rural roads from undivided two-lane designs to divided four-lane designs reduced crash rates by roughly 50 percent. In contrast, conversion without directional division was found to have only a minor impact.

Clarke and Sarasua (2003) and Lacy (2002) analyzed crash data for South and North Carolina. The data was collected primarily on rural two-lane roads, and their Bayesian approach resulted in crash reduction factors (CRFs) for specific countermeasures. Twenty-five countermeasures spanning pavement, roadway, roadside, and lighting improvements, as well as regulations, were examined. Clear-zone improvements and guardrail additions had the greatest effects on crash rates; geometric realignments and speed limit enforcement also had major effects. Fitzpatrick and colleagues' recent work for TxDOT (Fitzpatrick and Brewer 2002; Fitzpatrick et al. 2001 and 2002) reviewed a lot of literature regarding design treatments for rural, low-volume two-lane roads; however, they did not analyze any data or provide design or treatment recommendations.

While cross-classification of crash rates according to several variables at a time offers a glimpse of relationships between design practices, traffic conditions, and crash consequences, more sophisticated, multivariate methods are the standard (see, e.g., Duncan 1998; Golob et al. 1987; Kockelman and Kweon 2002; Kweon and Kockelman 2003; Quddus et al. 2002; Abdel-aty et al. 1998; Farmer et al. 1997; and O'Donnell and Connor 1996). For example, Duncan et al. (1998) investigated the effects of several factors on injuries endured during two-vehicle rear-end crashes using the ordered probit modeling technique. Their results suggested that darkness, high speed differentials, higher speed limits, and excessive grades contributed to substantially more severe crashes. Using the same statistical specification, Ma and Kockelman (2004) found that higher design speeds[5] (holding speed limits fixed) and speeding contribute to injury severity, while lighting and pavement wetness appear to play no role. And Kweon and Kockelman (2001) found that vehicle type can be key in protecting (or not protecting) occupants and crash partners.

There many other studies on traffic safety of rural two-lane roads; for example, Jaarsma et al. (2005) investigated passing bays for slow-moving vehicles; Schurr et al. (2004) modeled appropriate design speed for horizontal curves approaching stop-controlled intersections; Fitzpatrick (2004) summarized some important documents for users to conduct safety studies and select treatments; Gattis et al. (2003) proposed an alternate passing lane to improve the safety of rural two-lane roadways; Persaud et al. (2003) evaluated the effects of centerline rumble strips on crash reductions; Ranck (2003) applied safety and operational effects of highway design features to two-lane rural highways; Wooldridge et al. (2003) discussed geometric design consistency; Kindler et al. (2003) developing the Intersection Diagnostic Review Module (IDRM).

Notably, prior research on crash severity has not controlled for roadway design and traffic conditions, examining only a subset of severe crash causes. Thanks to the compilation of new, more extensive data sets, these analyses provide additional control variables, such as driveway density and the presence of improved roadside. Their inclusion allows for new insights and recommendations.

[5] Design speed is the minimum speed for which the road is designed. In other words, the most restrictive geometric features which make up a highway's design such as degree of curvatures and vertical grades should safely allow an operating speed at least equal to the design speed.

2.5 Summary

A review of early research on crash occurrence shows:

- Most crash occurrence modeling efforts are based on univariate count data models such as Poisson, NB and their extensions to address issues of overdispersion and/or heterogeneity; few of them address the issue of correlated counts of crashes at different levels of severity.
- Although empirical Bayesian approaches have seen wide application in traffic safety fields, the full Bayesian analysis of crash data has not found many applications in the field.
- Many studies on the safety of rural two-lane roadways have focused on different aspects of treatments, exhibiting a great deal of inconsistency in their findings.

CHAPTER 3 BAYESIAN STATISTICS

This chapter first briefly discusses Baye's theorem and Markov chain Monte Carlo (MCMC) simulation techniques such as the Gibbs sampler and Metropolis-Hastings (M-H) algorithms. Then, univariate Poisson regression models are introduced with a discussion of potential challenges to the application of correlated count data. After that, the discussion focuses on addressing the problems of correlated counts using multivariate Poisson-lognormal (MPLN) regression models. Finally, a Gibbs sampler and M-H algorithms are constructed to estimate the unknown parameters in the model.

3.1 Bayesian Theory

The essential part of the Bayesian paradigm is Bayes' theorem. Let (y, X) denote data, where y is a vector of dependent variables and X is a matrix of explanatory variables. Let θ be a vector of unknown parameters that, along with X, statistically determines the distribution of y values. Notationally, Bayes' theorem can be written as

$$\pi(\theta|y, X) = \frac{\pi(\theta|X)\pi(y|\theta, X)}{\pi(y|X)} \propto \pi(\theta|X)\pi(y|\theta, X) \qquad (1)$$

where $\pi(\theta|X) = \pi(\theta)$ (i.e., X is not relevant to θ) is the prior distribution of the random parameters θ to serve as one's beliefs about the parameters before taking a look at the data; $\pi(y|\theta, X)$ is the conditional probability of the observed response given a particular set of values for θ and X, commonly referred to as the likelihood function of y; and $\pi(y|X)$ is the marginal likelihood of y. Usually, $\pi(y|X)$ is dropped from the equality part of Equation (1) because this term does not explicitly involve the unknown parameters.

$\pi(\theta|y,X)$ is the density of the posterior distribution of θ, standing for one's updated prior beliefs about the parameters after incorporating the new information contained in data (y,X).

Bayesian statistics is an approach to statistical problems that aims to combine two pieces of information optimally: the beliefs or information one has before taking a look at the sample data and the information extracted from the sample data. Bayes' theorem is basically a rule for combining these two sources of information into the new information about the parameters or hypotheses of interest.

Figure 1 schematically shows how the two pieces of information are incorporated into the posterior distribution of θ via Bayes' theorem.

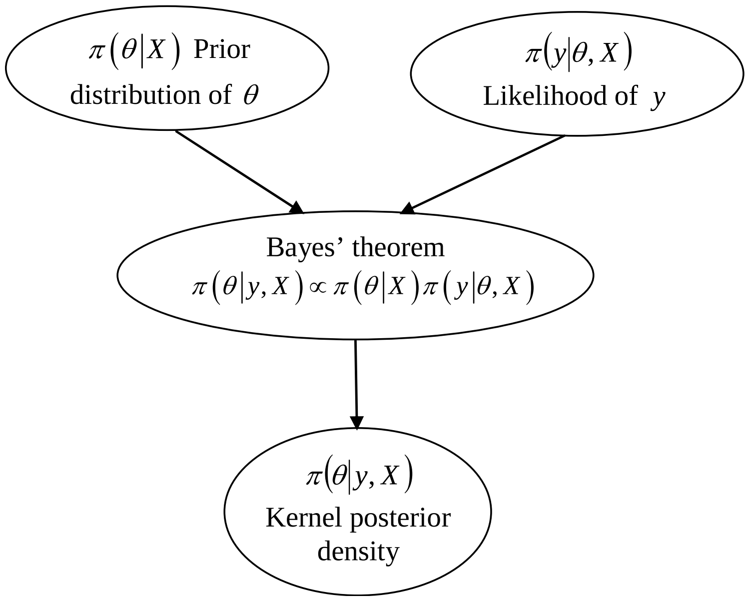


Figure 1 Schematic Diagram of Bayes' Theorem

The relative weight of the two pieces of information depends on their precision[6]; that is, the smaller the variance of the distribution, the larger its role in determining the posterior distribution. Let us explain this using an example.

Suppose $y = (y_1, y_2, \ldots, y_n)'$ is an independent random sample from a Poisson distribution with mean λ. Since the gamma distribution is a conjugate prior[7] for the Poisson parameter, let us assume that the prior distribution of λ is a gamma with parameters $\alpha \geq 0$ and $\beta \geq 0$. Applying Bayes' theorem, one can write the posterior distribution of λ as (Winkelmann 2003):

[6] Here, precision is defined as the inverse of the variance (Gelman et al. 2004).

[7] $\pi(\lambda)$ is said to be a conjugate prior if the posterior density $\pi(\lambda|y) \propto \pi(\lambda)\pi(y|\lambda)$ has the same class of distributions as $\pi(\lambda)$ (Lee 2004).

$$\pi(\lambda|y) \propto \pi(\lambda)\pi(y|\lambda) \propto$$

$$\left[\prod_{i=1}^{n} e^{-\lambda}\lambda^{y_i}\right]\frac{\alpha^{\beta}}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\lambda\beta} \propto e^{-\lambda(\beta+n)}\lambda^{\alpha+\sum_{i=1}^{n}y_i-1} = e^{-\lambda(\beta+n)}\lambda^{\alpha+n\overline{y}-1} \tag{2}$$

As can be seen from Equation (2), the posterior distribution of λ is a gamma with parameters $\tilde{\alpha} = \alpha + n\overline{y}$ and $\tilde{\beta} = \beta + n$. The mean of the prior gamma distribution is obtained using $E_0(\lambda) = \alpha/\beta$. The mean of the posterior distribution of λ is similarly obtained as:

$$E(\lambda|y,\alpha,\beta) = \frac{\alpha + n\overline{y}}{\beta + n} = \frac{\beta}{\beta + n}E_0(\lambda) + \frac{n}{\beta + n}\overline{y}$$

$$E(\lambda|y,\alpha,\beta) = \frac{\beta^2/\alpha}{\beta^2/\alpha + \beta n/\alpha}E_0(\lambda) + \frac{\beta n/\alpha}{\beta^2/\alpha + \beta n/\alpha}\overline{y} \tag{3}$$

The variance of the prior distribution of λ is $\sigma_0 = \alpha/\beta^2$. The precision of a random variable is the inverse of its variance; that is, the precision of λ's prior is $\tau_0 = \beta^2/\alpha$.

$$Var(\overline{y}) = Var\left(\frac{1}{n}\sum_{i=1}^{n}y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}Var(y_i) = \frac{1}{n^2}n\frac{\alpha}{\beta} = \frac{\alpha}{n\beta} \tag{4}$$

Hence, the precision of $\bar{y}$ is $\dfrac{n\beta}{\alpha}$. The posterior mean of λ is a weighted average of the prior mean and sample mean. The weights given to them ($\dfrac{\beta}{\beta+n}$ and $\dfrac{n}{\beta+n}$) are proportional to their precisions ($\dfrac{\beta^2}{\alpha}$ and $\dfrac{n\beta}{\alpha}$), shown as Equation (3). The posterior precision is the sum of the prior's precision and the sample average's precision, which is $\beta^2/\alpha + \beta n/\alpha$.

3.2 Bayesian Inference via Posterior Simulation

Bayesian inference is usually implemented by random draws from the posterior distribution of the model parameters. For example, percentiles of the posterior distribution of a specific parameter provide Bayesian credible intervals as well as skewness in its marginal posterior distribution. Furthermore, scatterplots and histograms of simulated values illustrate the posterior distributions.

In simple Bayesian models, such as the ones using a conjugate prior distribution, it is often very straightforward to make random draws from the posterior distribution directly. The Poisson example in the previous section can be conveniently implemented to make inferences about the unknown parameters, since a gamma (conjugate) prior is assumed for λ.

Taking one step further, if $y = (y_1, y_2, \ldots, y_n)'$ is an independent random sample from a Poisson distribution with a mean function $\lambda_i = \exp(x_i'\beta)$, the likelihood is proportional to

$$\pi(y|\beta, X) \propto \prod_{i=1}^{n} \exp\left[-\exp(x_i'\beta)\right]\left[\exp(x_i'\beta)\right]^{y_i} \tag{5}$$

where $X = (x_1, x_2, \ldots, x_n)'$ and the β's are coefficients for the explanatory variables in each vector x_i.

If a non-informative prior[8] is assumed for β, its posterior density can be written as proportional to this same likelihood:

$$\pi(\beta | y, X) \propto \prod_{i=1}^{n} \exp\left[-\exp(x_i'\beta)\right]\left[\exp(x_i'\beta)\right]^{y_i} \tag{6}$$

This expression is apparently not the kernel of any known parametric distribution for β. There are two approaches to estimate the unknown parameters β. One is the use of the Laplace Approximation (see, e.g., Albert and Pepple 1989; and Gelman et al. 2004).[9] The other is estimation of the posterior distribution of parameters using MCMC simulation techniques.

3.3 MCMC Simulation

Bayesian inference is almost exclusively based on MCMC simulation techniques, such as the Gibbs sampler and the M-H algorithm. (See, e.g., Metropolis et al. 1953; Hastings 1970; Tanner and Wong 1987; Gelfand and Smith 1990; Smith and Roberts 1993; Tierney 1994; Gelman et al. 2004; and Lee 2004.) The Gibbs

[8] Using informative prior distributions allows the incorporation of information available to traffic safety researchers from the literature and in light of their experience with crash analysis. However, using informative priors may lead to problems because of the subjective beliefs of the researchers. To avoid such issues, non-informative priors are commonly used. Uniform priors are generally chosen when non-informative priors are needed.

[9] The Laplace Approximation method is basically based on a Taylor series expansion of the logarithm of the posterior density around its mode. The term involving the first derivatives is zero at the mode; and the third and higher derivatives are approximately zero. Thus, the new transformed posterior density looks like a multivariate normal kernel. The statistical inference about the parameters of interest can be readily obtained using the multivariate normal density. Further details about normal approximation methods can be found in Gelman et al. (2004).

sampler and M-H algorithm set up a Markov chain[10] in the parameter space. MCMC simulation proceeds by making random draws of unknown parameters θ from approximate distributions to better approximate the target posterior distribution of the parameters $\pi\left(\theta \middle| y, X\right)$. The draws are sampled sequentially, with the distribution of the sampled draws depending on the most recent (or previous) values sampled. The Gibbs sampler and M-H algorithms provide marginal and joint distributions of all parameter estimates.

By sampling from a Markov chain whose stationary long-run distribution is the desired sample distribution, it is possible to generate observations from distributions that would otherwise be very difficult to sample. The drawbacks of this technique are: (1) it is generally unknown how long the chain must be run to reach a good approximation to the stationary distribution, and (2) the values generated are not independent.

First some background material on Markov chains will be reviewed, and then some specific methods for constructing Markov chains with a specified stationary distribution will be introduced.

The major applications driving development of MCMC methods have been to problems of Bayesian inference, but they are also useful for a variety of other problems where direct generation of independent observations from the joint distribution is difficult, such as in conditional frequentist inference problems for categorical data, where the conditional sampling distributions can have complex forms, and for Monte Carlo evaluation of integrals appearing in the E

[10] A Markov chain can be thought of as a stochastic process with a finite number of states, which moves randomly through the states without having any memory of where it has been (Lee 2004). That is, the probability of occurrence of a future state depends on only the current state.

(Expectation) step of EM algorithms (see, e.g., Chan and Ledolter 1995; and Booth and Hobert 1999).

3.3.1 Markov Chains

A Markov chain is a discrete state stochastic process. It is a sequence of random variables $\theta^{(1)}, \theta^{(2)}, \ldots$, where the distribution of current state $\theta^{(m)}$ at time (or step) m depends only on the most recent state $\theta^{(m-1)}$ at time (or step) $m-1$ (Gelman et al. 2004). That is, the Markov chain is memoryless. Only chains with a finite number of possible states will be considered formally here, although in applications it will often be convenient to think in terms of continuous distributions. The discrete state model is always technically true if the calculations are done on a computer, since there are only a finite number of values that can be represented on the computer. That is, when generating data from continuous distributions, such as the uniform or normal, the values are actually drawn from a discrete approximation to the continuous distribution.

Let $\Theta^{(m)}$ be the values that the state of the chain takes at time (or step) m. Let $S = \{\theta_1, \theta_2, \ldots, \theta_S\}$ be the state space, which is the set of possible values for $\Theta^{(m)}$. (Note: The number of possible states S could be enormously large, as long as it is finite.) The possible states θ_j could be virtually anything, but in statistical applications they usually can be thought of as points in R^p, where p is the dimension of Θ.

The chain starts from an initial state $\Theta^{(0)}$ at time (or step) 0. The distribution of the state of the chain at time (or step) m, given the state at time (or step) $m-1$, is given by a transition kernel $q_{\Theta^{(m)}}\left(\Theta^{(m)} \middle| \Theta^{(m-1)}\right)$. The Markov property states that the transition kernel depends on only the most recent draw $\Theta^{(m-1)}$. That is,

$$P\left(\Theta^{(m)} = \theta^{(m)} \middle| \Theta^{(i-1)} = \theta^{(i-1)}, i = 1,2,\ldots,m\right) = P\left(\Theta^{(m)} = \theta^{(m)} \middle| \Theta^{(m-1)} = \theta^{(m-1)}\right) \quad (7)$$

Let $p\left(\theta^{(m)} \middle| \theta^{(m-1)}\right) = P\left(\Theta^{(m)} = \theta^{(m)} \middle| \Theta^{(m-1)} = \theta^{(m-1)}\right)$. Throughout it will be assumed that the chain is homogeneous in time, meaning that the transition kernel is the same for all m. Let $p^{(\tau)}\left(\theta^{(m)} \middle| \theta^{(m-\tau)}\right) = P\left(\Theta^{(m)} = \theta^{(m)} \middle| \Theta^{(m-\tau)} = \theta^{(m-\tau)}\right)$ be the τ-step transition probabilities. A Markov chain is irreducible if there exists a value $\tau_{i,j}$ such that $p_{i,j}^{(\tau_{i,j})}\left(\theta_j \middle| \theta_i\right) > 0$ for each pair of possible states θ_i, θ_j. That is, for an irreducible chain, every state can be reached from every other state.

A state θ_i is periodic if there exists an integer $d > 1$ such that $p^{(\tau)}\left(\theta_i \middle| \theta_i\right) = 0$ whenever τ is not divisible by d. A state is aperiodic if it is not periodic. A chain is aperiodic if all states are aperiodic. For irreducible chains with a finite state space, it can be shown that either all states are aperiodic or all are periodic.

The one-step transition probabilities $p\left(\theta_j \middle| \theta_i\right)$ can be organized in an $S \times S$ matrix $P = \left(p_{ij}\right)$, where $p_{ij} = p\left(\theta_j \middle| \theta_i\right)$. Also, in general the τ-step transition probabilities can be put in a matrix $P^{(\tau)} = \left(p^{(\tau)}\left(\theta_j \middle| \theta_i\right)\right)$. Since the two-step transition probabilities satisfy $p^{(2)}\left(\theta_j \middle| \theta_i\right) = \sum_{l=1}^{S} p\left(\theta_l \middle| \theta_i\right) p\left(\theta_j \middle| \theta_l\right)$, it follows that $P^{(2)} = P^2 = P \times P$, the ordinary matrix product of P with itself. Continuing in the same fashion, it can be seen that $P^{(\tau)} = P^\tau$.

Let $\pi^{(0)}(\theta_j) = P(\Theta^{(0)} = \theta_j)$ be the initial distribution of the chain. (Note: If the chain always starts in a particular state θ_i, then this will be the degenerate distribution with $\pi^{(0)}(\theta_i) = 1$.) Also, let $\pi^{(m)}(\theta_j) = P(\Theta^{(m)} = \theta_j)$ be the marginal distribution of $\Theta^{(m)}$, and let $\pi^{(m)} = \left(\pi^{(m)}(\theta_1), \pi^{(m)}(\theta_2), \ldots, \pi^{(m)}(\theta_S)\right)'$. Since $\pi^{(1)}(\theta_j) = \sum_{l=1}^{S} \pi^{(0)}(\theta_l)\, p(\theta_j | \theta_l)$, it follows that $\pi^{(1)} = P\pi^{(0)}$, and continuing in the same fashion, $\pi^{(m)} = P^m \pi^{(0)}$.

For irreducible, aperiodic chains, there is a unique probability distribution with mass probabilities $\pi_j = \pi(\theta_j)$ (for state θ_j) satisfying $\pi = P\pi$, where $\pi = (\pi_1, \pi_2, \ldots, \pi_S)'$. This distribution is known as the stationary distribution. If the initial distribution $\pi^{(0)}$ is the stationary distribution π, then $\pi^{(1)} = P\pi^{(0)} = P\pi = \pi$, and continuing in the same fashion, $\pi^{(m)} = \pi$ for all m. Thus if the chain starts from its stationary distribution, the marginal distribution of the state at time m is again given by the stationary distribution.

Another important result is that for an irreducible, aperiodic chain with stationary distribution π, $\lim_{m \to \infty} \pi^{(m)} = \pi$, regardless of the initial distribution $\pi^{(0)}$. That is, the marginal distribution of $\Theta^{(m)}$ converges to the stationary distribution as $m \to \infty$. Thus if an irreducible, aperiodic Markov chain is started from some arbitrary state, then for sufficiently large m, the current state $\Theta^{(m)}$ is essentially generated from the stationary distribution on a state space S. Also, once the distribution $\pi^{(m)}$ converges to the stationary distribution the marginal distribution of the state at all future times is again given by the stationary distribution, so these values are an identically distributed sample from this distribution. (However, they generally are not independent.) Thus, a way to generate values from a distribution $g(\theta)$ on a state space S is to construct a Markov chain with $g(\theta)$ as its stationary distribution, and to run the chain from an arbitrary starting value until the distribution $\pi^{(m)}$ converges to $g(\theta)$. Two important problems are: (1) how to construct an appropriate Markov chain, and (2) how long the chain needs to be run to reach the stationary distribution.

For an irreducible, aperiodic chain, if

$$\pi_i p\left(\theta_j \middle| \theta_i\right) = \pi_j p\left(\theta_i \middle| \theta_j\right) \tag{8}$$

for all i and j, then π is the stationary distribution. This follows from Equation (8), since $\sum_{i=1}^{S} p\left(\theta_i \middle| \theta_j\right) = 1$, so by definition, π must be the stationary distribution. Equation (8) is called the reversibility condition, since it states that for the stationary distribution, the probability of being in state θ_i and moving to state θ_j on the next step is the same as the probability of being in state θ_j and moving to state θ_i. Equation (8) is usually easy to check, and will be very useful in helping to construct chains with arbitrary stationary distributions. However, not all chains have stationary distributions that satisfy this condition.

Several methods for constructing Markov chains are described next.

3.3.2 The Gibbs Sampler

The Gibbs sampler is logically simpler than the M-H algorithm, but requires knowledge of the conditional distributions of the unknown parameters. It generates random draws from a joint density $\pi(\theta) = \pi(\theta_1, \theta_2, \ldots, \theta_K)$, where θ is the parameter vector. Let $\pi(\theta_i | \theta_{-i})$ denote the full conditional posterior density of θ_i given values of other parameters $\theta_{-i}\left(= \theta_j,\, j \neq i,\, j \in \{1, 2, \ldots, K\}\right)$, $i \in \{1, 2, \ldots, K\}$. Taking a starting point $\theta^{(0)} = \left(\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_K^{(0)}\right)$, successive random draws are made from each of the conditional distributions $\pi(\theta_i | \theta_{-i})$ $i = 1, 2, \ldots, K$, using the following subroutine:

Draw a value $\theta_1^{(m+1)}$ from $\pi\left(\theta_1 \middle| \theta_{-1}^{(m)}\right)$;
Draw a value $\theta_2^{(m+1)}$ from $\pi\left(\theta_2 \middle| \theta_1^{(m+1)}, \theta_3^{(m)}, \ldots, \theta_K^{(m)}\right)$;
$\vdots$
Draw a value $\theta_K^{(m+1)}$ from $\pi\left(\theta_K \middle| \theta_{-K}^{(m+1)}\right)$.

where $m = 1, 2, \ldots, M$. Iterating the subroutine M times produces M draws from the joint density $\pi(\theta)$. Thus, the problem of sampling a multivariate distribution is reduced to the much easier problem of sampling from a series of univariate distributions. Under mild regularity conditions (Roberts and Smith, 1994), the sample $\left\{\theta^{(m)}; m = 1, 2, \ldots, M\right\}$ converges in distribution to $\pi(\theta)$.

3.3.3 The Metropolis-Hastings Algorithm

Practical application of MCMC sampling goes back at least to Metropolis et al (1953). Hastings (1970) extended the basic proposal from that paper and offered some of the first applications in the statistical literature. These methods did not become popular until the wide-spread availability of high-speed computers in the last decade. Beginning with Tanner and Wong (1987) and Gelfand and Smith (1990), there is now a sizable literature devoted to MCMC.

Hastings' extension of the Metropolis et al. algorithm is referred to as the M-H algorithm here. The M-H algorithm gives a general method for constructing a Markov chain with stationary distribution using an arbitrary mass function (or approximating density) $g(\theta)$.

In order to use the Gibbs sampler, one needs to have the conditional distributions for each of parameters, which are not always straightforward to obtain. For example, given a non-informative prior for β, its posterior expressed in Equation (6) cannot be written as any known parametric distribution (conditional or otherwise). Thus, the Gibbs sampler, thus, cannot be employed to make random draws from the posterior distributions. The M-H algorithm allows one to make random draws from such non-standard distributions.

Let $q\left(\theta^{(\tau_1)},\theta^{(\tau_2)}\right)$ be any Markov chain transition kernel whose state space is the same as the sample space of $\pi(\theta)$. Some specific proposals for $q\left(\theta^{(\tau_1)},\theta^{(\tau_2)}\right)$ are discussed below. To be useful, $q\left(\theta^{(\tau_1)},\bullet\right)$ should be easy to sample from, but generally q does not have $\pi(\theta)$ as its stationary distribution. Let us define an acceptance rate function as follows:

$$\alpha\left(\theta^{(m-1)},\theta^{(m)}\right)=$$
$$\begin{cases}\min\left\{\dfrac{\pi\left(\theta^{(m)}\right)q\left(\theta^{(m)},\theta^{(m-1)}\right)}{\pi\left(\theta^{(m-1)}\right)q\left(\theta^{(m-1)},\theta^{(m)}\right)},1\right\} & \pi\left(\theta^{(m-1)}\right)q\left(\theta^{(m-1)},\theta^{(m)}\right)>0 \\ 1 & \pi\left(\theta^{(m-1)}\right)q\left(\theta^{(m-1)},\theta^{(m)}\right)=0\end{cases} \tag{9}$$

Given the state of the chain at $m-1$, $\Theta^{(m-1)}$, the M-H algorithm samples a trial value $\Theta_q^{(m)}$ from $q\left(\Theta^{(m-1)},\bullet\right)$, sets $\Theta^{(m)}=\Theta_q^{(m)}$ with probability $\alpha\left(\Theta^{(m-1)},\Theta_q^{(m)}\right)$, and sets $\Theta^{(m)}=\Theta^{(m-1)}$ with probability $1-\alpha\left(\Theta^{(m-1)},\Theta_q^{(m)}\right)$. In practice this is accomplished by drawing a standard uniform random value $U^{(m)}\sim U\left(0,1\right)$ and setting $\Theta^{(m)}=\Theta_q^{(m)}I\left\{U^{(m)}\le\alpha\left(\Theta^{(m-1)},\Theta_q^{(m)}\right)\right\}+\Theta^{(m-1)}I\left\{U^{(m)}>\alpha\left(\Theta^{(m-1)},\Theta_q^{(m)}\right)\right\}$.

The transition kernel of the resulting chain is given by

$$p\left(\theta^{(m-1)},\theta^{(m)}\right)=\begin{cases}q\left(\theta^{(m-1)},\theta^{(m)}\right)\alpha\left(\theta^{(m-1)},\theta^{(m)}\right) & \theta^{(m-1)}\ne\theta^{(m)} \\ 1-\sum_{u\ne\theta^{(m-1)}}q\left(\theta^{(m-1)},u\right)\alpha\left(\theta^{(m-1)},u\right) & \theta^{(m-1)}=\theta^{(m)}\end{cases} \tag{10}$$

Approximately, if $\alpha\left(\Theta^{(m-1)},\Theta_q^{(m)}\right)<1$ then $\Theta^{(m-1)}$ is underrepresented relative to $\Theta_q^{(m)}$ in the chain generated by q, and occasionally rejecting $\Theta_q^{(m)}$ and keeping $\Theta^{(m-1)}$ adjusts for this underrepresentation. More formally, it can be shown that $\pi\left(\theta\right)$ satisfies the reversibility Equation (8) for the transition kernel (10) for all $\theta^{(m-1)}$ and $\theta^{(m)}$ in the sample space of $\pi\left(\theta\right)$, guaranteeing that $\pi\left(\theta\right)$ is the stationary distribution. For example, suppose $\theta^{(m-1)}\ne\theta^{(m)}$ are such that

$\dfrac{\pi\left(\theta^{(m-1)}\right)q\left(\theta^{(m-1)},\theta^{(m)}\right)}{\pi\left(\theta^{(m)}\right)q\left(\theta^{(m)},\theta^{(m-1)}\right)} > 1$. Then $\alpha\left(\theta^{(m-1)},\theta^{(m)}\right) = \dfrac{\pi\left(\theta^{(m)}\right)q\left(\theta^{(m)},\theta^{(m-1)}\right)}{\pi\left(\theta^{(m-1)}\right)q\left(\theta^{(m-1)},\theta^{(m)}\right)}$ and $\alpha\left(\theta^{(m)},\theta^{(m-1)}\right) = 1$, so

$$\begin{aligned}
\pi\left(\theta^{(m)}\right)q\left(\theta^{(m)},\theta^{(m-1)}\right) &= \pi\left(\theta^{(m-1)}\right)q\left(\theta^{(m-1)},\theta^{(m)}\right)\alpha\left(\theta^{(m-1)},\theta^{(m)}\right) \\
&= \pi\left(\theta^{(m-1)}\right)q\left(\theta^{(m-1)},\theta^{(m)}\right)\frac{\pi\left(\theta^{(m)}\right)q\left(\theta^{(m)},\theta^{(m-1)}\right)}{\pi\left(\theta^{(m-1)}\right)q\left(\theta^{(m-1)},\theta^{(m)}\right)} \\
&= \pi\left(\theta^{(m)}\right)q\left(\theta^{(m)},\theta^{(m-1)}\right) \qquad\qquad (11) \\
&= \pi\left(\theta^{(m)}\right)q\left(\theta^{(m)},\theta^{(m-1)}\right)\alpha\left(\theta^{(m)},\theta^{(m-1)}\right) \\
&= \pi\left(\theta^{(m)}\right)q\left(\theta^{(m)},\theta^{(m-1)}\right)
\end{aligned}$$

Note that since $\pi(\theta)$ terms appear in both the numerator and dominator of Equation (9), $\pi(\theta)$ only needs to be known to a normalizing constant.

The success of this algorithm depends on how $q\left(\theta^{(m-1)},\theta^{(m)}\right)$ is close to $\pi\left(\theta^{(m)}\right)$. If $\pi(\bullet)$ is small while $q\left(\theta^{(m-1)},\bullet\right)$ is large, then most trial points sampled will be rejected and the chain will remain in the same state for long periods of time. Thus, choosing an appropriate q is in general not a trivial problem.

Three specific methods for constructing transitional kernels $q\left(\theta^{(m-1)},\theta^{(m)}\right)$ will now be described. In the following, let $\hat{\theta}^{(m-1)}$ be the mode of $\pi\left(\theta^{(m-1)}\right)$, and let

$\hat{H} = -\left\{\partial^2 \ln\left[\pi\left(\hat{\theta}^{(m-1)}\right)\right] \Big/ \partial\theta^{(m-1)}\partial\theta^{(m-1)\prime}\right\}^{-1}$ [11]. If these quantities cannot be easily approximated, other approximations to the mean and variance of $\pi\left(\theta^{(m-1)}\right)$ may be considered.

3.3.3.1 Random Walk Chain

The first of three methods for constructing transitional kernels is called the random walk chain. Let $h(\theta)$ be a density defined on the same space as $\pi(\theta)$, and set $q\left(\theta^{(m-1)}, \theta^{(m)}\right) = h\left(\theta^{(m)} - \theta^{(m-1)}\right)$. One choice for $h(\theta)$ is the normal density with zero mean and a variance covariance matrix $\hat{H}$. A multivariate t distribution with a variance covariance matrix $\hat{H}$ may also be a better choice since its degrees of freedom can be used as a tuning factor (Lee 2004).

If $h(\theta)$ is symmetric about zero, then

$$\alpha\left(\theta^{(m-1)}, \theta^{(m)}\right) =$$

$$\begin{cases} \min\left\{ \dfrac{\pi\left(\theta^{(m)}\right) h\left(\theta^{(m)} - \theta^{(m-1)}\right)}{\pi\left(\theta^{(m-1)}\right) h\left(\theta^{(m-1)} - \theta^{(m)}\right)}, 1 \right\} & \pi\left(\theta^{(m-1)}\right) h\left(\theta^{(m-1)} - \theta^{(m)}\right) > 0 \\ 1 & \pi\left(\theta^{(m-1)}\right) h\left(\theta^{(m-1)} - \theta^{(m)}\right) = 0 \end{cases} \qquad (12)$$

$$= \begin{cases} \min\left\{ \dfrac{\pi\left(\theta^{(m)}\right)}{\pi\left(\theta^{(m-1)}\right)}, 1 \right\} & \pi\left(\theta^{(m-1)}\right) > 0 \\ 1 & \pi\left(\theta^{(m-1)}\right) = 0 \end{cases}$$

---

[11] The mode and Hessian of log-likelihood can be readily obtained using the Newton-Raphson method shown in Appendix A.

Equation (12) is the algorithm originally proposed by Metropolis et al (1953).

3.3.3.2 Independence Chain

A second algorithm is the independence chain, since trial values are independent of all prior draws. Here, $q\left(\theta^{(m-1)}, \theta^{(m)}\right) = h\left(\theta^{(m)}\right)$. Again, common choices for $h(\theta)$ are a normal or t distribution, but with a mean value of $\hat{\theta}^{(m-1)}$ and a variance covariance matrix $\hat{H}$ (Gelman et al. 2004; and Lee 2004). In this case $\alpha\left(\theta^{(m-1)}, \theta^{(m)}\right) = \min\left\{\frac{\pi\left(\theta^{(m)}\right)h\left(\theta^{(m-1)}\right)}{\pi\left(\theta^{(m-1)}\right)h\left(\theta^{(m)}\right)}, 1\right\} = \min\left\{\frac{\pi\left(\theta^{(m)}\right)/h\left(\theta^{(m)}\right)}{\pi\left(\theta^{(m-1)}\right)/h\left(\theta^{(m-1)}\right)}, 1\right\}$, so $\alpha\left(\Theta^{(m-1)}, \Theta_q^{(m)}\right)$ is the ratio of the importance sampling weights at the current and trial point values.

3.3.3.3 Rejection Sampling Chain

For rejection sampling, it is necessary to find a function $h(\theta)$ which everywhere dominates the density of $\pi(\theta)$. It is often difficult to prove a function dominates everywhere. For example, let $h(\theta)$ be a t density function with relatively few degrees of freedom, a mean value of $\hat{\theta}^{(m-1)}$ and a variance covariance matrix $c\hat{H}$ for some $c > 1$, rescaled so that $h\left(\theta^{(m-1)}\right) > \pi\left(\theta^{(m-1)}\right)$. Such a function often dominates $\pi(\theta)$ everywhere (or nearly everywhere), but this dominance generally is difficult to prove. Tierney (1994) offered a method for using the standard M-H algorithms to correct for possible non-dominance in the proposed dominating function $h(\theta)$.

Suppose then that there is a rejection-sampling algorithm that samples from a density proportional to a function $h(\theta)$, which may not actually dominate $\pi(\theta)$ everywhere. In Tierney's method, at each step in the M-H algorithm, the rejection-sampling algorithm is run, and the M-H trial value $\Theta_q^{(m)}$ is the first value not rejected in the rejection sampling. If $h(\theta)$ does not actually dominate, then the density/mass function for $\Theta_q^{(m)}$ is $f\left(\theta^{(m-1)}\right) \propto \min\left\{\pi\left(\theta^{(m-1)}\right), h\left(\theta^{(m-1)}\right)\right\}$ and the M-H transition kernel $q\left(\theta^{(m-1)}, \theta^{(m)}\right) = f\left(\theta^{(m)}\right)$. Then, defining

$$\alpha\left(\theta^{(m-1)}, \theta^{(m)}\right) = \begin{cases} \min\left\{\dfrac{\pi\left(\theta^{(m)}\right) f\left(\theta^{(m-1)}\right)}{\pi\left(\theta^{(m-1)}\right) f\left(\theta^{(m)}\right)}, 1\right\} & \pi\left(\theta^{(m-1)}\right) f\left(\theta^{(m)}\right) > 0 \\ 1 & \pi\left(\theta^{(m-1)}\right) f\left(\theta^{(m)}\right) = 0 \end{cases}$$

gives an M-H chain that corrects the sample for possible non-dominance of $h(\theta)$. If $h(\theta)$ does dominate $\pi(\theta)$, then $\alpha\left(\theta^{(m-1)}, \theta^{(m)}\right) \equiv 1$, and this algorithm is identical to rejection sampling. If $h(\theta)$ does not dominate, then points where $h\left(\theta^{(m-1)}\right) < \pi\left(\theta^{(m-1)}\right)$ will have $\alpha\left(\theta^{(m-1)}, \theta^{(m)}\right) < 1$ for some values of $\theta^{(m)}$ (assuming $h(\theta)$ does dominate at some points), so when these non-dominant points do occur, the M-H algorithm will sometimes reject the new trial point, increasing the frequency of the non-dominant points. It is straightforward to verify that $\pi(\theta)$ satisfies the reversibility Equation (8) for the transition kernel of this chain, guaranteeing that $\pi(\theta)$ is the stationary distribution.

Given the similarities among importance sampling, independence chain sampling, and rejection chain sampling, it is an interesting question as to which can be implemented more efficiently in particular applications.

This research adopts the independence chain since it is difficult to find the function $h(\theta)$ which dominates the posterior density of $\pi(\theta)$.

3.3.4 Gibbs Sampler Using Auxiliary Variables

As discussed earlier, practitioners usually turn to the M-H algorithm when facing complicated posteriors, rather than known parametric distributions. However, the M-H algorithm usually is difficult to set up and requires "tuning" to achieve satisfactory performance (Chi and Greenberg 1995; and Damien et al. 1999). A Gibbs sampler using auxiliary variables (also known as the slice sampler) in an MCMC simulation allows one to construct Markov chains that mix faster and are easier to simulate than the M-H algorithm. (See, e.g., Swendsen and Wang 1987; Edwards and Sokal 1988; Besag and Green 1993; Mira and Tierney 1997; Higdon 1998; and Damien et al. 1999.)

The introduction of auxiliary variable allows one to explore a complicated lower dimensional problem, which may become more tractable once it is embedded in a higher dimensional framework. The lower dimensional problem solution is obtained by discarding the auxiliary variables from the higher dimensional problem solutions.

Let $\pi(\theta)$ denote the posterior density specified up to a normalizing constant.

The key idea of the auxiliary variable methods is to introduce a new variable, V, and to construct the joint distribution of Θ and V by keeping the marginal distribution of Θ unchanged and defining the conditional distribution of V given

Θ: $\pi(\theta, v) = \pi(\theta)\pi(v|\theta)$. It is necessary to specify a transition kernel $K_{\theta,v}((\theta, v),(\theta', v'))$ which is irreducible, aperiodic, and has $\pi(\theta, v)$ as its stationary distribution. Thus, the marginal distribution of Θ converges to $\pi(\theta)$. Typically, the joint transitional kernel is obtained by specifying the two kernels $K_\theta((\theta, v),(\theta', v))$ and $K_v((\theta, v),(\theta, v'))$. This allows one to use the Gibbs sampler to update Θ and V conditionally on each other. The simplest choice for the conditional transition functions is (Higdon 1998):

$$K_\theta((\theta, v),(\theta', v)) = \pi(\theta'|v) \text{ and } K_v((\theta, v),(\theta, v')) = \pi(v'|\theta) \tag{13}$$

This framework allows one to choose: (1) the conditional distribution of the auxiliary variable $\pi(v|\theta)$, (2) the transitional kernels that define the Markov chain, and (3) the specification of the updating scheme between the transitional kernels.

Following Damien et al. (1999), specification of the conditional distribution, $\pi(v|\theta)$, is given in the following. Suppose the posterior density, $\pi(\theta)$, can be decomposed into two parts, $\pi(\theta) \propto q(\theta)l(\theta)$, where $q(\theta)$ is a density (e.g., the prior) and $l(\theta)$ might be the likelihood. Given the decomposition, the conditional distribution, $\pi(v|\theta)$, can be taken as uniformly distributed on the interval $(0, l(\theta))$. It leads to a joint distribution of Θ and V with density proportional to $\pi(\theta, v) \propto q(\theta) I(v < l(\theta))$, where I is an indicator function. Given the transition kernels in Equation (13), Θ and V are iteratively sampled from $q(\theta)$

(though restricted to the set $A_v = \{\theta : l(\theta) > v\}$, i.e., a truncated $q(\theta)$), and from a uniform distribution on the interval $(0, l(\theta))$, respectively. This sampling scheme is called the Gibbs sampler using auxiliary variables (or the auxiliary variable method).

The auxiliary variable method has proven very efficient for non-conjugate and hierarchical models (Damien et al. 1999). As pointed out in Besag and Green (1993), this method is also very appealing for multidimensional problems, when $q(\theta)$ has a simpler structure than $\pi(\theta)$. As mentioned in Damien et al. (1999), several decompositions $\pi(\theta) \propto q(\theta) l(\theta)$ may be available. It is difficult to determine an optimal decomposition (Mira and Tierney 1997). However, a significant improvement can be achieved if sampling the truncated $q(\theta)$ is more convenient than sampling $\pi(\theta)$ directly. Damien et al. (1999) provide several examples for decomposition of several Bayesian non-conjugate models.

3.3.5 Sampling Strategy

If the above algorithms were the full extent of what could be achieved with Markov chain sampling, then there might be little to be gained over other methods such as importance sampling. The real power of Markov chain sampling is that, when using a chain to generate observations from a multivariate distribution, it is not necessary to update all components simultaneously, so that a complex problem can be broken down into a series of simpler problems. This leads to Gibbs sampling techniques.

Suppose $\Theta \sim \pi(\theta)$ can be divided into U and V, $\Theta = (U,V)$, where U and V can also be vectors. Let $\pi_{U|V}(u|v)$ and $\pi_{V|U}(v|u)$ be the conditional distributions of $U|V$ and $V|U$, respectively. Suppose $q_{U|V}(\cdot,\cdot|v)$ is a Markov chain transition kernel with stationary distribution $\pi_{U|V}(u|v)$, and $q_{V|U}(\cdot,\cdot|u)$ is a Markov chain transition kernel with stationary distribution $\pi_{V|U}(v|u)$.

Given the current state $\Theta^{(m-1)} = \left(U^{(m-1)},V^{(m-1)}\right)$, consider the two-step update

1. generate $U^{(m)}$ from $q_{U|V}\left(U^{(m-1)},\cdot\middle|V^{(m-1)}\right)$;
2. generate $V^{(m)}$ from $q_{V|U}\left(V^{(m-1)},\cdot\middle|U^{(m-1)}\right)$.

These two steps can be thought of as generating a single update, $\Theta^{(m)}$. This update has transition kernel

$$q\left(\theta^{(m-1)},\theta^{(m)}\right) = q\left((u,v),(w,z)\right) = q_{U|V}(u,w|v)\,q_{V|U}(v,z|w).$$

This transition kernel generally does not satisfy the reversibility Equation (8), but it does have $\pi(\theta)$ as its stationary distribution. To see this, note that

$$\begin{aligned}
\sum_u\sum_v \pi(u,v)\,q\left((u,v),(w,z)\right) &= \sum_u\sum_v \pi_{U|V}(u|v)\,\pi_V(v)\,q_{U|V}(u,w|v)\,q_{V|U}(v,z|w) \\
&= \sum_v \pi_{U|V}(w|v)\,\pi_V(v)\,q_{V|U}(v,z|w) \\
&= \sum_v \pi_{V|U}(v|w)\,\pi_U(w)\,q_{V|U}(v,z|w) \\
&= \pi_{V|U}(z|w)\,\pi_U(w) = \pi(w,z)
\end{aligned} \tag{14}$$

where the second line follows because $\pi_{U|V}(w|v)$ is the stationary distribution of $q_{U|V}(u,w|v)$, and the last line because $\pi_{V|U}(z|w)$ is the stationary distribution of $q_{V|U}(v,z|w)$. Thus, to find a Markov chain with stationary distribution $\pi(\theta)$, it is only necessary to find transition kernels for the conditional distributions of blocks of components. These can be simpler to construct and to sample from.

This approach can be extended to any number of blocks (i.e., subsets of parameters). Any M-H-type update can be used within each block, and different types of updates can be used in different blocks.

The advantage of the block-at-a-time algorithm is that it is often much easier to find good approximations to the conditional distributions to use in M-H and other Markov chain updating schemes, leading to simpler methods of generating new values and greater acceptance rates of generated trial values. In some applications the conditional distributions can be sampled directly as described in the following subsection. However, separately updating blocks will often induce greater autocorrelation in the resulting Markov chain, leading to slower convergence (Gelman et al. 2004). Transformations to reduce the correlation between blocks can greatly improve the performance of block-at-a-time algorithms, although there are no simple general recipes for finding appropriate transformations.

Gibbs sampling is a block-at-a-time update scheme, where the new values for each block are generated directly from the full conditional distributions. That is, in terms of the previous notation, $q_{U|V}(u,w|v) = \pi_{U|V}(w|v)$ and $q_{V|U}(v,z|w) = \pi_{V|U}(z|w)$.

It turns out that a variety of problems can be put in a framework where the conditional distributions are easy to sample. Gelfand et al. (1990) offer several examples. This is especially true of hierarchical normal random-effects models and incomplete data problems involving normal distributions (Gelfand et al. 1995).

The idea of generating data from a series of conditional distributions was discussed by Geman and Geman (1984), and independently by Tanner and Wong (1987). The origin of the term ‘Gibbs sampling’ is not completely clear, since neither of these papers used this term. However, Geman and Geman did use Gibbs distributions in their paper.

3.3.6 Implementation Issues

The goal of Markov chain sampling is to generate observations from a specified distribution $\pi(\theta)$. Having constructed a transition kernel with stationary distribution $\pi(\theta)$, the usual approach to is to start from an arbitrary point and run the chain until it is thought to have converged, and to discard the values generated prior to convergence (Gelman et al. 2004). The values generated before convergence (and discarded) are referred to as the “burn-in.”[12]

[12] Gelfand and Smith (1990) proposed just keeping a single value from the chain and starting a new chain to generate each new observation from $\pi(\theta)$. In this way they obtain an independent series of observations from $\pi(\theta)$. It is now generally recognized that this is inefficient. If the marginal distributions of the Markov chain have converged, then more information would be obtained by continuing to sample from the same chain than from starting a new one, since each time the chain is restarted from an arbitrary value there is a new burn-in period that needs to be discarded.

Empirical monitoring for convergence of a Markov chain is not a simple problem, and examples can be constructed that will fool any particular test. Gelman and Rubin (1992) advocate running several parallel chains, starting from widely dispersed values, and offer a convergence diagnostic based on an analysis of variance procedure for comparing the within-chain and between-chain variations. At a minimum, it seems sensible to run several chains from different starting values and to compare inferences from the different chains. There have been a number of other convergence diagnostics proposed. Cowles and Carlin (1996) provide a comparative review. Simple plots of the trace of generated values over the iterations ($\Theta^{(m)}$ versus m) and of cumulative sums $\left(\sum_m \left(\Theta^{(m)} - \bar{\Theta}\right)\right)$, can reveal serious problems, and apparently "good" results in selected plots do not guarantee overall convergence.

Perhaps the greatest danger is that a chain run for some finite period of time will completely miss some important region of the sample space. For example, if a density has distinct local modes, with little probability mass between these, then the chain will tend to become trapped in one of the modes and could completely miss the others. In this case, although technically the chain may be irreducible, practically speaking it is not, since the probability of a transition from the neighborhood of one mode to another is very small. Gelman and Rubin (1992) also advocated doing an extensive search for local modes and starting chains within the distinct modes. Once the modes are located, M-H sampling based on mixtures of components centered on each mode could be considered.

Additional issues arising in implementation of the M-H algorithm is detailed by Smtih and Roberts (1993), Besag et al. (1995), and Gilks et al. (1996) among many others.

3.3.7 Precision of Estimates

If $\Theta^{(1)}$, $\Theta^{(2)}$, $\ldots$, $\Theta^{(M)}$ are the values generated from a single long run of a Markov chain with stationary distribution $\pi(\theta)$, and if the chain has approximately converged by the τ^{th} iteration, then an approximately unbiased estimator for $\int b(\theta^{(m)})\pi(\theta^{(m)})d\theta^{(m)}$ is $\tilde{b} = \frac{1}{M-\tau}\sum_{m=\tau+1}^{M} b(\Theta^{(m)})$ (see e.g. Tierney, 1994, for a more precise statement of consistency and asymptotic normality of average of sequences from Markov chains). Since the $\Theta^{(m)}$ are generally not independent, determining the precision of this estimator is not a trivial problem. Let $\sigma^2 = Var\left[b\left(\Theta^{(m)}\right)\right]$ be the marginal variance and $\rho_k = cor\left[b\left(\Theta^{(m)}\right), b\left(\Theta^{(m+k)}\right)\right]$ be the lag k autocorrelation in the sequence. If the chain is assumed to have converged by the τ^{th} step, σ^2 and ρ_k should not depend on m for $m > \tau$. If the autocorrelations die out reasonably fast and can be assumed to be negligible for $k > K$ for some K, then

$$
\begin{aligned}
Var\left(\tilde{b}\right) &= \frac{1}{(M-\tau)^2}\left(\sum_{m=\tau+1}^{M}\sigma^2 + \sum_{m<j} 2\rho_{j-m}\sigma^2\right) \\
&= \frac{\sigma^2}{(M-\tau)^2}\left(M-\tau+2\sum_{j=1}^{K}(M-\tau-j)\rho_j\right)
\end{aligned}
\qquad (15)
$$

If the autocorrelations do go to zero reasonably quickly, then the usual empirical moments will be consistent (although somewhat biased) for σ^2 and ρ_k, so that this quantity is straightforward to estimate.

Since Markov chain Monte Carlo runs are often quite long, a simpler approach is to group the data into blocks and estimate the variance from each block's means.

That is, suppose $M - \tau = Jl$, and let $\tilde{b}_j = \sum_{m=\tau+1+(j-1)l}^{\tau+jl} b\left(\Theta^{(m)}\right) \big/ l$, $j = 1,2,\ldots,J$ be the means of groups of l consecutive values. Note that $\tilde{b} = \sum_{j=1}^{J} \tilde{b}_j \big/ J$. If l is large relative to the point at which the autocorrelations die out, then the correlations among the $\tilde{b}_j$ should be negligible, and the variance can be estimated as if the $\tilde{b}_j$ were independent. If the correlation is slightly larger, then it might be reasonable to assume that the correlation between $\tilde{b}_j$ and $\tilde{b}_{j+1}$ is some value ρ to be estimated, but that correlations at larger lags are negligible. In this case

$$Var\left(\tilde{b}\right) \approx Var\left(\tilde{b}_j\right)\left(1 + 2\rho\right) \big/ J \tag{16}$$

and ρ and $Var\left(\tilde{b}_j\right)$ can be estimated using empirical moments.

3.4 Summary

This chapter introduces the basics of Bayesian statistics, the most important of which is Bayes' theorem. Bayesian theorem allows one to incorporate one's prior beliefs and observed data into the decision-making process. This chapter also presents the two most widely used Markov chain simulation methods: the Gibbs sampler and the Metropolis-Hastings algorithm. Additionally, some implementation issues in applying Markov chain Monte Carlo (MCMC) simulation techniques are briefly discussed.

CHAPTER 4 MULTIVARIATE POISSON COUNT DATA MODELS

Poisson regression models and their extensions have been widely used in modeling traffic crash occurrence. In this section, the univariate Poisson regression model is briefly introduced before discussing the multivariate Poisson-lognormal regression model. A Gibbs sampler and M-H algorithms are developed to estimate the unknown parameters.

4.1 Mathematical Formulation

4.1.1 Univariate Poisson Regression Models

The Poisson regression model plays an essential role in analyzing count data in much the same way as the normal linear model plays a role in modeling continuous data. A comprehensive account of both methodological contributions and applications of existing methods for count data can be found in several references. (See, e.g., Johnson et al. 1997; Cameron and Trivedi 1998; and Winkelmann 2003.)

Let y_{is} denote the crash count for roadway segment i and severity level s, for $i = 1, 2, \ldots, n$ and $s = 1, 2, \ldots, S$, where n is the number of roadway segments and S is the number of severity levels. Let $\vec{\mathbf{y}}_i = \left(y_{i1}, y_{i2}, \ldots, y_{iS}\right)'$ denote the vector of crash counts for roadway segment i over the different levels of severity. For example, if a segment of roadway exhibits 10 PDO crashes, 4 possible injury crashes, 3 non-disabling injury crashes, 1 disabling injury crashes, and 0 fatal crashes in a given year, its $\vec{\mathbf{y}}_i = (10, 4, 3, 1, 0)'$. Let $\mathbf{y} = (\vec{\mathbf{y}}_1', \vec{\mathbf{y}}_2', \ldots, \vec{\mathbf{y}}_n')'$ denote the vector of crash counts by severity across all roadway segments.

Suppose there is no spatial correlation among roadway segments (e.g., more proximate segments are not more likely to perform similarly ceteris paribus) and that crash counts at different levels of severity for segment i are independent of one another. Based on these two assumptions, we have the following variance-covariance matrix for crash counts in the sample:

$$Var(\mathbf{y}) = \begin{bmatrix} \Omega_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \Omega_2 & \cdots & \mathbf{0} \\ & & \vdots & \\ \mathbf{0} & \mathbf{0} & \cdots & \Omega_n \end{bmatrix} \tag{17}$$

$$\text{where } \Omega_i = \begin{bmatrix} \omega_{11}^i & 0 & \cdots & 0 \\ 0 & \omega_{22}^i & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & \omega_{SS}^i \end{bmatrix} \text{ for } i = 1, 2, \ldots, n \tag{18}$$

This is equivalent to modeling crash counts at different levels of severity separately, as univariate Poisson models. We have the following specifications for common univariate Poisson regression models:

$$y_{is} \,|\, x_{is}, \beta_s \sim Poisson\left[\lambda(x_{is}; \beta_s)\right] \text{ for } i = 1, 2, \ldots, n \text{ and } s = 1, 2, \ldots, S \tag{19}$$

where $\lambda(x_{is}; \beta_s) = \exp(x_{is}' \beta_s)$

Therefore, given x_{is} and β_s, the conditional probability distribution of y_{is} can be written as

$$p\left(y_{is} \middle| x_{is}, \beta_s\right) = \frac{\exp\left[-\exp\left(x_{is}'\beta_s\right)\right]\exp\left(x_{is}'\beta_s\right)^{y_{is}}}{y_{is}!} \text{ for } y_{is} = 0,1,2,\ldots \qquad (20)$$

The conditional expectation and variance functions can be obtained as follows.

$$E\left(y_{is} \middle| x_{is}, \beta_s\right) = \exp\left(x_{is}'\beta_s\right)$$

and

$$Var\left(y_{is} \middle| x_{is}, \beta_s\right) = \exp\left(x_{is}'\beta_s\right)$$

This specification of univariate Poisson regression models can be readily estimated in both a Bayesian paradigm (via MCMC simulation) and frequentist framework (via maximum likelihood estimation (MLE)) techniques.

To account for correlations, Tsionas (2001) developed a multivariate Poisson (MVP) regression model to examine forest damage. Ma and Kockelman (2006) applied the MVP specification to predict crash counts by severity simultaneously. More details about this application of MVP regression model can be found in Appendix B. However, the MVP specification allows for a common positive correlation among crash counts by severity and a very specific data pattern where all counts are equally shifted (by δ) which is very unreasonable for crash counts (e.g., common PDO vs. rare fatal crashes). This specification is too restrictive for modeling crash counts by severity. Specification with a more general correlation structure is pursued next.

4.1.2 Multivariate Poisson-Lognormal Regression Models

In practice, omitted variables (such as driveway density and sight distances) may simultaneously affect all crash counts at different levels of severity for a particular roadway segment, including correlation. The univariate Poisson regression models cannot account for such correlations. One needs to turn to multivariate count data models to address the issue of correlated count data. There have been several multivariate count data regression models developed (see, e.g., Arbous and Kerrich 1951; King 1989; Winkelmann 2000; Kockelman 2001; Tsionas 2001; and Karlis 2003). However, these specifications only support a common positive correlation among counts.

Here, the focus is placed on the correlated counts within individual roadway segments. Crash counts across roadway segments are assumed to be independent, i.e., no spatial correlation. The variance-covariance matrix of $\mathbf{y}$ can be expressed as below:

$$Var(\mathbf{y}) = \begin{bmatrix} \Omega_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \Omega_2 & \cdots & \mathbf{0} \\ & & \vdots & \\ \mathbf{0} & \mathbf{0} & \cdots & \Omega_n \end{bmatrix} \tag{21}$$

$$\text{where } \Omega_i = \begin{bmatrix} \omega_{11}^i & \omega_{12}^i & \cdots & \omega_{1S}^i \\ \omega_{21}^i & \omega_{22}^i & \cdots & \omega_{2S}^i \\ & & \vdots & \\ \omega_{S1}^i & \omega_{S2}^i & \cdots & \omega_{SS}^i \end{bmatrix} \text{ for } i = 1, 2, \ldots, n \tag{22}$$

Let $\vec{\boldsymbol{\varepsilon}}_i = \left(\varepsilon_{i1}, \varepsilon_{i2}, \ldots \varepsilon_{iS}\right)'$ denote the severity-level-specific unobserved heterogeneity for roadway segment i, and let $\boldsymbol{\varepsilon} = \left(\vec{\boldsymbol{\varepsilon}}_1', \vec{\boldsymbol{\varepsilon}}_2', \ldots, \vec{\boldsymbol{\varepsilon}}_n'\right)'$ denote the severity-level-specific unobserved heterogeneity across roadway segments.

Assume that crash counts y_{is}, conditioned on $\vec{\boldsymbol{\varepsilon}}_i$, the severity-level-specific explanatory variables x'_{is} and their coefficients of β_s, are independent Poisson distributed for $s = 1, 2, \ldots, S$.

$$y_{is} \left| \vec{\boldsymbol{\varepsilon}}_i, \beta_s, x_{is} \sim Poisson\left(\lambda_{is}\right) \right. \tag{23}$$

where $\lambda_{is} = \exp\left(x'_{is}\beta_s + \varepsilon_{is}\right)$ for $s = 1, 2, \ldots, S$ and $i = 1, 2, \ldots, n$. The unobserved heterogeneity $\vec{\boldsymbol{\varepsilon}}_i$ are assumed to be uncorrelated with the control (i.e., explanatory) variables, for $i = 1, 2, \ldots, n$.

Aitchison and Ho (1989) developed a multivariate Poisson-lognormal (MVPLN) distribution by mixing Poisson counts with lognormally distributed (unobserved) rates. This MVPLN distribution takes advantage of the rich covariance structure of the multivariate lognormal distribution. An MVPLN regression model can be further developed by specifying the mean and variance of counts as functions of explanatory variables.

The unobserved heterogeneity $\vec{\boldsymbol{\varepsilon}}_i$ are assumed to be multivariate normally distributed with a mean vector $\mathbf{0}$ and an unrestricted variance-covariance matrix Σ. Notationally,

$$\vec{\boldsymbol{\varepsilon}}_i \left| \Sigma \sim \phi_S\left[\mathbf{0}, \Sigma\right] \right., \text{ for } i = 1, 2, \ldots, n \tag{24}$$

where $\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1S} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2S} \\ & & \vdots & \\ \sigma_{S1} & \sigma_{S2} & \cdots & \sigma_{SS} \end{bmatrix}$ and $\sigma_{ij} = \sigma_{ji}$ for $i, j \in \{1, 2, \ldots, S\}$.

Let $\vec{\mathbf{u}}_i = \exp(\vec{\boldsymbol{\varepsilon}}_i)$, where $\vec{\mathbf{u}}_i = (u_{i1}, u_{i2}, \ldots, u_{iS})'$. Based on this setting, $\vec{\mathbf{u}}_i$ follow a multivariate lognormal distribution with mean $\boldsymbol{\mu} = E(\vec{\mathbf{u}}_i) = \exp\left[diag(\Sigma)/2\right]$ and the variance-covariance matrix

$Z = Var(\vec{\mathbf{u}}_i) = \left[diag(\boldsymbol{\mu})\right]\left[\exp(\Sigma) - \mathbf{11'}\right]\left[diag(\boldsymbol{\mu})\right]$ [13]. Thus, one obtains the following expression:

$$y_{is} \left| \xi_{is}, u_{is} \right. \sim f_{Poisson}(\xi_{is} u_{is}) \tag{25}$$

where $\xi_{is} = \exp(x'_{is}\beta_s)$ for $i = 1, 2, \ldots, n$ and $s = 1, 2, \ldots, S$; $f_{Poisson}(\cdot)$ is the univariate Poisson probability mass function..

Let $\Lambda_i = diag(\vec{\boldsymbol{\lambda}}_i)$ (=S×S matrix), where $\vec{\boldsymbol{\lambda}}_i = (\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{iS})$, and $\lambda_{is} = \xi_{is} u_{is}$ for $i = 1, 2, \ldots, n$ and $s = 1, 2, \ldots, S$. Conditioning on β and Σ, the mean and covariance matrix of the marginal distribution of $\vec{\mathbf{y}}_i$ can be obtained as follows:

$$E\left(\vec{\mathbf{y}}_i \left| \beta, x_i, \Sigma \right.\right) = E_{\vec{\mathbf{u}}_i}\left(E_{\vec{\mathbf{y}}_i|\vec{\mathbf{u}}_i}\left(\vec{\mathbf{y}}_i \left| \beta, x_i, \vec{\mathbf{u}}_i, \Sigma \right.\right)\right)$$

$$= E_{\vec{\mathbf{u}}_i}\left(diag\left(\vec{\boldsymbol{\xi}}_i\right)\vec{\mathbf{u}}_i\right) = \vec{\boldsymbol{\lambda}}_i \tag{26}$$

$$VC\left(\vec{\mathbf{y}}_i \left| \beta, x_i, \Sigma \right.\right) = E_{\vec{\mathbf{u}}_i}\left(Var_{\vec{\mathbf{y}}_i|\vec{\mathbf{u}}_i}\left(\vec{\mathbf{y}}_i \left| \beta, x_i, \vec{\mathbf{u}}_i, \Sigma \right.\right)\right) + Var_{\vec{\mathbf{u}}_i}\left(E_{\vec{\mathbf{y}}_i|\vec{\mathbf{u}}_i}\left(\vec{\mathbf{y}}_i \left| \beta, x_i, \vec{\mathbf{u}}_i, \Sigma \right.\right)\right)$$

(Greene, 2003)

$$= E_{\vec{\mathbf{u}}_i}\left(diag\left(diag\left(\vec{\boldsymbol{\xi}}_i\right)\vec{\mathbf{u}}_i\right)\right) + Var_{\vec{\mathbf{u}}_i}\left(diag\left(\vec{\boldsymbol{\xi}}_i\right)\vec{\mathbf{u}}_i\right)$$

$$= \Lambda_i + \Lambda_i\left[\exp(\Sigma) - \mathbf{11'}\right]\Lambda_i \tag{27}$$

[13] $11'$ is an S×S matrix with all elements equal to 1; $diag(\mu)$ also is an S×S matrix with diagonal elements equal to the vector μ and off-diagonal elements equal to zero.

where $\beta = (\beta_1, \beta_2, \ldots, \beta_S)'$, $x_i = (x_{i1}, x_{i2}, \ldots, x_{iS})'$ and $\vec{\boldsymbol{\xi}}_i = (\xi_{i1}, \xi_{i2}, \ldots, \xi_{iS})'$. The length of β is $k = k_1 + k_2 + \cdots + k_S$, where k_s is the length of β_s. From Equation (27), the covariance between the counts can be obtained as follows:

$$
\begin{aligned}
Cov(y_{is}, y_{il}) &= 0 + \lambda_{is}\left[\exp(\sigma_{sl}) - 1\right]\lambda_{il} \\
&= \xi_{is}\exp(\sigma_{ss}/2)\left[\exp(\sigma_{sl}) - 1\right]\xi_{il}\exp(\sigma_{ll}/2), \text{ for } s \neq l \qquad (28)
\end{aligned}
$$

$$
Var(y_{is}, y_{is}) = \lambda_{is} + \lambda_{is}\left[\exp(\sigma_{ss}) - 1\right]\lambda_{is}
$$

The correlation between crash counts within segments is obtained as follows:

$$
\begin{aligned}
&Corr(y_{is}, y_{il}) = \\
&\frac{\xi_{is}\left[\exp(\sigma_{sl}) - 1\right]\xi_{il}}{\sqrt{\xi_{is}\exp(-\sigma_{ss}/2) + \xi_{is}^2\left[\exp(\sigma_{ss}) - 1\right]}\sqrt{\xi_{il}\exp(-\sigma_{ll}/2) + \xi_{il}^2\left[\exp(\sigma_{ll}) - 1\right]}} \qquad (29) \\
&= \frac{\exp(\sigma_{sl}) - 1}{\sqrt{\xi_{is}^{-1}\exp(-\sigma_{ss}/2) + \exp(\sigma_{ss}) - 1}\sqrt{\xi_{il}^{-1}\exp(-\sigma_{ll}/2) + \exp(\sigma_{ll}) - 1}}
\end{aligned}
$$

where $s \neq l$.

The correlation between crash counts within segments can be positive or negative, depending on the sign of σ_{sl}, which is the (s,l) element of Σ. That is, a positive correlation between y_{is} and y_{il} is expected for a positive σ_{sl}, and vice versa. The correlation structure of the crash counts is thus unrestricted. Moreover, this specification also implies overdispersion[14], since $\sigma_{ss} > 0$ for $s = 1, 2, \ldots, S$.

[14] Overdispersion refers to the situation in which variance is greater than mean.

Based on Equation (25), the likelihood of observation i can be represented by the following equation:

$$P\left(\vec{\mathbf{y}}_i \left| \vec{\boldsymbol{\varepsilon}}_i, \beta, x_i \right.\right) = \prod_{s=1}^{S} f_{Poisson}\left(y_{is} \left| \lambda_{is} \right.\right) \tag{30}$$

where $\lambda_{is} = \xi_{is} u_{is} = \exp\left(x'_{is}\beta_s + \varepsilon_{is}\right)$.

Unfortunately, the *marginal* distribution of the crash counts $\vec{\mathbf{y}}_i$ cannot be obtained by direct computation. Obtaining the marginal distribution requires the evaluation of an S -variate integral of the Poisson distribution with respect to the distribution of $\vec{\boldsymbol{\varepsilon}}_i$,

$$P\left(\vec{\mathbf{y}}_i \left| \vec{\boldsymbol{\lambda}}_i, \Sigma \right.\right) = \int \prod_{s=1}^{S} f_{Poisson}\left(y_{is} \left| x_{is}, \beta_s, \varepsilon_{is} \right.\right) \phi_S \left[\vec{\boldsymbol{\varepsilon}}_i \left| \mathbf{0}, \Sigma \right.\right] d\vec{\boldsymbol{\varepsilon}}_i \tag{31}$$

where ϕ_S is the S -variate normal distribution. This S -dimensional integral cannot be algebraically implemented in closed form for arbitrary Σ.

4.2 Parameter Estimation of MVPLN Model via MCMC Simulation

4.2.1 Preparations for Estimation

The multivariate Poisson-lognormal (MVPLN) regression model was developed in the previous section. In order to appreciate crash behaviors, the unknown parameters in the model need to be estimated. Conditioning on a ($S \times 1$) vector of roadway segment and severity-level-specific random effects $\vec{\boldsymbol{\varepsilon}}_i = \left(\varepsilon_{i1}, \varepsilon_{i2}, \ldots \varepsilon_{iS}\right)'$, $\vec{\mathbf{y}}_i$ follows an independent Poisson distribution:

$$\pi\left(\vec{\mathbf{y}}_i \left|\vec{\boldsymbol{\varepsilon}}_i, x_i, \beta\right.\right) = \prod_{s=1}^{S} f_{Poisson}\left(y_{is}\left|\exp\left(x_{is}'\beta_s\right)\exp\left(\varepsilon_{is}\right)\right.\right) \tag{32}$$

where $f_{Poisson}\left(y_{is}\left|\xi_{is}\exp\left(\varepsilon_{is}\right)\right.\right) = \dfrac{\left[\xi_{is}\exp\left(\varepsilon_{is}\right)\right]^{y_{is}}\exp\left[-\xi_{is}\exp\left(\varepsilon_{is}\right)\right]}{y_{is}!}$.

and

$$\pi\left(\vec{\boldsymbol{\varepsilon}}_i \left|\Sigma\right.\right) = \phi_S\left(\mathbf{0}, \Sigma\right) \tag{33}$$

where $f_{Poisson}(\cdot)$ is Poisson probability mass function with a rate $\xi_{is}\exp\left(\varepsilon_{is}\right)$, and ϕ_S is the S-variate normal density with covariance matrix Σ. Calculating the likelihood function requires the evaluation of an S-variate integral of the Poisson distribution with respect to the distribution of $\vec{\boldsymbol{\varepsilon}}_i$; that is,

$$\pi\left(\vec{\mathbf{y}}_i \left|x_i, \beta, \Sigma\right.\right) = \int \prod_{s=1}^{S} f_{Poisson}\left(y_{is}\left|x_i, \beta, \varepsilon_{is}\right.\right)\phi_S\left(\vec{\boldsymbol{\varepsilon}}_i\left|\mathbf{0}, \Sigma\right.\right) d\vec{\boldsymbol{\varepsilon}}_i \tag{34}$$

Related to this, Chib et al. (1998) showed how to estimate a posterior distribution of unknown parameters for their panel count data models[15], and Plassmann and Tideman (2001) developed a Gibbs sampler to estimate parameters in a univariate Poisson-lognormal model.

Based on Press (1982) and Gelman et al. (2004), the Wishart distribution has been commonly used as a conjugate prior for the inverse of variance-covariance parameters. According to Press (1982), the Wishart and normal distributions are

[15] Estimation of β in the panel count data models is similar to estimation of β_s in the MVPLN model.

very helpful for multivariate analysis. Suppose that the parameters (β, Σ) independently follow the prior distributions:

$$\beta \sim \phi_k\left(\beta_0, V_{\beta_0}\right),\ \Sigma^{-1} \sim f_W\left(\nu_\Sigma, V_\Sigma\right) \tag{35}$$

where $\beta_0 = \left(\beta_{01}, \beta_{02}, \ldots, \beta_{0S}\right)'$, $V_{\beta_0} = \begin{bmatrix} V_{\beta_{01}} & 0 & \cdots & 0 \\ 0 & V_{\beta_{02}} & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & V_{\beta_{0S}} \end{bmatrix}$, $f_W(\cdot,\cdot)$ is the Wishart distribution with ν_Σ degrees of freedom and scale matrix V_Σ, and $\beta_0, V_{\beta_0}, \nu_\Sigma$ and V_Σ are known hyperparameters. Since there is correlation between $V_{\beta_{0s}}$ and $V_{\beta_{0l}}$, the prior distribution for β can written as $\beta_s \sim \phi_{k_s}\left(\beta_{0s}, V_{\beta_{0s}}\right)$ for $s = 1, 2, \ldots, S$.

According to Bayes' theorem ($posterior \propto prior \times likelihood$), the posterior kernel can be written as follows:

$$\pi\left(\Sigma, \beta \middle| y, X\right) \propto \phi_k\left(\beta_0, V_{\beta_0}\right) f_W\left(\nu_\Sigma, V_\Sigma\right) \prod_{i=1}^{n} \int \prod_{s=1}^{S} f_{Poisson}\left(y_{is} \middle| x_{is}, \beta_s, \varepsilon_{is}\right) \phi_S\left(\vec{\boldsymbol{\varepsilon}}_i \middle| \mathbf{0}, \Sigma\right) d\vec{\boldsymbol{\varepsilon}}_i$$

Using data augmentation[16], the latent effects $\boldsymbol{\varepsilon}$ can be thought of as parameters to be estimated. Therefore, the joint posterior density of Σ, $\boldsymbol{\varepsilon}$, and β is written as follows:

[16] Data augmentation uses unobserved data or latent variables as unknown parameters (to be estimated) to establish iterative algorithms.

$$\pi\left(\Sigma,\boldsymbol{\varepsilon},\beta \middle| y,X\right)\propto$$
$$\phi_k\left(\beta_0,V_{\beta_0}\right) f_W\left(\nu_\Sigma,V_\Sigma\right)\prod_{i=1}^{n}\prod_{s=1}^{S} f_{Poisson}\left(y_{is}\middle|x_{is},\beta_s,\varepsilon_{is}\right)\phi_S\left(\vec{\boldsymbol{\varepsilon}}_i\middle|\mathbf{0},\Sigma\right) \tag{36}$$

Therefore, the parameters can be "blocked" as Σ, $\boldsymbol{\varepsilon}$, and β, after which the joint posterior is simulated by iteratively sampling from the following three conditional distributions: $\pi^p\left[\Sigma^{-1}\middle|\boldsymbol{\varepsilon}\right]$, $\pi^p\left[\boldsymbol{\varepsilon}\middle|y,X,\beta,\Sigma\right]$, and $\pi^p\left[\beta\middle|y,X,\boldsymbol{\varepsilon},\Sigma\right]$, where $\pi^p\left(\cdot\middle|\cdot\right)$ is the posterior density function.

The draws are sampled sequentially using the most recent values of the conditioning variables at each step. A schematic flowchart for sampling the parameters of interest is shown as Figure 2.

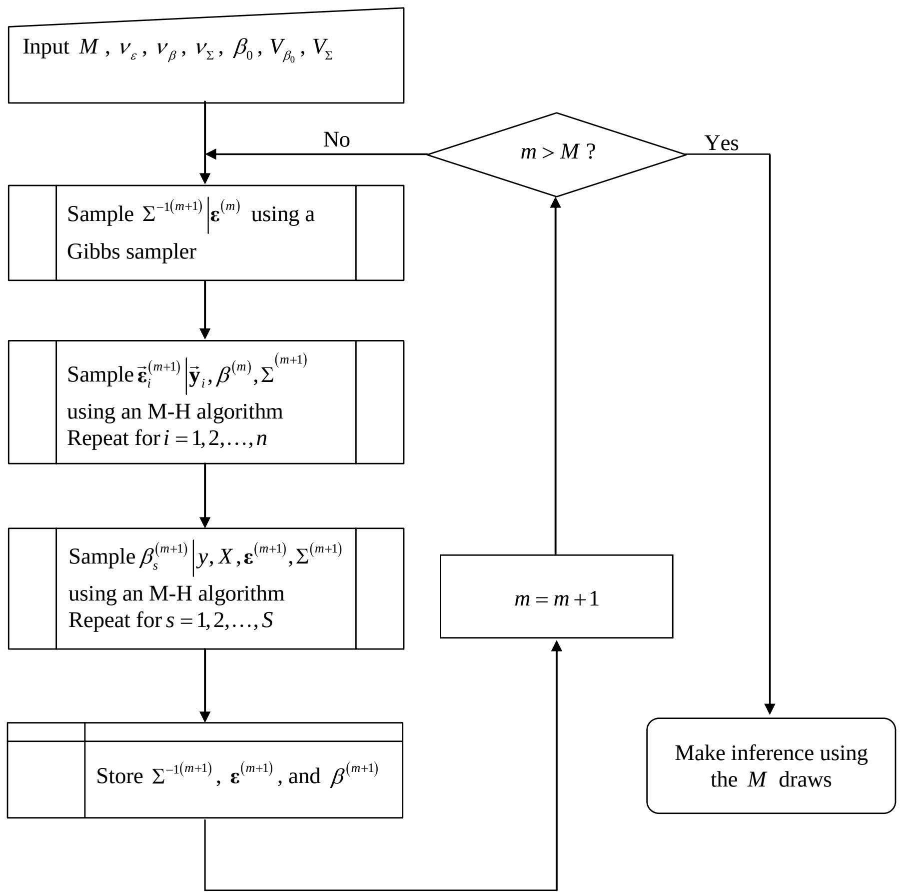


Figure 2 A Schematic Flowchart for the MCMC Simulation

4.2.2 Gibbs Sampler with Embedded M-H Algorithms

4.2.2.1 Gibbs Sampler to Draw Σ^{-1}

After manipulating the posterior equation (36), the posterior of Σ^{-1} conditional on data and other parameters can be written as

$$\pi\left(\Sigma^{-1} \middle| \boldsymbol{\varepsilon}\right) \propto f_W\left(\Sigma^{-1} \middle| \nu_\Sigma, V_\Sigma\right) \prod_{i=1}^{n} \phi_S\left(\vec{\boldsymbol{\varepsilon}}_i \middle| \mathbf{0}, \Sigma\right) \qquad (37)$$

where f_W denotes the Wishart density with ν_Σ degrees of freedom and scale matrix V_Σ.

After manipulating Equation (37), this density can be written as a Wishart kernel with degrees of freedom $n+\nu_\Sigma$ and scale matrix $\left[V_\Sigma^{-1}+\sum_{i=1}^{n}\left(\vec{\boldsymbol{\varepsilon}}_i\vec{\boldsymbol{\varepsilon}}'\right)_i\right]^{-1}$. In other words,

$$\Sigma^{-1}\left|\varepsilon \sim f_W\left(n+\nu_\Sigma,\left[V_\Sigma^{-1}+\sum_{i=1}^{n}\left(\vec{\boldsymbol{\varepsilon}}_i\vec{\boldsymbol{\varepsilon}}'\right)_i\right]^{-1}\right)\right. \tag{38}$$

This is a known parametric distribution and thus can be sampled using a Gibbs sampler.

4.2.2.2 M-H Algorithm to Draw $\vec{\boldsymbol{\varepsilon}}_i$

In order to sample $\boldsymbol{\varepsilon}$ from its posterior density $\pi\left(\boldsymbol{\varepsilon}\middle|y,\beta,\Sigma\right)=\prod_{i=1}^{n}\pi\left(\vec{\boldsymbol{\varepsilon}}_i\middle|\vec{\mathbf{y}}_i,\beta,\Sigma\right)$, consider simply the i^{th} posterior kernel density of $\vec{\boldsymbol{\varepsilon}}_i$ since the assumption of no spatial correlation across segments was made in Section 4.1.2.

$$\pi\left(\vec{\boldsymbol{\varepsilon}}_i\middle|\vec{\mathbf{y}}_i,x_i,\beta,\Sigma\right)=C_i\phi_S\left(\vec{\boldsymbol{\varepsilon}}_i\middle|\Sigma\right)\prod_{s=1}^{S}\exp\left(-\lambda_{is}\right)\lambda_{is}^{y_{is}}=C_i\pi^p\left(\vec{\boldsymbol{\varepsilon}}_i\middle|\vec{\mathbf{y}}_i,x_i,\beta,\Sigma\right) \tag{39}$$

where $\lambda_{is}=\exp\left(x_{is}'\beta_s+\varepsilon_{is}\right)$. Draws from this conditional density can be obtained by developing an M-H algorithm in the remainder of this section.

Following Chib et al. (1998), the multivariate t distribution is used as the proposal density[17]. Let $\hat{\vec{\boldsymbol{\varepsilon}}}_i = \arg\max_{\vec{\boldsymbol{\varepsilon}}_i} \left[\ln \pi^p \left(\vec{\boldsymbol{\varepsilon}}_i \middle| \vec{\mathbf{y}}_i, x_i, \beta, \Sigma \right) \right]$ and $V_{\varepsilon_i} = \left(-H_{\varepsilon_i} \right)^{-1}$ be the inverse of the Hessian of $\ln \pi^p \left(\vec{\boldsymbol{\varepsilon}}_i \middle| \vec{\mathbf{y}}_i, x_i, \beta, \Sigma \right)$ at the mode $\hat{\vec{\boldsymbol{\varepsilon}}}_i$. The mode $\hat{\vec{\boldsymbol{\varepsilon}}}_i$ and variance-covariance matrix V_{ε_i} can be obtained using the Newton-Raphson algorithm (see Appendix A) with the gradient vector

$\vec{\mathbf{g}}_{\varepsilon_i} = -\Sigma^{-1} \vec{\boldsymbol{\varepsilon}}_{\mathbf{i}} + \left[\vec{\mathbf{y}}_i - \exp\left(x_i \beta + \vec{\boldsymbol{\varepsilon}}_{\mathbf{i}} \right) \right]$ and Hessian matrix

$H_{\varepsilon_i} = -\Sigma^{-1} - diag\left[\exp\left(x_i \beta + \vec{\boldsymbol{\varepsilon}}_i \right) \right]$, where $x_i = \begin{bmatrix} x'_{i1} & 0 & \dots & 0 \\ 0 & x'_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & x'_{iS} \end{bmatrix}$ and $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_S \end{bmatrix}$.

Then, the proposal density is given by $f_T \left(\vec{\boldsymbol{\varepsilon}}_i \middle| \hat{\boldsymbol{\varepsilon}}_i, V_{\varepsilon_i}, \nu_\varepsilon \right)$, a multivariate-$t$ distribution with ν_ε degrees of freedom (where ν_ε can be used as a tuning parameter in the M-H algorithms to make sure that the acceptance rate[18] lies between 20 and 45 percent[19]). A proposal value $\vec{\boldsymbol{\varepsilon}}_i^*$ is drawn from $f_T \left(\vec{\boldsymbol{\varepsilon}}_i \middle| \hat{\boldsymbol{\varepsilon}}_i, V_{\varepsilon_i}, \nu_\varepsilon \right)$, and the chain moves to $\vec{\boldsymbol{\varepsilon}}_i^*$ from the current point $\vec{\boldsymbol{\varepsilon}}_i$ with probability

$$\alpha\left(\vec{\boldsymbol{\varepsilon}}_i, \vec{\boldsymbol{\varepsilon}}_i^* \middle| \vec{\mathbf{y}}_i, x_i, \beta, \Sigma \right) = \min \left\{ \frac{\pi^p \left(\vec{\boldsymbol{\varepsilon}}_i^* \middle| \vec{\mathbf{y}}_i, x_i, \beta, \Sigma \right) f_T \left(\vec{\boldsymbol{\varepsilon}}_i \middle| \hat{\boldsymbol{\varepsilon}}_i, V_{\varepsilon_i}, \nu_\varepsilon \right)}{\pi^p \left(\vec{\boldsymbol{\varepsilon}}_i \middle| \vec{\mathbf{y}}_i, x_i, \beta, \Sigma \right) f_T \left(\vec{\boldsymbol{\varepsilon}}_i^* \middle| \hat{\boldsymbol{\varepsilon}}_i, V_{\varepsilon_i}, \nu_\varepsilon \right)}, 1 \right\} \quad (40)$$

[17] See Section 3.3.3 for the discussions of proposal density functions.
[18] The acceptance rate is the fraction of proposed samples that is accepted. If the proposal steps are too small, the chain will move around the space slowly and thus converge slowly on the true posterior density. If the proposal steps are too large, the acceptance rate will be very low because the proposals are likely to land in regions of much lower probability density.
[19] Chib and Greenberg (1995) believe that an acceptance rate of 23 percent is desirable as the number of dimensions approaches infinity, and an acceptance rate of 45 percent is desirable for a one-dimensional random-walk chain.

If $\alpha\left(\vec{\boldsymbol{\varepsilon}}_i, \vec{\boldsymbol{\varepsilon}}_i^* \middle| \vec{\mathbf{y}}_i, x_i, \beta, \Sigma\right)$ is greater than U (where U is uniformly distributed on $[0,1]$), the proposal value $\vec{\boldsymbol{\varepsilon}}_i^*$ is accepted; otherwise, the current value $\vec{\boldsymbol{\varepsilon}}_i$ is kept as the new draw for the Markov chain.

4.2.2.3 M-H Algorithm to Draw β_s

The samples of β_s, conditional on $\boldsymbol{\varepsilon}$, y, X, Σ, and, β_{-s} (where $\beta_{-s} = [\beta_1, \beta_2, \ldots, \beta_{s-1}, \beta_{s+1}, \ldots, \beta_S]$) are drawn from the posterior distribution, which is proportional to

$$
\begin{aligned}
&\pi^p\left(\beta_s \middle| y, X, \boldsymbol{\varepsilon}, \Sigma\right) \\
&= \pi^p\left(\beta_s \middle| y_{\cdot s}, X, \varepsilon_{\cdot s}, \Sigma\right) \prod_{j=1, j \neq s}^{S} \pi^p\left(\beta_j \middle| y_{\cdot j}, X, \varepsilon_{\cdot j}, \Sigma\right) \\
&= C_{-s} \pi^p\left(\beta_s \middle| y_{\cdot s}, X, \varepsilon_{\cdot s}, \Sigma\right) \\
&\propto \phi_{k_s}\left(\beta_s \middle| \beta_{0s}, V_{\beta_{0s}}\right) \prod_{i=1}^{n} \exp\left[-\exp\left(x_{is}'\beta_s + \varepsilon_{is}\right)\right]\left[\exp\left(x_{is}'\beta_s + \varepsilon_{is}\right)\right]^{y_{is}} \\
&\propto \phi_{k_s}\left(\beta_s \middle| \beta_{0s}, V_{\beta_{0s}}\right) p\left(y_{\cdot s} \middle| \beta_s, \varepsilon_{\cdot s}\right) \qquad (41)
\end{aligned}
$$

where $C_{-s} = \prod_{j=1, j \neq s}^{S} \pi^p\left(\beta_j \middle| y_{\cdot j}, X, \varepsilon_{\cdot j}, \Sigma\right)$ (which does not involve β_s and thus serves as a constant), and

$p\left(y_{\cdot s} \middle| X, \beta_s, \varepsilon_{\cdot s}\right) = \prod_{i=1}^{n} \exp\left[-\exp\left(x_{is}'\beta + \varepsilon_{is}\right)\right]\left[\exp\left(x_{is}'\beta + \varepsilon_{is}\right)\right]^{y_{is}}$ is the probability mass function of $y_{\cdot s} = \left(y_{1s}, y_{2s}, \ldots, y_{ns}\right)$ given β_s, X and $\varepsilon_{\cdot s} = \left(\varepsilon_{1s}, \varepsilon_{2s}, \ldots, \varepsilon_{ns}\right)$.

Note that β_s's ($s \in \{1, 2, \ldots, S\}$) are assumed to be independent of one another.

A scheme similar to the one sampling $\vec{\boldsymbol{\varepsilon}}_i$ is developed here to sample β_s. The multivariate-t is also used as the proposal density. Let

$\hat{\beta}_s = \underset{\beta_s}{\arg\max}\left[\ln \pi^p\left(\beta_s \left| y_{\cdot s}, X, \varepsilon_{\cdot s}, \beta_{-s}, \Sigma\right.\right)\right]$ be the mode, and $V_{\beta_s} = \left(-H_{\beta_s}\right)^{-1}$ the inverse of the Hessian of $\ln \pi^p\left(\beta_s \left| y, X, \boldsymbol{\varepsilon}, \beta_{-s}, \Sigma\right.\right)$ at the mode $\hat{\beta}_s$. The mode $\hat{\beta}_s$ and variance-covariance matrix V_{β_s} can be obtained using the Newton-Raphson algorithm with the gradient vector $\vec{\mathbf{g}}_{\beta_s} = -V_{\beta_{0s}}^{-1}\left(\beta_s - \beta_{0s}\right) + \sum_{i=1}^{n}\left[y_{is} - \exp\left(x_{is}'\beta_s + \varepsilon_{is}\right)\right]x_{is}$ and Hessian matrix $H_{\beta_s} = -V_{\beta_{0s}}^{-1} - \sum_{i=1}^{n}\left[\exp\left(x_{is}'\beta_s + \vec{\boldsymbol{\varepsilon}}_{is}\right)\right]x_{is}x_{is}'$. Then, the proposal density is given by $f_T\left(\beta_s \left| \hat{\beta}_s, V_{\beta_s}, \nu_\beta\right.\right)$, a multivariate-$t$ distribution with ν_β degrees of freedom (where ν_β can be used as a tuning parameter in the M-H algorithms to make sure that the acceptance rate lies between 20 and 45 percent). A proposal value β_s^* is drawn from $f_T\left(\beta_s \left| \hat{\beta}_s, V_{\beta_s}, \nu_\beta\right.\right)$, and the chain moves to β_s^* from the current point β_s with probability

$$\alpha\left(\beta_s, \beta_s^* \left| y, X, \boldsymbol{\varepsilon}, \beta_{-s}, \Sigma\right.\right)\min\left\{\frac{\pi^p\left(\beta_s^* \left| y, X, \boldsymbol{\varepsilon}, \beta_{-s}, \Sigma\right.\right) f_T\left(\beta_s \left| \hat{\beta}_s, V_{\beta_s}, \nu_\beta\right.\right)}{\pi^p\left(\beta_s \left| y, X, \boldsymbol{\varepsilon}, \beta_{-s}, \Sigma\right.\right) f_T\left(\beta_s^* \left| \hat{\beta}_s, V_{\beta_s}, \nu_\beta\right.\right)}, 1\right\} \quad (42)$$

If $\alpha\left(\beta_s, \beta_s^* \left| y, X, \boldsymbol{\varepsilon}, \beta_{-s}, \Sigma\right.\right)$ is greater than U (where U is uniformly distributed on $[0,1]$), the proposal value β_s^* is accepted; otherwise, the current value β_s is kept as the new draws for the Markov chain.

4.2.3 Gibbs Sampler Using Auxiliary Variables

The M-H algorithms are commonly used to sample draws from the posterior density which cannot be written as a known parametric distribution. However, much effort is needed to make the acceptance rate lying between 20 and 45 percent. The Gibbs sampler using auxiliary variables (or slice sampler) avoids the M-H algorithms by introducing auxiliary variables to transform the posterior density into a series known conditional distributions. Thus, the parameters can be estimated by sampling draws from the conditional distributions.

4.2.3.1 Gibbs Sampler to Draw Σ^{-1}

The posterior distribution of Σ^{-1} can be written as a Wishart distribution and thus can be sampled using a Gibbs sampler. Please see Section 4.2.2.1 for details about sampling Σ^{-1}.

4.2.3.2 Gibbs Sampler Using Auxiliary Variables to Draw $\vec{\boldsymbol{\varepsilon}}_i$

Please refer to Section 3.3.4 for details about the Gibbs sampler using auxiliary variables (i.e., the Slice Sampler). Equation (39) can be written as

$$\pi\left(\vec{\boldsymbol{\varepsilon}}_i \mid \vec{\mathbf{y}}_i, x_i, \beta, \Sigma\right) \propto \phi_S\left(\vec{\boldsymbol{\varepsilon}}_i \mid \Sigma\right) \prod_{s=1}^{S} \exp\left(-\exp\left(x'_{is}\beta_s + \varepsilon_{is}\right)\right) \exp\left(x'_{is}\beta_s + \varepsilon_{is}\right)^{y_{is}} \qquad (43)$$

Let $l_{is,1}\left(\varepsilon_{is}\right) = \exp\left(-\exp\left(x'_{is}\beta_s + \varepsilon_{is}\right)\right)$ and $l_{is,2}\left(\varepsilon_{is}\right) = \exp\left(x'_{is}\beta_s + \varepsilon_{is}\right)^{y_{is}}$. We can introduce the variables U_{is} and V_{is} whose joint distribution with ε_{is} is given by

$$f\left(\vec{\boldsymbol{\varepsilon}}_i, \vec{\mathbf{u}}_i, \vec{\mathbf{v}}_i\right) \propto \phi_S\left(\vec{\boldsymbol{\varepsilon}}_i \mid \Sigma\right) \prod_{s=1}^{S} I\left(u_{is} < l_{is,1}\left(\varepsilon_{is}\right)\right) I\left(v_{is} < l_{is,2}\left(\varepsilon_{is}\right)\right) \qquad (44)$$

where $\vec{\boldsymbol{\varepsilon}}_i = \left(\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iS}\right)'$, $\vec{\mathbf{u}}_i = \left(u_{i1}, u_{i2}, \ldots, u_{iS}\right)'$, and $\vec{\mathbf{v}}_i = \left(v_{i1}, v_{i2}, \ldots, v_{iS}\right)'$. The full conditionals are given by

$$f\left(u_{is}\left|\varepsilon_{is}\right.\right)=Uniform\left(0,l_{is,1}\left(\varepsilon_{is}\right)\right),$$

$$f\left(v_{is}\left|\varepsilon_{is}\right.\right)=Uniform\left(0,l_{is,2}\left(\varepsilon_{is}\right)\right)$$

$$f\left(\vec{\boldsymbol{\varepsilon}}_i\left|\vec{\mathbf{u}}_i,\vec{\mathbf{v}}_i\right.\right)\propto\phi_S\left(\vec{\boldsymbol{\varepsilon}}_i\left|\Sigma\right.\right)\text{ with }\varepsilon_{is}\text{ restricted to }M_{is}\text{ for }s=1,2,\ldots,S$$

where $M_{is}=M_{is,1}\bigcap M_{is,2}=\left\{\varepsilon_{is}:y_{is}^{-1}\log\left(v_{is}\right)-x_{is}'\beta_s<\varepsilon_{is}<\log\left[-\log\left(u_{is}\right)\right]-x_{is}'\beta_s\right\}$,

$M_{is,1}=\left\{\varepsilon_{is}:\varepsilon_{is}<\log\left[-\log\left(u_{is}\right)\right]-x_{is}'\beta_s\right\}$, and

$M_{is,2}=\left\{\varepsilon_{is}:\varepsilon_{is}>y_{is}^{-1}\log\left(v_{is}\right)-x_{is}'\beta_s\right\}$.

The conditional distribution of $\vec{\boldsymbol{\varepsilon}}_i$, $f\left(\vec{\boldsymbol{\varepsilon}}_i\left|\vec{\mathbf{u}}_i,\vec{\mathbf{v}}_i\right.\right)$, is in fact a truncated multivariate normal kernel. The difficulty lies in obtaining random draws from the truncated multivariate normal distribution. One way is to sample from a multivariate normal distribution and determine if the sampled values meet the truncation criteria. This algorithm tends to be inefficient when drawing many samples. Geweke (1991) and Rodriguez-Yam (2003) proposed two efficient ways to sample from truncated multivariate normal distributions. This research adopts their method.

4.2.3.3 Gibbs Sampler Using Auxiliary Variables to Draw β_s

Based on Equation (41), the posterior kernel of β_s (conditional on data and other parameters) can be written as

$$\pi^p\left[\beta_s\left|y,X,\boldsymbol{\varepsilon},\Sigma\right.\right]\propto\phi_{k_s}\left(\beta_s\left|\beta_{0s},V_{\beta_{0s}}\right.\right)\prod_{i=1}^{n}\exp\left[-\exp\left(x_{is}'\beta_s+\varepsilon_{is}\right)\right]\left[\exp\left(x_{is}'\beta_s+\varepsilon_{is}\right)\right]^{y_{is}}$$

Similarly, let $l_{i,1}(\beta_s) = \exp\left(-\exp\left(x'_{is}\beta_s + \varepsilon_{is}\right)\right)$ and $l_{i,2}(\beta_s) = \exp\left(x'_{is}\beta_s + \varepsilon_{is}\right)^{y_{is}}$. One can introduce the variables U_i and V_i whose joint distribution with β_s is given by

$$f(\beta_s, \vec{\mathbf{u}}, \vec{\mathbf{v}}) \propto \phi_{k_s}\left(\beta_s \middle| \beta_{0s}, V_{\beta_{0s}}\right) \prod_{i=1}^{n} I\left(u_i < l_{i,1}(\beta_s)\right) I\left(v_i < l_{i,2}(\beta_s)\right) \qquad (45)$$

where $\vec{\mathbf{u}} = (u_1, u_2, \ldots, u_n)'$, and $\vec{\mathbf{v}} = (v_1, v_2, \ldots, v_n)'$. The full conditionals are given by

$f\left(u_i \middle| \beta_s\right) = Uniform\left(0, l_{i,1}(\beta_s)\right)$,

$f\left(v_i \middle| \beta_s\right) = Uniform\left(0, l_{i,2}(\beta_s)\right)$

$f\left(\beta_s \middle| \vec{\mathbf{u}}, \vec{\mathbf{v}}\right) \propto \phi_{k_s}\left(\beta_s \middle| \beta_{0s}, V_{\beta_{0s}}\right)$ with β_s restricted to the sample space M

where

$M = M_1 \cap M_2 =$

$\left\{\beta_s : \left[x_{is}x'_{is}\right]^{-1} x_{is}\left[y_{is}^{-1}\log(v_i) - \varepsilon_{is}\right] < \beta_s < \left[x_{is}x'_{is}\right]^{-1} x_{is}\left[\log\left[-\log(u_i)\right] - \varepsilon_{is}\right]\right\}$[20],

$M_1 = \left\{\beta_s : \beta_s < \left[x_{is}x'_{is}\right]^{-1} x_{is}\left[\log\left[-\log(u_i)\right] - \varepsilon_{is}\right]\right\}$, and

$M_2 = \left\{\beta_s : \beta_s > \left[x_{is}x'_{is}\right]^{-1} x_{is}\left[y_{is}^{-1}\log(v_i) - \varepsilon_{is}\right]\right\}$.

The conditional distribution of β_s, $f\left(\beta_s \middle| \vec{\mathbf{u}}, \vec{\mathbf{v}}\right)$, is in fact a truncated multivariate normal kernel which can be sampled using the methods proposed in Rodriguez-

---

[20] In order to make this inequality work, one has to transform all of explanatory into positive values. Given the data sets in the next chapter, the explanatory variables only take non-negative values. To guarantee the positive explanatory variables, all of zeros in observations are replaced with a small number 10^{-10}.

Yam (2003). Damien and Walker (2001) also developed a Gibbs sampler within the Gibbs sampler to sample random variates from truncated normal densities. The slice sampler developed in this research eliminates the need to “tune” proposal distribution as in general M-H algorithms. However, it requires stronger assumptions to construct full conditions. In addition, the acceptance rate for new points is always 100% in the slice sampler.

4.3 Summary

This chapter starts from univariate Poisson regression models and then moves to multivariate Poisson models. The MVP regression model is briefly discussed since it only allows for a common positive correlation between each pair of counts. The MVPLN regression specification is developed to allow for a more general correlation structure for traffic crash counts at different levels of severity on a particular segment. Additionally, two MCMC simulation procedures are established to estimate the unknown parameters in the MVPLN model.

CHAPTER 5 EMPIRICAL ANALYSES

Therefore, this research conducts an empirical analysis using Washington State rural two-lane crash data sets in the Puget Sound region. An MVPLN regression model is estimated using the data sets.

5.1 Data Description

The crash data sets used here were collected from Washington State through the Highway Safety Information System (HSIS). After filtering off unreasonable observations (such as segments with zero speed limits), a total of 103,106 Washington State homogeneous highway segments are available for analysis (with an average segment length of 0.0814 miles and a total distance of 8,400 centerline miles). In the year 2002, there were 177 fatal crashes, 1,013 disabling injury crashes, 5,630 non-disabling injury crashes, 13,060 possible injury crashes and 29,006,100 property-damage-only (PDO) crashes along these segments. These segments serve as distinct observational units, and the HSIS dataset contains information on their design features, including curve attributes, shoulder and surface width, speed limit, and average annual daily traffic (AADT). A total of 12 explanatory variables are controlled for in the model.

In order to examine traffic crashes patterns on rural two-lane roadways, this research considers crashes occurring on rural two-lane roads in the Puget Sound region. A random sample of 60%[21] of rural two-lane road segments in this region was used for model estimation. A total of 7,773 rural two-lane highway segments (with an average segment length of 0.0655 miles and a total of 510 miles) are available for analysis. This sample contains 16 fatal crashes, 50 disabling-injury

[21] A preliminary implementation of the MCMC algorithms shows that it would take more than a year to get 6,000 draws using the entire Washington State data.

crashes, 180 non-disabling-injury crashes, 175 possible-injury crashes and 532 property-damage-only (PDO). Table 5 reports summary statistics for the dependent and independent variables employed in the analysis.

Table 5 Summary Statistics of Variables for Puget Sound Roadway Segments 2002

Variable Name	Mean	Std. Dev.	Min	Max
Dependent Variables				
Number of fatal crashes	0.002058	.04533	0	1
Number of disabling injury crashes	0.006433	.07995	0	1
Number of non-disabling injury crashes	0.02316	.1587	0	3
Number of possible injury crashes	0.02251	.2045	0	11
Number of PDO crashes	0.06844	.3345	0	12
Independent Variables				
Segment length (miles)	0.0655	.08689	.00	1.92
Horizontal curve length (feet)	247.6	475.4	.00	4715
Degree of curvature (°/100feet)	2.337	5.462	.00	100.5
Vertical curve length (feet)	302.7	376.0	.00	3200
Vertical grade (%)	1.805	1.991	.00	16.13
Average shoulder width (feet)	2.087	1.298	.00	16.50
Surface width (feet)	24.00	4.461	16.0	73.0
Posted speed limit (miles/hour)	49.62	8.163	25.0	60.0
Posted speed limit squared (miles^2/hour^2)	2528	715.5	625	3600
Average annual daily traffic (AADT)	3757	2,729	254	28,624
Indicator for principal arterial: 1=yes, 0=otherwise	0.48	0.499	0	1
Indicator for minor arterial: 1=yes, 0=otherwise	0.28	0.451	0	1
Indicator for collector: 1=yes, 0=otherwise	0.24	0.430	0	1
Indicator for level terrain: 1=yes, 0=otherwise	0.36	0.482	0	1
Indicator for rolling terrain: 1=yes, 0=otherwise	0.60	0.491	0	1
Indicator for mountainous terrain: 1=yes, 0=otherwise	0.04	0.194	0	1
Vehicle miles traveled (VMT) in 2002	88,106	142,830	.00	2,679,710
The natural logarithm of VMT	10.45	2.737	-22.35	14.80
Number of observations				7,773

5.2 Model Estimation and Results

5.2.1 Model Estimation

The MVPLN regression model described in Chapter 4 was estimated using a Bayesian approach. The starting values for β came from distinct univariate Poisson models (using the method of maximum likelihood estimation (MLE)). The starting values for Σ are $I_5 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$. Tables C-1 through C-5 in Appendix C provide the MLE estimates for the five univariate Poisson models. A Gibbs sampler and two M-H algorithms were coded in R language (an open-source statistical computing environment described at http://www.r-project.org/). The priors for the estimation are defined by the hyperparameters ν_Σ =10, $V_\Sigma^{-1} = I_5$, $\beta_{0s} = (0, 0, \ldots, 0)'$, and $V_{\beta_{0s}} = 100 \times I_{14}$. The Gibbs sampler was implemented to obtain M = 8,000 draws for Σ. The two M-H algorithms were implemented to obtain M = 8,000 draws for each of the $5 \times 14 = 70$ $\beta's$ and each of the $7{,}773 \times 5 = 38{,}865$ ε's, respectively. The initial 1,000 draws were discarded as "burn-ins." As described in Section 3.3.6, an adequate burn-in period eliminates the influence of the starting values. To help ensure chain convergence, the Gibbs sampler and the two M-H algorithms were implemented using two sets of initial values[22] and both converged at the same posterior distribution of parameters. Estimation results are presented in Tables 6 through 10.

[22] Zero was used as the initial values for β in the second chain.

Table 6 PDO Crash Frequency MVPLN Model for the Puget Sound Region in 2002

Variable definition	Mean	Std. Err.	The 95% (2.5-97.5%) HDR	
Constant	-12.64	0.4562	-13.38	-11.88
Horizontal curve length (feet)	*2.09E-05*	*1.35E-05*	*-1.31E-06*	*4.27E-05*
Degree of curvature (°/100feet)	0.1241	6.31E-03	0.1136	0.1344
Vertical curve length (feet)	-2.05E-04	1.97E-05	-2.37E-04	-1.73E-04
Vertical grade (%)	0.1377	0.01441	0.1134	0.1609
Average shoulder width (feet)	-0.01125	3.54E-03	-0.01694	-5.28E-03
Surface width (feet)	-0.01520	5.25E-04	-0.01607	-0.01434
Posted speed limit (miles/hour)	0.01493	2.89E-03	0.01014	0.01972
Posted speed limit squared (miles^2/hour^2)	-1.53E-04	8.64E-05	-2.97E-04	-1.33E-05
Average annual daily traffic (AADT)	4.79E-05	2.03E-06	4.46E-05	5.13E-05
Indicator for minor arterial: 1=yes, 0=otherwise	*-0.01112*	*0.01631*	*-0.03759*	*0.01568*
Indicator for collector: 1=yes, 0=otherwise	*-0.009441*	*0.01872*	*-0.04049*	*0.02080*
Indicator for rolling terrain: 1=yes, 0=otherwise	0.03929	0.01439	0.01526	0.06240
Indicator for mountainous terrain: 1=yes, 0=otherwise	0.6120	0.04687	0.5355	0.6888
Number of observations	7,773			

Note: Smaller, lighter font is used for parameters that do **not** differ from zero in a statistically significant way, based on the 95% (2.5-97.5) high density region (HDR).

Table 7 Possible-Injury Crash Frequency MVPLN Model for the Puget Sound Region in 2002

Variable definition	Mean	Std. Err.	The 95% (2.5-97.5%) HDR	
Constant	-15.85	0.8120	-17.22	-14.53
Horizontal curve length (feet)	*2.90E-05*	*2.37E-05*	*-8.46E-06*	*6.90E-05*
Degree of curvature (°/100feet)	0.1031	7.09E-03	0.09136	0.1147
Vertical curve length (feet)	-2.97E-04	1.30E-05	-3.18E-04	-2.76E-04
Vertical grade (%)	0.1616	9.20E-03	0.1465	0.1766
Average shoulder width (feet)	-8.71E-03	9.48E-04	-0.01027	-7.17E-03
Surface width (feet)	-0.01258	7.16E-04	-0.01371	-0.01139
Posted speed limit (miles/hour)	0.03116	5.25E-03	0.02238	0.03970
Posted speed limit squared (miles^2/hour^2)	*-1.40E-05*	*1.57E-05*	*-4.02E-05*	*1.19E-05*
Average annual daily traffic (AADT)	1.08E-04	3.28E-06	1.03E-04	1.13E-04
Indicator for minor arterial: 1=yes, 0=otherwise	0.2257	0.02809	0.1799	0.2729
Indicator for collector: 1=yes, 0=otherwise	0.4971	0.03114	0.4448	0.5478
Indicator for rolling terrain: 1=yes, 0=otherwise	-0.2344	0.02530	-0.2756	-0.1934
Indicator for mountainous terrain: 1=yes, 0=otherwise	-0.3552	0.1301	-0.5677	-0.1452
Number of observations	7,773			

Note: Smaller, lighter font is used for parameters that do **not** differ from zero in a statistically significant way, based on the 95% (2.5-97.5) high density region (HDR).

Table 8 Non-disabling Injury Crash Frequency MVPLN Model for the Puget Sound Region in 2002

Variable definition	Mean	Std. Err.	The 95% (2.5-97.5%) HDR	
Constant	-15.37	0.9321	-16.89	-13.81
Horizontal curve length (feet)	-2.01E-05	2.41E-06	-2.41E-05	-1.61E-05
Degree of curvature (°/100feet)	0.1576	6.04E-03	0.1477	0.1676
Vertical curve length (feet)	-2.04E-04	1.12E-05	-2.22E-04	-1.85E-04
Vertical grade (%)	0.1850	0.01532	0.1602	0.2110
Average shoulder width (feet)	-4.69E-03	9.17E-04	-6.22E-03	-3.22E-03
Surface width (feet)	-0.01079	1.25E-03	-0.01287	-8.72E-03
Posted speed limit (miles/hour)	0.01335	1.73E-03	0.01051	0.01621
Posted speed limit squared (miles^2/hour^2)	*-2.30E-04*	*1.56E-04*	*-4.82E-04*	*3.38E-05*
Average annual daily traffic (AADT)	*2.37E-06*	*3.55E-06*	*-3.46E-06*	*8.24E-06*
Indicator for minor arterial: 1=yes, 0=otherwise	0.2489	0.02867	0.2025	0.2963
Indicator for collector: 1=yes, 0=otherwise	0.4896	0.03679	0.4292	0.5508
Indicator for rolling terrain: 1=yes, 0=otherwise	0.1341	0.02343	0.09553	0.1733
Indicator for mountainous terrain: 1=yes, 0=otherwise	*-0.1685*	*0.1100*	*-0.3428*	*0.01523*
Number of observations	7,773			

Note: Smaller, lighter font is used for parameters that do **not** differ from zero in a statistically significant way, based on the 95% (2.5-97.5) high density region (HDR).

Table 9 Disabling Injury Crash Frequency MVPLN Model for the Puget Sound Region in 2002

Variable definition	Mean	Std. Err.	The 95% (2.5-97.5%) HDR	
Constant	-16.73	2.182	-20.37	-13.12
Horizontal curve length (feet)	6.49E-05	3.97E-05	3.70E-07	1.30E-04
Degree of curvature (°/100feet)	0.02029	6.64E-03	9.62E-03	0.03097
Vertical curve length (feet)	-3.69E-04	3.63E-05	-4.28E-04	-3.10E-04
Vertical grade (%)	0.1431	0.01101	0.1255	0.1607
Average shoulder width (feet)	*6.27E-03*	*0.01656*	*-0.02102*	*0.03334*
Surface width (feet)	-9.85E-03	1.47E-03	-0.01226	-7.41E-03
Posted speed limit (miles/hour)	0.01040	1.81E-03	7.42E-03	0.01344
Posted speed limit squared (miles^2/hour^2)	*3.48E-04*	*3.22E-04*	*-1.94E-04*	*8.64E-04*
Average annual daily traffic (AADT)	5.34E-04	5.78E-05	4.38E-04	6.30E-04
Indicator for minor arterial: 1=yes, 0=otherwise	0.3470	0.04676	0.2700	0.4243
Indicator for collector: 1=yes, 0=otherwise	0.4106	0.05675	0.3171	0.5033
Indicator for rolling terrain: 1=yes, 0=otherwise	0.2814	0.04212	0.2133	0.3498
Indicator for mountainous terrain: 1=yes, 0=otherwise	*167.6*	*115.3*	*-24.93*	*355.2*
Number of observations	7,773			

Note: Smaller, lighter font is used for parameters that do **not** differ from zero in a statistically significant way, based on the 95% (2.5-97.5) high density region (HDR).

Table 10 Fatal Crash Frequency MVPLN Model for the Puget Sound Region in 2002

Variable definition	Mean	Std. Err.	The 95% (2.5-97.5%) HDR	
Constant	-24.46	6.780	-35.61	-13.63
Horizontal curve length (feet)	-3.56E-05	5.67E-06	-4.47E-05	-2.63E-05
Degree of curvature (°/100feet)	0.02080	1.23E-03	0.01868	0.02274
Vertical curve length (feet)	3.67E-05	1.07E-05	1.93E-05	5.39E-05
Vertical grade (%)	-0.05849	0.02737	-0.1032	-0.01380
Average shoulder width (feet)	*0.01766*	*0.03147*	*-0.03503*	*0.06981*
Surface width (feet)	0.05338	0.02102	0.01937	0.08909
Posted speed limit (miles/hour)	0.01463	2.27E-03	0.01073	0.01835
Posted speed limit squared (miles^2/hour^2)	*1.78E-04*	*9.08E-04*	*-1.34E-03*	*1.64E-03*
Average annual daily traffic (AADT)	*1.64E-05*	*1.30E-05*	*-4.62E-06*	*3.83E-05*
Indicator for minor arterial: 1=yes, 0=otherwise	0.1532	0.09024	3.70E-03	0.3053
Indicator for collector: 1=yes, 0=otherwise	0.4176	0.1206	0.2263	0.6169
Indicator for rolling terrain: 1=yes, 0=otherwise	-0.1714	0.07712	-0.2997	-0.04648
Indicator for mountainous terrain: 1=yes, 0=otherwise	1.801	0.2251	1.436	2.172
Number of observations	7,773			

Note: Smaller, lighter font is used for parameters that do **not** differ from zero in a statistically significant way, based on the 95% (2.5-97.5) high density region (HDR).

Figures 3 through 7 illustrate the estimates of posterior distributions for these regression coefficients. Based on the posterior density of Σ (shown in Figures 8a and 8b), positive correlations between crash counts at different levels of severity within the segment do appear to exist, in a statistically significant way. The univariate models are a special case of the MVPLN, with off-diagonal elements of Σ equal to zero. Given the MVPLN predictions' added flexibility to represent such pattern, it is expected that they will offer somewhat better predictions.

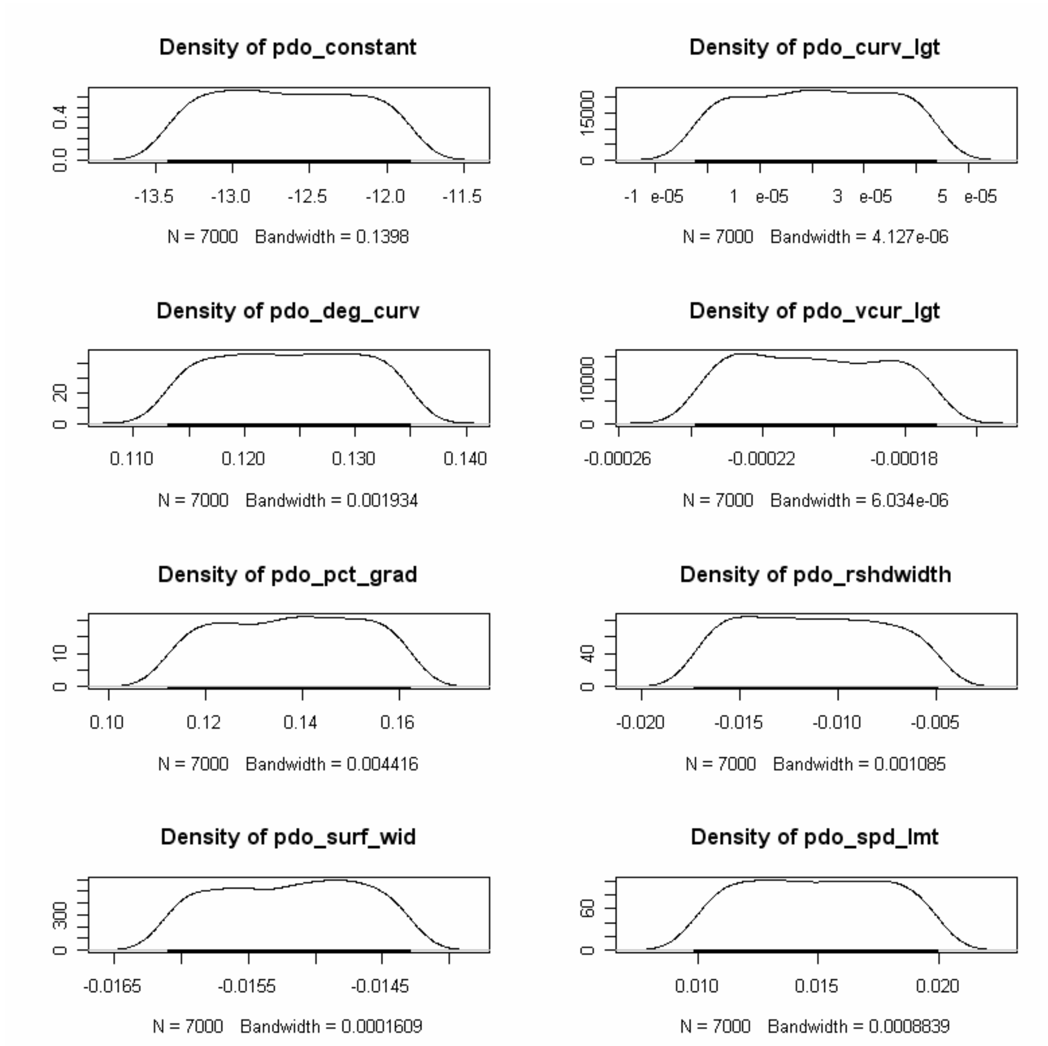


Figure 3a Posterior Density of Variables of Interest for PDO Crash Frequency

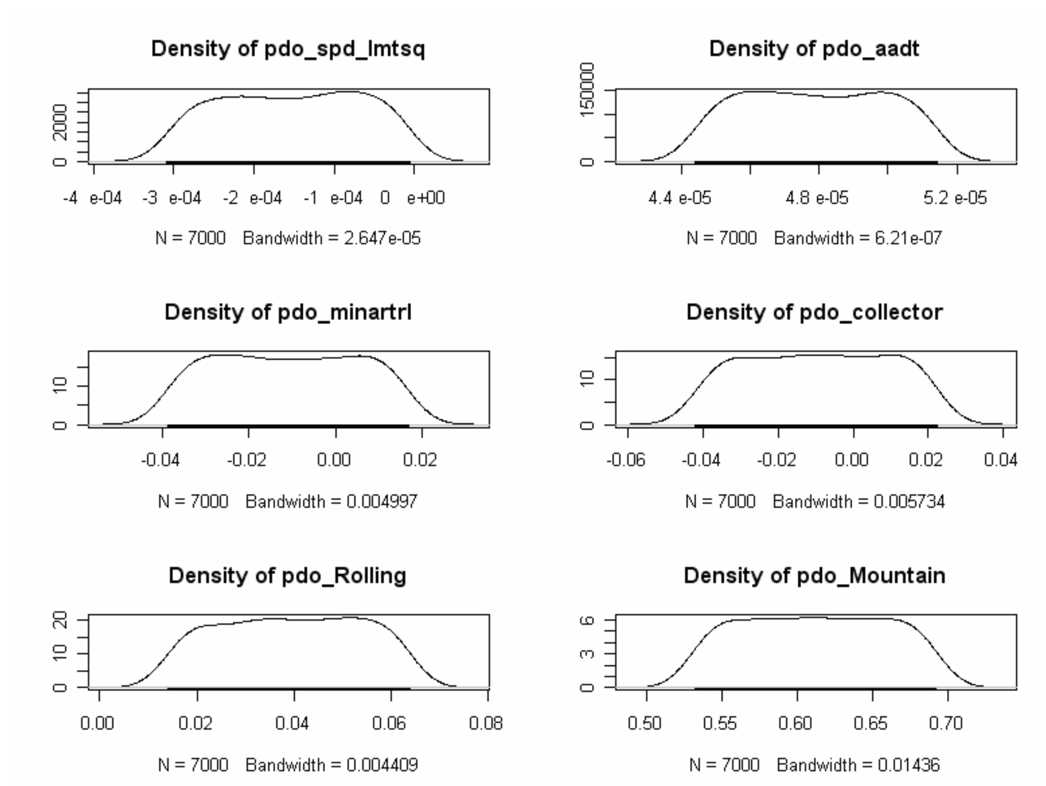


Figure 3b Posterior Density of Variables of Interest for PDO Crash Frequency

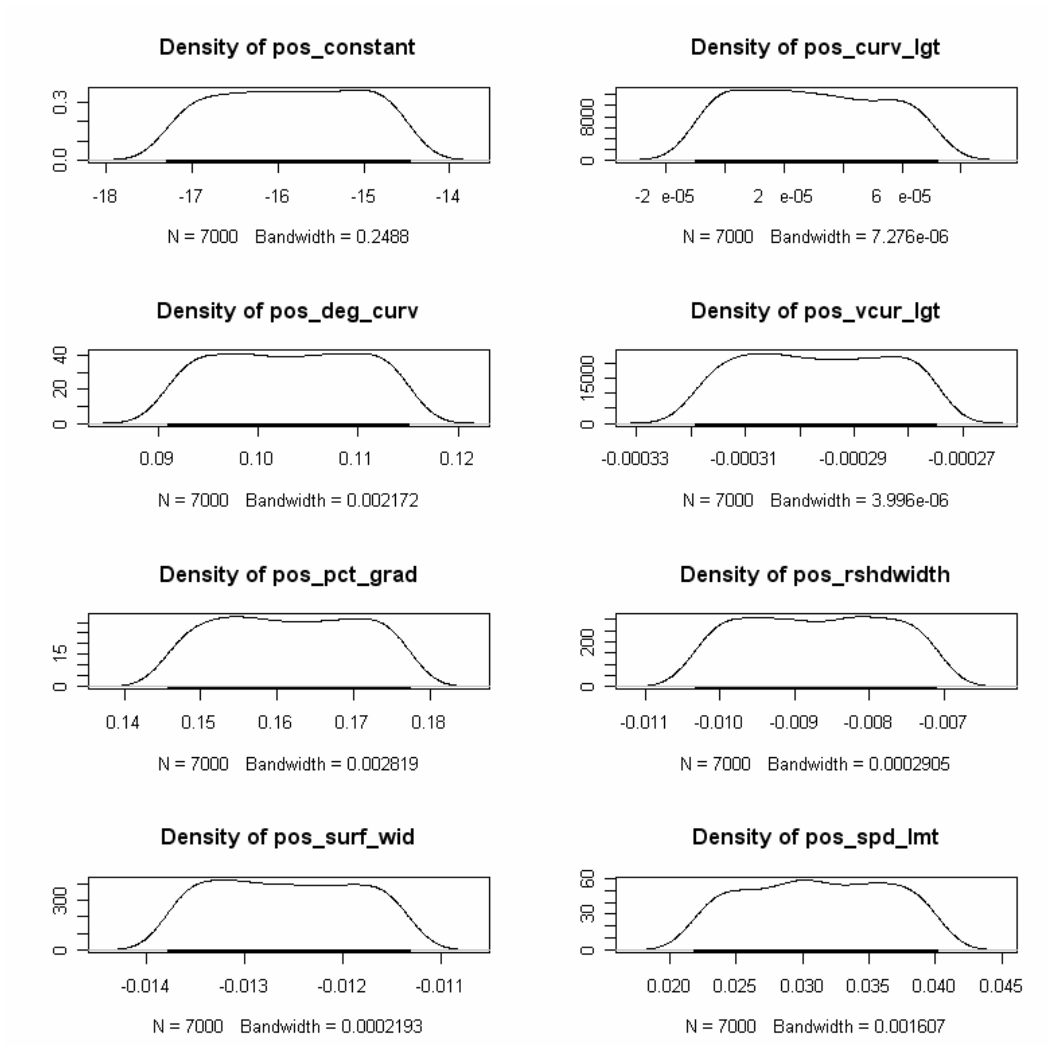


Figure 4a Posterior Density of Variables of Interest for Possible-Injury Crash Frequency

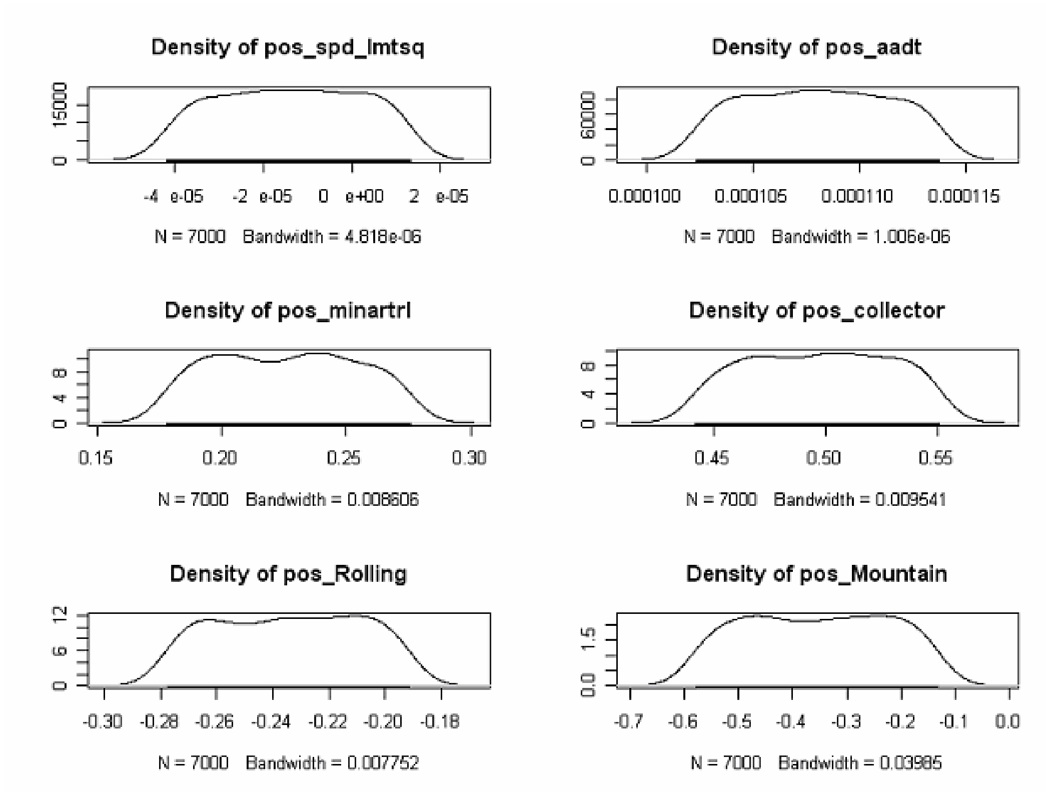


Figure 4b Posterior Density of Variables of Interest for Possible-Injury Crash Frequency

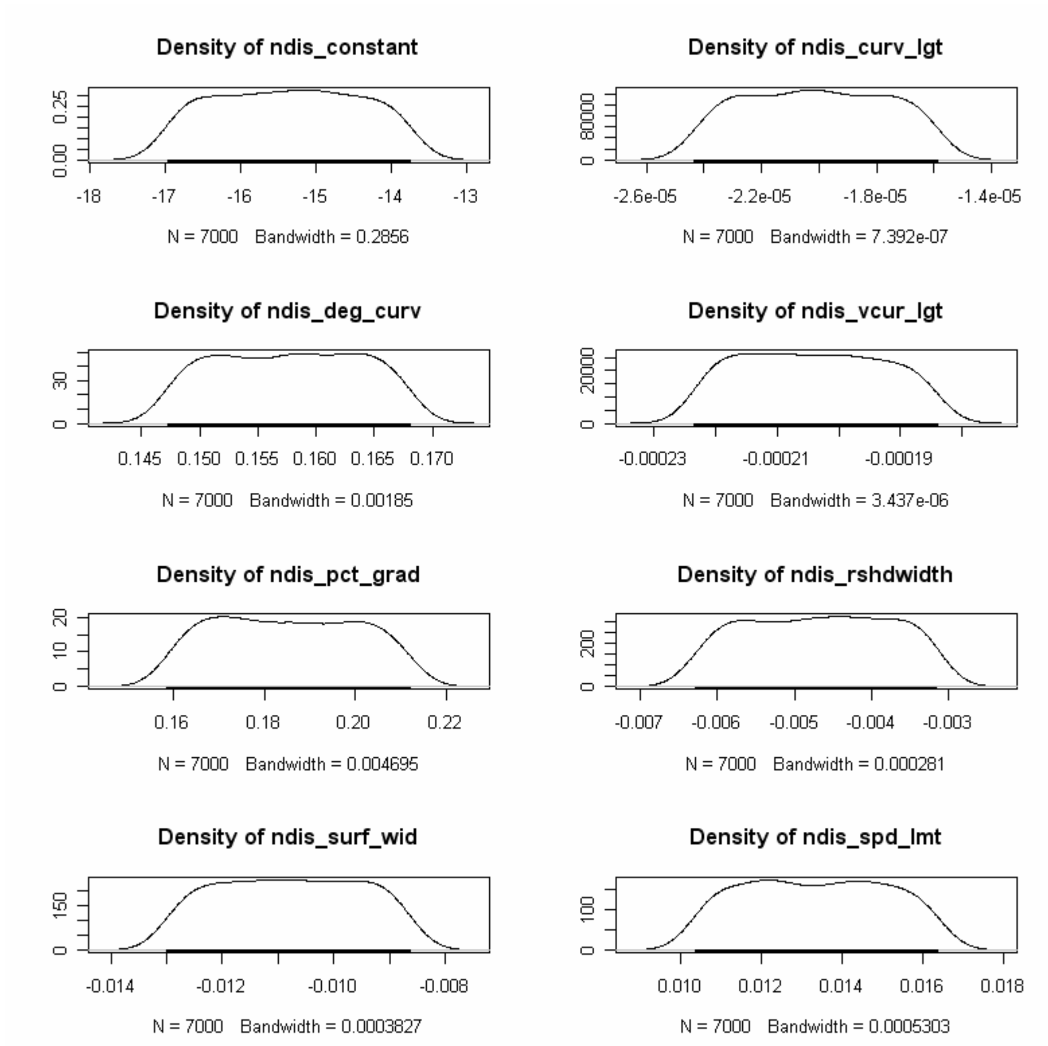


Figure 5a Posterior Density of Variables of Interest for Non-disabling Injury Crash Frequency

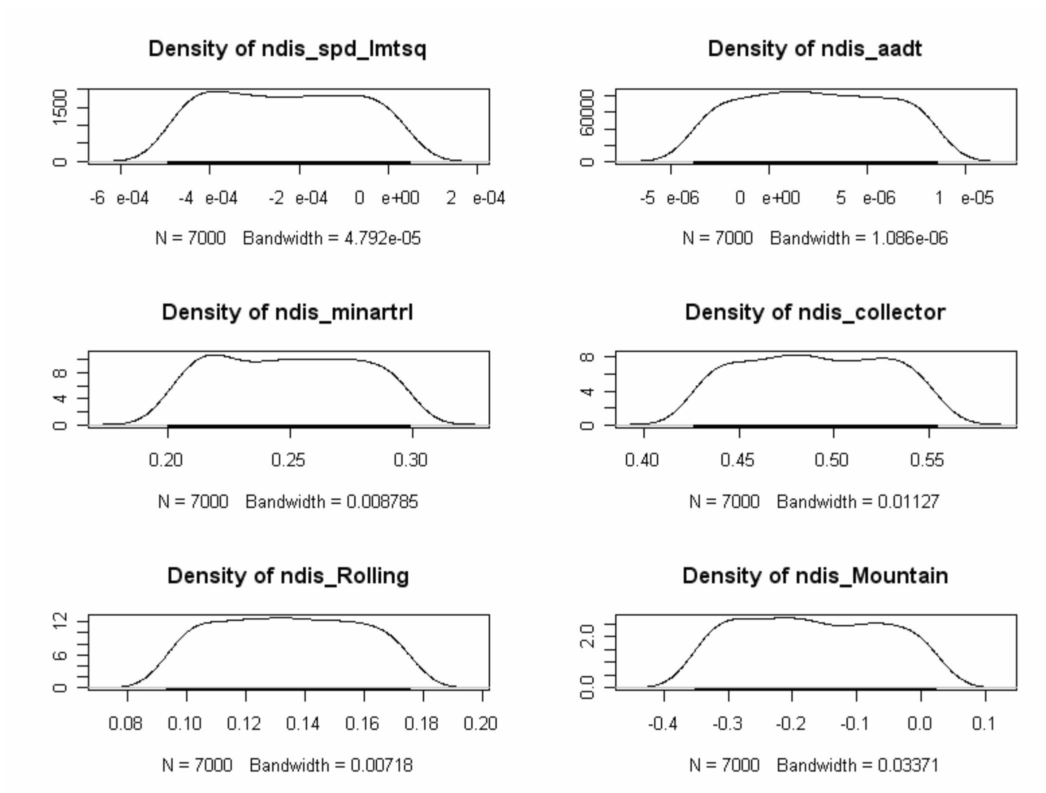


Figure 5b Posterior Density of Variables of Interest for Non-disabling Injury Crash Frequency

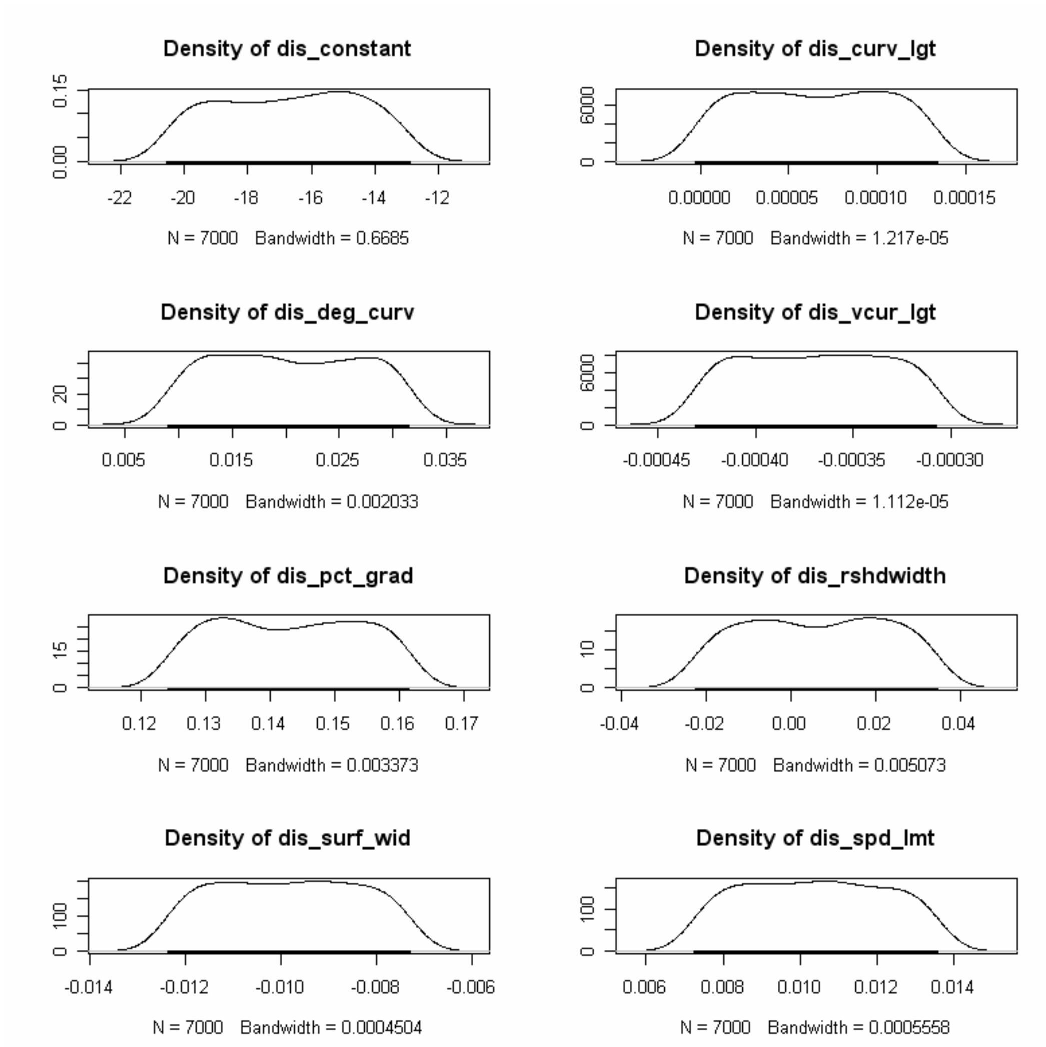


Figure 6a Posterior Density of Variables of Interest for Disabling Injury Crash Frequency

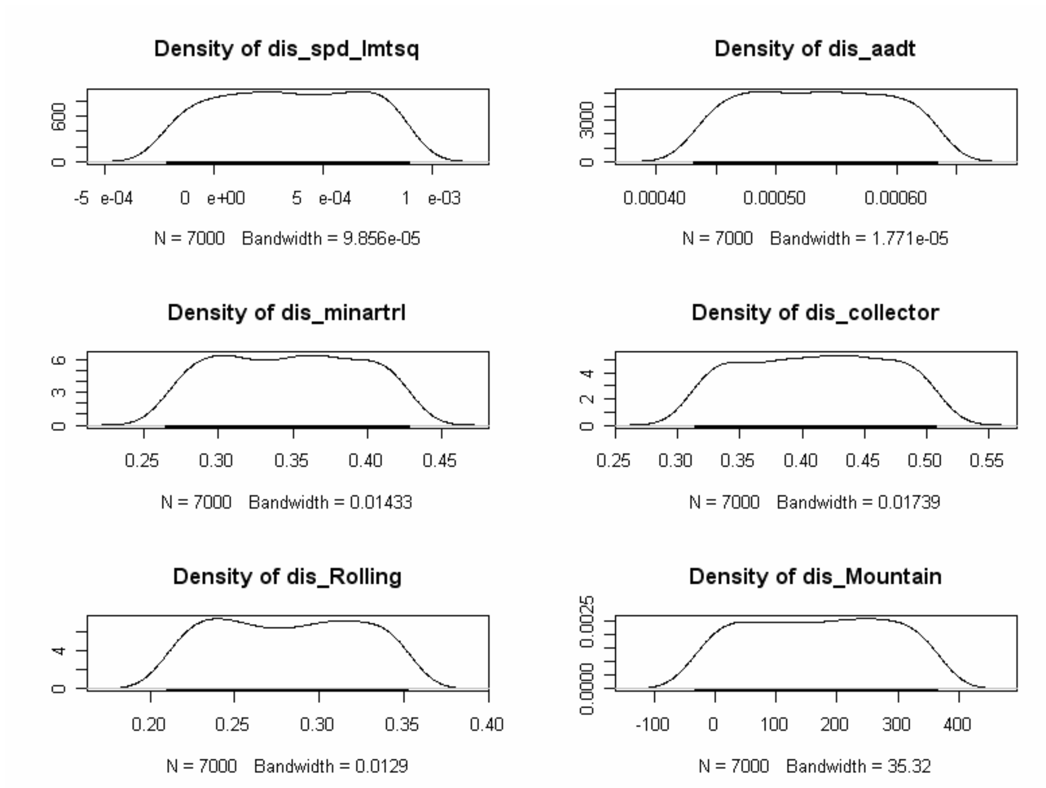


Figure 6b Posterior Density of Variables of Interest for Disabling Injury Crash Frequency

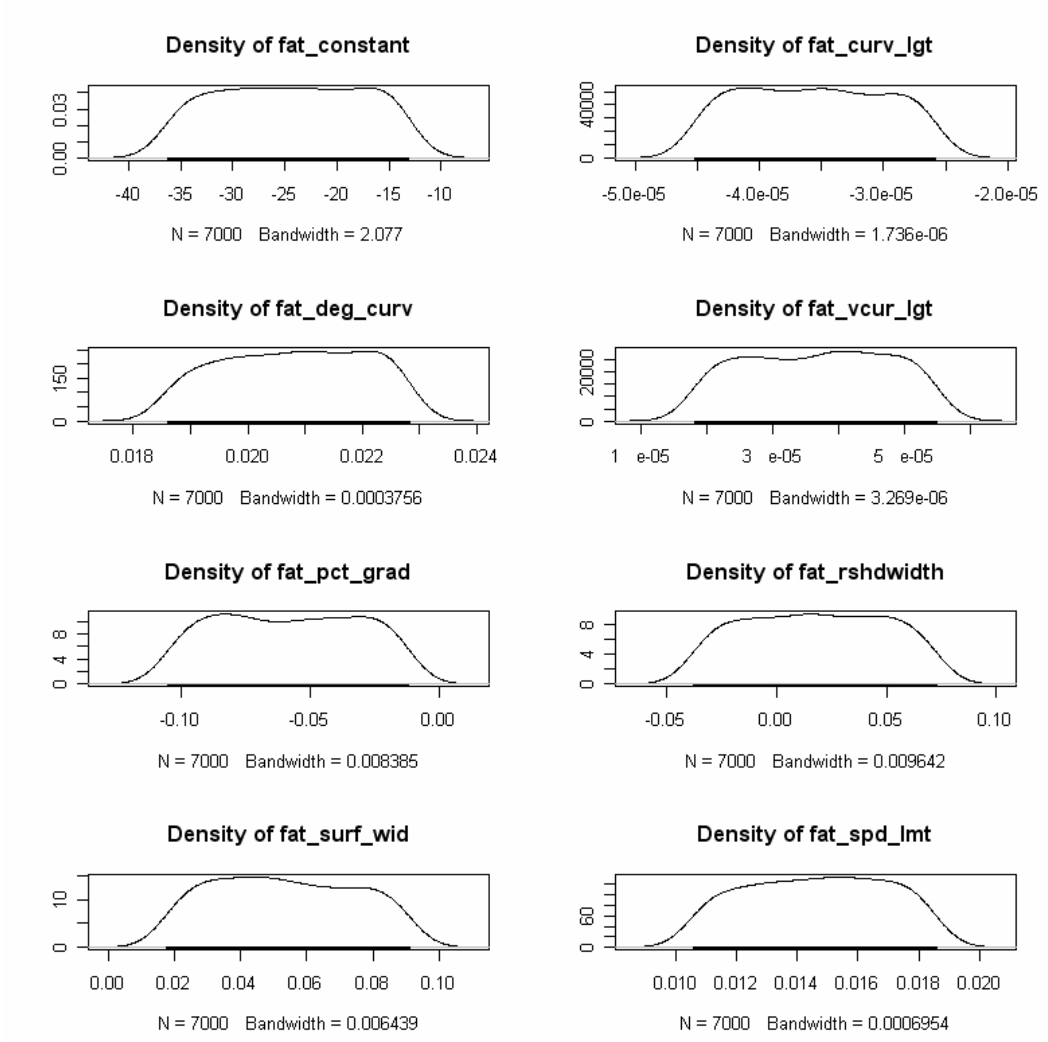


Figure 7a Posterior Density of Variables of Interest for Fatal Crash Frequency

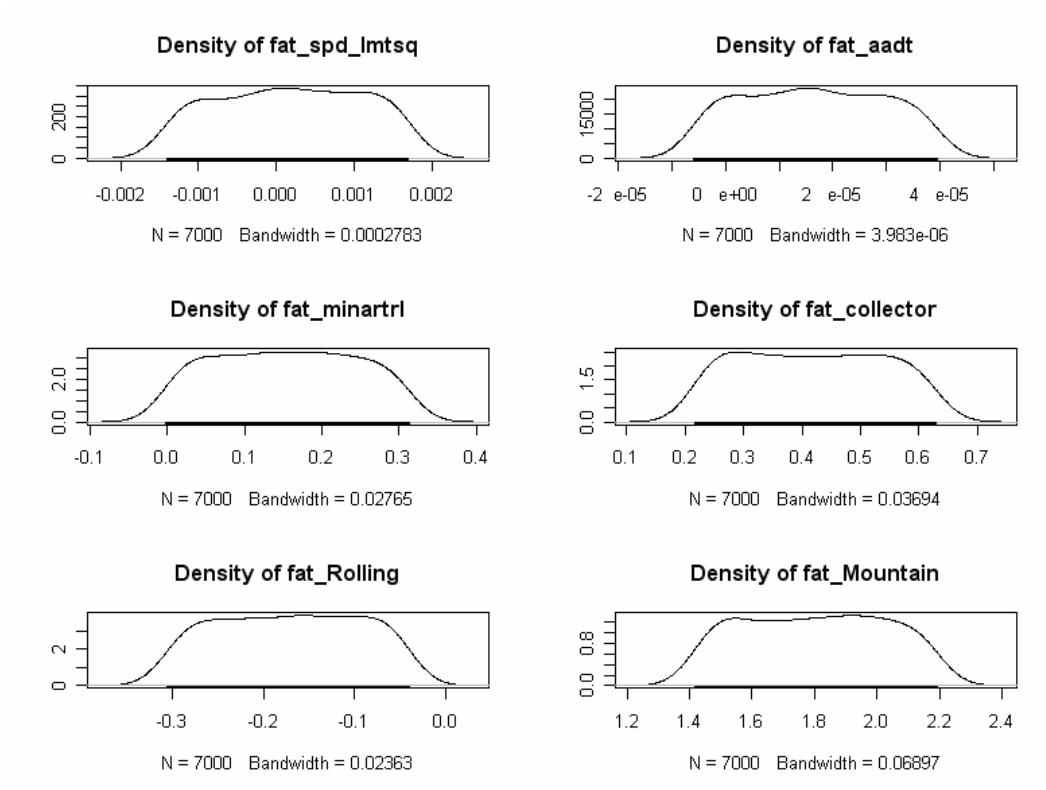
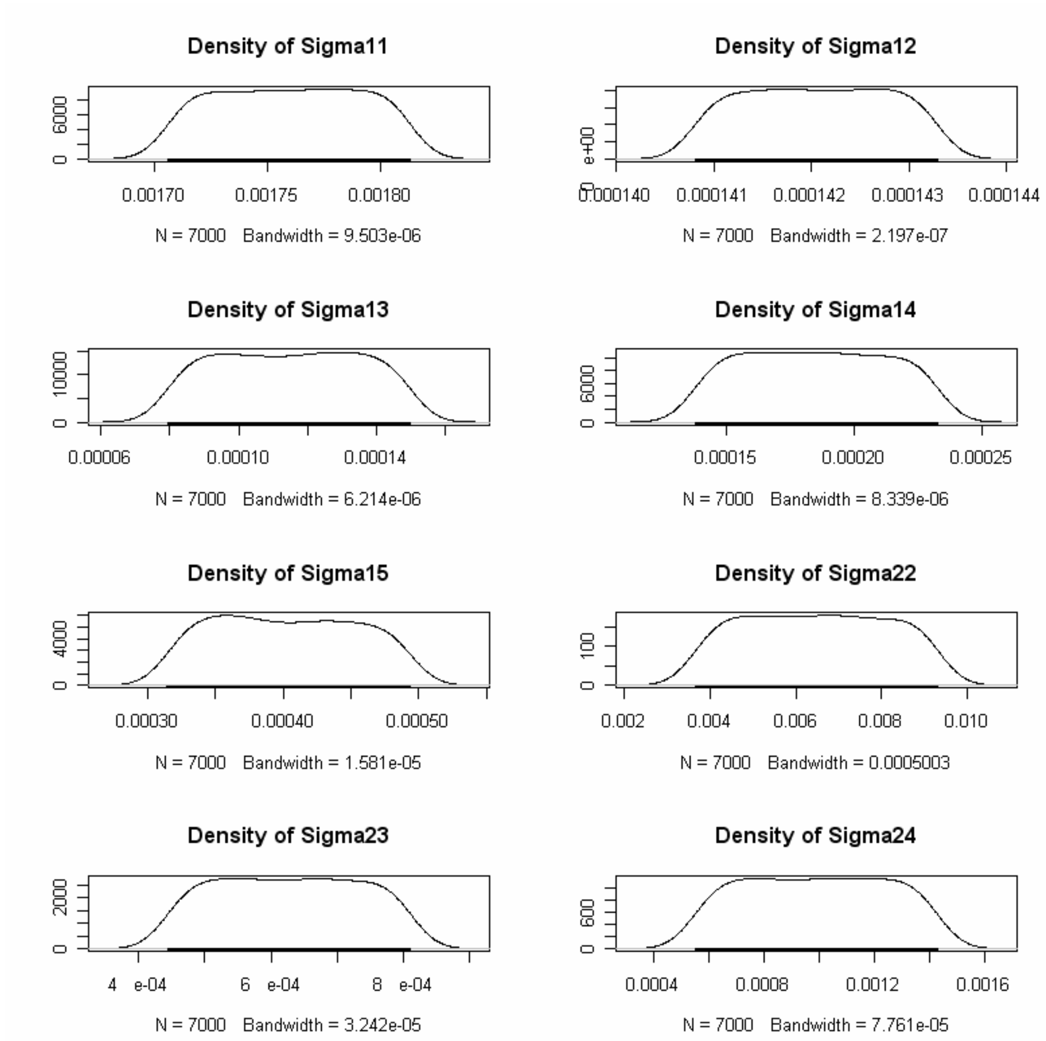


Figure 7b Posterior Density of Variables of Interest for Fatal Crash Frequency



Note: 1 stands for fatal crashes; 2 stands for disabling injury crashes; 3 stands for non-disabling injury crashes; 4 stands for possible injury crashes; 5 stands for PDO crashes.

Figure 8a Posterior Density of Variance-Covariance for ε_{is}

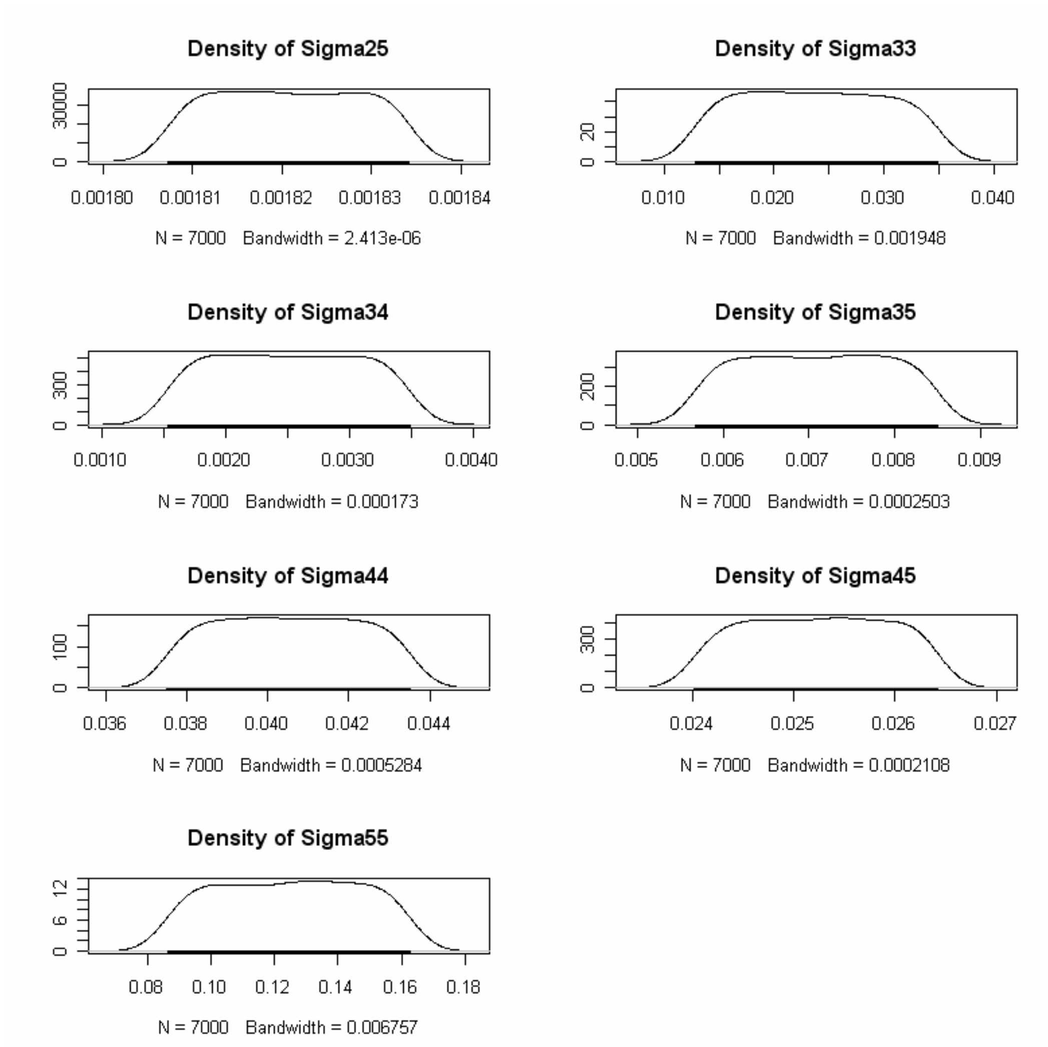


Figure 8b Posterior Density of Variance-Covariance for ε_{is}

Figures C-1 through C-6 in Appendix C display the traces of all of samples for five groups of severity-specific parameters, as well as the variance-covariance matrix, Σ.

5.2.2 Interpretation of Results

The following discussion of results emphasizes disabling and fatal injuries (Tables 9 and 10), since these arguably are of greatest concern to agencies and policymakers. Moreover, the data on such outcomes are more likely to be reported and more reliably recorded than that for other crash outcomes (Blincoe et al. 2002). Tables 6 through 8 provide crash count model estimates for the other three severity levels. The signs of most coefficients are consistent throughout the models, indicating robust directions of effect for most control variables.

Parameter estimates shown in Tables 6 through 10 suggest that roadway design plays an important role in predicting crash counts. For example, holding all other factors fixed, more severe injury crashes are expected on sharper horizontal curves, while wider shoulders tend to reduce rates of less severe crashes (perhaps by offering added maneuverability space for crash avoidance). Based on an average road segment's attributes and the MVPLN model's average parameter estimates, Table 11 provides estimates of percentage changes in crash rates as a function of various design details. For example, a 5-feet increase in (average) right shoulder width (from 2 to 7 feet) is predicted to result in 7.04% fewer crashes (total) per 100 million VMT. Higher average annual daily traffic levels (rising from 3757 to 4757 vehicles) are predicted to increase total crash count by 16.41% — while reducing the total crash *rate* by 5.51%. In this way, the MVPLN model offers statistically (and practically) significant insights into crash counts' dependence on roadway design.

Table 11 Expected Percentage Changes in Crash Rates Corresponding to Changes in Variables

Variables	Averages	Changes in Variable	Percentage change in crash rates (per 100 million VMT)					
			Fatal	Disabling	Non-disabling	Possible	PDO	Total
CURV_LGT	248 (ft)	+100	-0.36%	0.65%	-0.20%	—	—	0.30%
DEG_CURV	2.3 (°/100ft)	+2	4.08%	3.98%	27.04%	18.63%	21.98%	18.58%
VCUR_LGT	303 (ft)	+100	0.37%	-3.76%	-2.06%	-3.01%	-2.08%	-2.52%
PCT_GRAD	1.805	+2	-12.41%	24.88%	30.93%	27.62%	24.07%	24.86%
SHLDWID	2.1 (ft)	+5	—	—	-5.54%	-6.49%	-7.89%	-7.04%
SURF_WID	24 (ft)	+5	-12.52%	-58.65%	-5.36%	-6.49%	4.76%	0.04%
SPD_LIMT	50 (mi/h)	+10	28.97%	38.56%	-12.72%	25.64%	-1.95%	12.99%
AADT	3757	+1000	—	41.37%	—	10.24%	4.68%	16.42%

The magnitudes of the parameter estimates for the MVPLN specification are not directly comparable to those of univariate Poisson models (shown in Appendix C) or those of univariate Negative Binomial (UVNB) models (shown in Appendix C). The reason for this is that the MVPLN model accounts for correlations across crash counts (by severity), and is therefore somewhat different from the univariate models. However, a comparison of parameter signs shows that sharper curves are associated with more fatal crashes in all three models (MVPLN, UVP, and UVNB). The rest of control variables are not statistically significant in both the UVP and UVNB models; however, some of these control variables remain showing a statistically significant effect on fatal crash occurrence in the MVPLN model. For example, speed limit is not statistically significant in the univariate models but is expected to increase fatal crash rates in the MVPLN model. Vertical curve length and segment grade show the same pattern of effects on disabling-injury crashes in all three models. For example, long vertical curves are predicted to reduce disabling-injury crashes, but steeper segments are associated more disabling-injury crashes. The coefficients on remaining control variables are not in agreement across all three models.

The MVPLN specification yields a superior crash prediction model because the crash counts by severity on the same segment of roadway are found to be correlated with one another shown as Table 12. It also allows for overdispersion, as does the MVNB approach. The correlations may be caused by omitted variables (such as pavement quality, sight distance, driveway density, and surrounding land use), which can influence crash occurrence at all levels of severity. Essentially, higher crash rates of one type are associated with higher crash rates of other types. Negative correlations are not likely in models of crash prediction since crash likelihood for all crash types is likely to rise due to the same deficiencies in roadway design, or other unobserved factors.

Table 12 Correlation-Coefficients of $\vec{\boldsymbol{\varepsilon}}_{\mathbf{i}}$

	Fatal	Disabling	Non-Disabling	Possible injury	PDO
Fatal	1	0.04207	0.01777	0.02191	0.02718
Disabling	0.04207	1	0.05061	0.06100	0.4328
Non-Disabling	0.01777	0.05061	1	0.08071	0.1304
Possible injury	0.02191	0.06100	0.08071	1	0.3552
PDO	0.02718	0.4328	0.1304	0.3552	1

So far, our discussion has concentrated on the regression coefficients and the correlation structure. In addition, the model can be used to obtain predictive distributions. The probability functions of the model depend on the slope parameters β_s, the latent effects ε_{is}, and the control variables x_{is}. Theoretically, one can compute the (average) probability of crash counts (by severity) by integrating $\pi\left(y_{is}\left|\varepsilon_{is}, x_i, \beta_s\right.\right)$ for a given count y_{is} and set of control variables over the joint posterior distribution of β_s and ε_i. In practice, this approach is simple to implement because it requires only the output from the MCMC simulation. For each iteration's set of β_s^m and $\left\{\varepsilon_{is}^m\right\}$ from the MCMC simulation, one can compute $\lambda_{is}^m = \exp\left(x_{is}'\beta_s^m + \varepsilon_{is}^m\right)$. To predict marginal probabilities, one can simply average the associated probabilities:

$$\hat{p}_{y_s} = \frac{1}{n}\frac{1}{M}\sum_{i=1}^{n}\sum_{m=1}^{M}\pi_p\left(y_s\left|\lambda_{is}^m\right.\right),\ y_s = 0,1,\ldots, \tag{46}$$

The prediction of joint probabilities (e.g., $P\left(y_{i1} = Y_{i1}, y_{i2} = Y_{i2}, \ldots, y_{iS} = Y_{iS}\right)$) is even more interesting and is generally difficult in such cases that proper modeling of the correlation structure is important. However, in this Bayesian MCMC instance, predictions of joint probabilities for the MVPLN model are quite simple because, conditional on $\vec{\boldsymbol{\varepsilon}}_i$, the crash counts at each level of severity are independently Poisson distributed. Thus, the joint probabilities can be obtained via direct multiplication of these average probabilities.

In addition, out-of-sample predictions can be conducted. In this case, $\vec{\boldsymbol{\varepsilon}}_i$ for the new data points must be drawn from a normal distribution with zero mean and variance-covariance Σ.

$\lambda_{is}^m = \exp\left(x_{is}'\beta_s^m + \varepsilon_{is}^m\right)$ for $m = burnin + 1, burnin + 2, \cdots, M$

Therefore, the prediction distribution can be readily obtained using the above equation.

5.2.3 Example Application: A Cost-Benefit Analysis of Raised Speed Limits

Results in Tables 6 through 10 offer several suggestions for design changes that transportation agencies may consider. As indicated in Table 11, a speed limit increase of 10 mi/h (from 50 mi/h to 60 mi/h, on the "average" roadway section in the database) is predicted to increase fatal crash rates by 28.97% and disabling injury crash rates by 38.56% (according to the MVPLN model's average parameter values). Total crash rates are predicted to increase 12.99% given the same amount of increase in speed limits, everything else constant. One might argue that travel time savings due to a raise in limits can offset the costs of increases in these and other crash outcomes. This section considers this question, as an example application of the model results.

Table C-8 in Appendix C presents estimates of injury costs. Its first two rows summarize a National Highway Traffic Safety Administration (NHTSA) study by Blincoe et al. (2002). The first row presents the "market costs" of injuries (based on medical treatment, emergency services, losses in market and household productivity, insurance administration, workplace cost, and legal costs). The second row gives comprehensive costs, accounting for injured persons' Quality-Adjusted Life Years (QALYs) and accounting for pain and suffering by family members. Since the HSIS database recognizes five injury levels (rather than NHTSA's six), injury costs were calculated using a weighted average of the six MAIS (Maximum Abbreviated Injury Scale)[23] costs.

---

[23] MAIS denotes the highest (maximum) abbreviated injury severity score (AIS) that corresponds to a crash victim's incurred injuries. It can take on values from 0 (minor injury) to 5 (fatal injury).

Table C-9 in Appendix C presents driving speed increases that have been observed in a variety of published studies following speed limit increases[24]. Based on Table C-9, there is approximately a 3.1 mi/h increase in average, observed traffic speeds if speed limits are raised 10 mph. Thus, the time savings per 100 million VMT due to a 10 mph increase in speed limits is estimated to be 94,043 hours. This time savings is equivalent to $1,414,420, assuming a $15.04/vehicle-hour value of travel time savings (USDOT, 1997, 2003). A 10 mph increase in speed limits is predicted to result in 0.68, 2.82, and 6.55 more fatal, disabling injury, and possible injury cases, respectively, and in 3.34 and 1.51 fewer non-disabling injury and no injury cases (per 100 million VMT), respectively. The equivalent average cost estimate for such shifts in crash types is estimated to be $9.02 million (in 2000 dollars, using the values of crash costs in the last row of Table C-8[25]). Therefore, the estimated cost-benefit ratio is 6.4:1. These results suggest that raising speed limits does not offer adequate time-savings benefits. Moreover, if actual travel speeds were to increase one-to-one with speed limits (i.e., by 10 mi/h, rather than 3.1 mi/h), this ratio would fall to 2:1 – but remain greater than 1:1. Thus, the results consistently suggest that a speed limit change (an increase from 50 to 60 mile/hour) on this set of two-lane roadways cannot be justified in terms of time savings benefits.

Let's consider another scenario. Extending the average shoulder width from 2 to 7 feet is expected to result in 10, 11, and 42 fewer non-disabling injury, possible injury and no injury cases on those 7,773 segments annually, respectively. The

[24] Most of the studies listed here (except that in NCHRP Project 17-23) examined speeds on rural interstate highways, following a change from 55 mi/h to 65 mi/h. Kockelman and Bottom's NCHRP study (2005) examined an urban and rural site, both experiencing a 5 mi/h increase. (The resulting average speed change was therefore doubled in that case, to estimate the change that would have occurred had the speed limit change been 10 mi/h.)

[25] Mrozek and Taylor (2002) investigated the value of a statistical life (VOSL) using a meta-analysis. Based on 33 previous studies, they recommended a VOSL of $1.5 to $2.5 million, which is considerably lower than NHTSA's $3.37 million recommendation.

equivalent average benefit estimate for such shits in crash types is estimated to be $462,259 per year. If the cost of one foot of shoulder addition (on both sides of the roadway) is assumed to be $20,000 per mile, then the total cost for treating those 7,773 segments is estimated to $10.18 million. The treatment is a one-time investment. Without any discounting of future crash benefits, it would pay for itself in 22 years. With discounting, it would take longer for expected crash benefits to offset the costs. Thus, adding shoulder width may not be justified in the long run.

5.3 Goodness of Fit and Model Comparison

Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) are used compare alternative models. The AIC and BIC model selection criteria penalize models with additional parameters (Box, Jenkins, and Reinsel 1994).

$$AIC = -2\log\left[L\left(\hat{\theta}\middle|y, X\right)\right] + 2K$$

$$BIC = -2\log\left[L\left(\hat{\theta}\middle|y, X\right)\right] + K \cdot \log\left(n\right)$$

where $L\left[\hat{\theta}\middle|y, X\right]$ refers to the likelihood at convergence, K is the number of parameters, and n is the number of observations.

Calculations of AIC and BIC are shown in Table C-15. Models with lower AIC and BIC criteria are preferred. According to Table C-15, the MVPLN model has both the lowest AIC and BIC measure, compared to the UVP and UVNB models.

In addition, out-of-sample predictions from both univariate and multivariate models are compared one another. Table C-16 suggests that the MVPLN model with MCMC draws predicts better than the univariate models (UVP and UVNB).

This is simply because MVPLN model addresses the issue of unobserved heterogeneity and allows for correlations among crash counts at all levels of severity.

5.4 Conclusions

This study developed a model that allows one to model crash outcomes by severity simultaneously, based on an MVPLN specification that can be estimated within a Bayesian framework using Metropolis-Hastings algorithms within a Gibbs sampler. Crash counts for 7,773 homogeneous segments of rural two-lane roadways in the Puget Sound region of Washington State in 2002 were used to estimate the model.

Thanks to MCMC simulation techniques, the marginal posterior distributions of all parameters of interest were obtained. As expected, positive correlations in unobserved factors affecting count outcomes were found to exist across severity levels, resulting in statistically significant additive latent terms. The non-zero diagonal elements suggest the existence of overdispersion in crash counts at all levels of severity. The estimation results from the MVPLN approach offered more intuitive interpretations and better predictions than those from the univariate Poisson models. As anticipated, the results lend themselves to several recommendations for highway safety treatments and design policies. For example, wide lanes and shoulders are helpful for reducing the total crash frequencies, as are long horizontal and vertical curves. Moreover, using a cost-benefit approach and assumptions about travel speed changes, model results suggest that time savings from raising speed limits 10 mi/h (from 50 to 60 mi/h) may not be worth the added crash cost.

5.5 Summary

This chapter illustrates the merits of an MVPLN model of count data for investigations of the correlated crash counts. The MVPLN specification used here allows for a very general correlation structure for each pair of crash counts at different levels of severity for a particular roadway segment. The empirical results from Metropolis-Hastings algorithms within a Gibbs sampler are summarized. Finally, cost-benefit analyses of raised speed limits and shoulder widening are performed, based on the results. These suggest that speed limit increases along such roads probably is not worth the time savings, and shoulder widening costs may not be worth the crash savings benefits obtained.

CHAPTER 6 CONCLUSIONS AND FURTHER WORK

6.1 Summary

Roadway safety is a major concern for the general public and public agencies. Roadway crashes claim many lives and cause substantial economic losses each year. The situation is of particular interest on rural two-lane roadways, which experience significantly higher fatality rates than urban roads. There have been numerous efforts devoted to investigating crash occurrence as related to roadway design features, environmental conditions and traffic levels. However, most such research has relied on univariate count models; that is, traffic crash counts at different levels of severity have been estimated separately. The widely used univariate count data models ignore the interdependence of crash counts at different levels of severity for a specific segment of roadway.

This research simultaneously models correlated crash counts at different levels of severity using multivariate Poisson-lognormal (MVPLN) regression models. The MVPLN specification allows for a more general correlation structure as well as overdispersion. With recent advancements in crash modeling and Bayesian statistics, the parameter estimation is done within the Bayesian paradigm, using a Gibbs Sampler and Metropolis-Hastings algorithms.

Crash counts for over 7,773 homogeneous segments of rural two-lane Washington State roadways in the Puget Sound region in 2002 were used to estimate the model. Thanks to MCMC simulation techniques, the marginal posterior distributions of all parameters of interest were obtained, and estimation results from the MVPLN approach offered better predictions than those from univariate Poisson and negative binomial models.

As anticipated, the results lend themselves to several recommendations for highway safety treatments and design policies. For example, adding shoulder width is predicted to be highly cost-effective, in terms of the crash cost reductions over the long run.

The MVPLN model's estimation was conducted using MCMC simulation techniques. Could an MLE estimation be performed for the same specification? Theoretically, it is feasible. For example, one can use the technique of maximum simulated likelihood estimator (MSLE) to estimate the parameters in the MVPLN specification. The intricate part is integrating Equation (34) with respect to $\vec{\boldsymbol{\varepsilon}}_i$. According to Greene (2003), the likelihood for segment i can be written as:

$$\begin{aligned} L_i &= \pi\left(\vec{\mathbf{y}}_i \middle| \vec{\boldsymbol{\lambda}}_i, \Sigma\right) \\ &= \int \prod_{s=1}^{S} \pi_p\left(y_{is} \middle| \xi_{is}, \varepsilon_{is}\right) \phi_S\left(\vec{\boldsymbol{\varepsilon}}_i \middle| \mathbf{0}, \Sigma\right) d\vec{\boldsymbol{\varepsilon}}_i \\ &= E_{\vec{\boldsymbol{\varepsilon}}_i}\left[\prod_{s=1}^{S} \pi_p\left(y_{is} \middle| \xi_{is}, \varepsilon_{is}\right)\right] \\ &\approx \frac{1}{M} \sum_{m=1}^{M} \prod_{s=1}^{S} \pi_p\left(y_{is} \middle| \xi_{is}, \varepsilon_{is}^{(m)}\right) \end{aligned} \tag{47}$$

The multivariate normal latent effects $\vec{\boldsymbol{\varepsilon}}_i^{(m)} = \left(\varepsilon_{i1}^{(m)}, \varepsilon_{i2}^{(m)}, \ldots \varepsilon_{iS}^{(m)}\right)$ can be sampled through $\vec{\boldsymbol{\varepsilon}}_i^{(m)} = L\vec{\boldsymbol{\eta}}_i^{(m)}$, where $\vec{\boldsymbol{\eta}}_i^{(m)}$ is a vector of S independent draws from the standard normal distribution and L is the lower triangular matrix of Σ's Cholesky decomposition. Therefore, the log-likelihood of the data sample can be written as follows:

$$\ln L_{data} = \sum_{i=1}^{n} \frac{1}{M} \sum_{m=1}^{M} \prod_{s=1}^{S} \pi_p \left(y_{is} \middle| \xi_{is}, L \right) \tag{48}$$

Then, a maximization procedure can be applied to maximize the log-likelihood (Equation (48)) with respect to β and L.

However, when the dimension of the (intractable) likelihood's integral is large, numerical approximation methods are generally not recommended. For example, Evans and Swartz (1995) conclude that numerical techniques can be very inefficient for integral dimensions exceeding 4. Moreover, they also point out that error estimates for numerical approximation uncertainties are typically hard to come by. For these reasons, Monte Carlo methods are an attractive alternative.

In addition to estimation and application benefits, Monte Carlo methods allow one to make use of Bayes' theorem and incorporate one's prior beliefs and knowledge into the modeling process. Of course, some limitations remain in the current specification.

6.2 Limitations

The current MVPLN specification assumes no spatial correlation across roadway segments. Various unobserved variables may play very similar roles in determining crash frequency on adjacent roadway segments. The assumption of no spatial correlation is actually too strong in this case. These uncontrolled (or simply unobserved) factors may also render significant spatial correlations over time. (See, e.g. Meliker et al. 2004; Miaou, Song, and Mallick 2003; Pawlovich, Souleyrette, and Strauss 2000.)

The framework of this research is established in its parametric assumptions. Parametric methods can be implemented using assumptions of underlying distributions and relationships. Misspecification of the distribution may lead to serious errors in the subsequent data analysis. Semi-parametric and nonparametric regression analysis relaxes these assumptions[26] (see, e.g., Gurmu, Rilstone, and Stern 1999; Wooldridge 1999; Alfò and Trovato 2004). For example, Gurmu, Rilstone and Stern (1999) developed a semiparametric estimation approach to investigate overdispersed count data using series expansion of unknown density of the unobserved heterogeneity. In their work, the distribution of the unobserved heterogeneity was approximated by the Laguerre expansion.

The cost of relaxing such assumption requires more computation and, in some instances, a more difficult-to-understand result. The benefits of nonparametric methods include a potentially more accurate estimate of the regression function

[26] Damien (2005) suggests that nonparametric distributions actually have an infinite-dimensional parameter space. That is, they have too many parameters to be described in the way that parametric distributions are.

and often “exact” probability statements, regardless of the shape of the population distribution from which the random sample was drawn (Damien, 2005).

The MVPLN model estimated here incorporates the safety effects of several roadway design and traffic features of interest to traffic and transportation engineers. However, several features of interest that are not available have been omitted from the model, including, for example, driveway density and sight distance. In addition, the model generally treats the effects of individual geometric design features as independent of one another and ignores potential interactions among them. It is likely that such interactions exist (such as combinations of horizontal and vertical curves on the same segment), and they should be accounted for in the crash prediction model in the future endeavor.

6.3 Extensions

Based on the above limitations and potential applications of the model, several recommendations can be made. These include:

- Modelers should seek to relax the assumption that the latent effects $\vec{\boldsymbol{\varepsilon}}_i$ are independent of the explanatory variables by letting the mean of $\vec{\boldsymbol{\varepsilon}}_i$ be a function of one or more of the available explanatory variables;
- Use a multivariate t distribution, instead of a multivariate normal, as the parametric distribution of $\vec{\boldsymbol{\varepsilon}}_i$ in order to allow for thicker tails;
- Relax the parametric assumption of $\vec{\boldsymbol{\varepsilon}}_i$ using nonparametric methods;
- Account for the spatial and temporal correlations in crash data obtained over several years, across a connected network;
- Calibrate the model using crashes occurring on the entire Washington State highway system;

- Improve data quality and accessibility (for example, driver-based crash data sets can be used to investigate the safety effects of driver attributes and driving behavior);
- Apply the calibrated model to identify critical high-crash sites for design enhancements/treatments;
- Develop a software package to integrate model calibration, crash prediction, and treatment evaluation;
- Extend the model to intersection-related crash predictions.

Several model extensions were discussed in Section 6.2. Calibrating the models using region-wide data is a prerequisite for agencies seeking to identify potential segments for treatment. The crash prediction models developed in this research target roadway segments. In order to obtain a general estimate of an area's or region's crash frequency, intersection- (or driveway-) related crashes should be considered as well. Developing crash prediction models for intersections involves adjusting the assumptions and collecting intersection-related data, such as intersection geometric features, types of traffic control, and designation of travel lanes.

Estimating the models requires a lot of coding work and may not be convenient for practitioners and engineers. To make estimating and updating models easier, the development of a software package oriented towards consolidating all possible models into a single platform is needed.

Data improvement and accessibility efforts would be helpful to further application of the Bayesian models. Improved roadway inventory data, both segment-based and intersection-based, are valuable for the enhanced performance of crash prediction models (e.g., AADT improvements).

6.4 Policy Implications

The crash prediction model is intended to help traffic and transportation engineers, law enforcement agencies and policy makers make unbiased estimates of the expected safety performance of various roadway designs. Crash prediction helps users evaluate one or more proposed design alternatives by comparing the predictions of safety performance.

The MVPLN crash model can better predict crash frequencies than its univariate counterparts by recognizing correlations among crash counts at different levels of severity on an individual roadway segment. The unbiased prediction of crash counts at different levels allows highway agencies to calculate the rankings of safety performance of individual sites more precisely and thus more judicially prioritize safety improvement projects. For example, sharper curves are predicted to result in higher crash rates at all levels of severity. The expected crash frequencies are calculated using the crash prediction model for all sites with horizontal curves. Roadway improvement alternatives and the costs associated with those alternatives are provided based on site-specific attributes. An economic benefit ranking of all sites is estimated by comparing their benefit-cost (b/c) ratios (i.e., the benefit refers to the crash reduction and the cost is the expenditure associated with the treatment, [e.g., elimination of tight horizontal curves]). Higher b/c ratios receive higher rankings and thus higher priority is granted for planning safety improvement programs.

6.5 Concluding Remarks

Traffic crashes remain a major health problem for the U.S. as well as for other countries. Roadway design and speed limit policies are important determinants of crash outcomes. The MVPLN specification yields a superior crash prediction model because crash counts by severity on the same segment of roadway are found to be correlated. The correlations may be caused by omitted variables such as sight distance, pavement quality, land use type, and driveway density, which can influence crash occurrence at all levels of severity. Moreover, the correlation structure is found to be heterogeneous: more PDO crashes generally mean more crashes of other types, ceteris paribus. PDO crashes are found to be most closely correlated with disabling-injury and possible-injury crashes. Negative correlations are not likely to be the case in modeling crash prediction; if one type of crashes exhibits a higher probability of occurrence, the other types of crashes are more likely to occur due to the same deficiencies in roadway design or other crash-causing factors. In addition, the MVPLN model readily allows one to predict the joint probabilities. Bayesian methods offers a valuable framework for tackling this complicated, yet flexible, model of behavior, illuminating relationships that have not been rigorously quantified previously.

APPENDIX A NEWTON-RAPHSON METHOD

The Newton-Raphson method (Mathworld, 2006) can be used to find local maxima (or minima) of a twice-differentiable function $f(x)$ with respect to x, where x can be a scalar or a vector. Its schematic iterative steps are as follows:

$$\max_x f(x)$$

1. Set $x^{(0)} = x_0$ and δ_0 (a vector of thresholds (convergence criteria))
2. $x^{(n+1)} = x^{(n)} - \left[Hf\left(x^{(n)}\right)\right]^{-1} \nabla f\left(x^{(n)}\right)$
3. $n = n + 1$
4. $\delta = x^{(n+1)} - x^{(n)}$

Repeat steps 2 through 4, until $\delta < \delta_0$.

The geometric interpretation of this method is that at each iteration one approximates $f(x)$ via a quadratic function around $x^{(n)}$, and then takes an appropriate step towards the maximum (or minimum) of that function.

APPENDIX B TSIONAS' MULTIVARIATE POISSON REGRESSION MODELS

The following describes Tsionas' (2001) multivariate model of Poisson counts. For ease of presentation, a trivariate MVP mathematical formulation is shown, to analyze counts of crash-involved persons across three levels of injury severity. Extending the specification to accommodate additional levels of severity (e.g., 5 levels) is conceptually and mathematically straightforward. Suppose we have a sample $\{\mathbf{y}_i; i = 1,2,\ldots,n\}$ from a trivariate Poisson distribution, where $\mathbf{y}_i = [y_{i1}, y_{i2}, y_{i3}]'$ denotes the number of crash-involved persons on the i^{th} roadway segment in the sample experiencing no injury (y_{i1}), injury (y_{i2}), and fatal injury (y_{i3}), over a given time period (such as a year). According to Karlis (2003), a rather general trivariate Poisson model can be specified as follows:

$$\mathbf{y}_i = A\mathbf{z}_i \tag{B-1}$$

where $A = [A_1 \quad A_2 \quad A_3]$, $A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $A_2 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$, and $A_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

Substituting matrix A into Equation (B-1), one arrives at the following:

$$\begin{aligned} y_{i1} &= z_{i1} + z_{i12} + z_{i13} + z_{i123} \\ y_{i2} &= z_{i2} + z_{i12} + z_{i23} + z_{i123} \\ y_{i3} &= z_{i3} + z_{i13} + z_{i23} + z_{i123} \end{aligned} \tag{B-2}$$

where all z_{ik} 's are independently Poisson distributed random variables with parameters θ_{ik}, $k \in \{1,2,312,13,23,123\}$. Parameters θ_{ikj} are actually covariance parameters between Y_{ik} and Y_{ij}, and θ_{ikjl} is a common, 3-way covariance parameter among Y_{ik}, Y_{ij}, and Y_{il}.

For ease of implementation, the following assumption is made for the trivariate Poisson distribution, as employed by Tsionas (2001) for his models of forest damage:

$$
\begin{aligned}
y_{i1} &= z_{i1} + \delta_i \\
y_{i2} &= z_{i2} + \delta_i \\
y_{i3} &= z_{i3} + \delta_i
\end{aligned}
\qquad
\begin{cases}
\delta_i = z_{i123} \\
z_{i12} = z_{i13} = z_{i23} = 0
\end{cases}
\tag{B-3}
$$

where $z_{i1}, z_{i2}, z_{i3}, \delta_i$ have independent Poisson distributions with parameters $\theta_{i1}, \theta_{i2}, \theta_{i3}, \lambda$, respectively for each $i = 1,2,\ldots,n$.

Like the univariate Poisson regression, the MVP regression model is constructed so that the parameters depend on explanatory variables $\mathbf{x}_{is}$ $(s = 1,2,3)$.

$$
\begin{aligned}
\theta_{i1} &= E^{\alpha_1} \exp\left(\mathbf{x}'_{i1}\boldsymbol{\gamma}_1\right) \\
\theta_{i2} &= E^{\alpha_2} \exp\left(\mathbf{x}'_{i2}\boldsymbol{\gamma}_2\right) \\
\theta_{i3} &= E^{\alpha_3} \exp\left(\mathbf{x}'_{i3}\boldsymbol{\gamma}_3\right)
\end{aligned}
\tag{B-4}
$$

where $\mathbf{x}_{is}$ and $\boldsymbol{\gamma}_s$ are $p_s \times 1$ column vectors. E^{α_s} denotes an exposure measure (such as VMT), and the exponential transformation ensures non-negativity of crash rates. Equation (B-4) can be further expressed as follows:

$$\begin{aligned} \theta_{i1} &= \exp\left(\mathbf{x}_{i1}'\boldsymbol{\gamma}_1 + \alpha_1 \ln(E)\right) \\ \theta_{i2} &= \exp\left(\mathbf{x}_{i2}'\boldsymbol{\gamma}_2 + \alpha_2 \ln(E)\right) \\ \theta_{i3} &= \exp\left(\mathbf{x}_{i3}'\boldsymbol{\gamma}_3 + \alpha_3 \ln(E)\right) \end{aligned} \quad \Rightarrow \quad \begin{aligned} \theta_{i1} &= \exp\left(\mathbf{x}_{i1}'\boldsymbol{\beta}_1\right) \\ \theta_{i2} &= \exp\left(\mathbf{x}_{i2}'\boldsymbol{\beta}_2\right) \\ \theta_{i3} &= \exp\left(\mathbf{x}_{i3}'\boldsymbol{\beta}_3\right) \end{aligned}$$

where $\begin{aligned} \mathbf{x}_{i1}'\boldsymbol{\beta}_1 &= \mathbf{x}_{i1}'\boldsymbol{\gamma}_1 + \alpha_1 \ln(E) \\ \mathbf{x}_{i2}'\boldsymbol{\beta}_2 &= \mathbf{x}_{i2}'\boldsymbol{\gamma}_2 + \alpha_2 \ln(E) \\ \mathbf{x}_{i3}'\boldsymbol{\beta}_3 &= \mathbf{x}_{i3}'\boldsymbol{\gamma}_3 + \alpha_3 \ln(E) \end{aligned}$.

In this way the set of regressors (and their number) may differ across θ_{is}'s. It also is assumed that δ_i is independent of the $\mathbf{x}_{is}$'s.

For application of computational Bayesian models, the MVP regression model requires a distributional assumption for δ_i, as well as knowledge of each observational unit's contribution to the likelihood, $\mathbf{y}_i \left| \delta_i, \boldsymbol{\beta}, \mathbf{x}_i\right.$, where $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3\right)'$ and $\mathbf{x}_i = \left(\mathbf{x}_{i1}', \mathbf{x}_{i2}', \mathbf{x}_{i3}'\right)'$. Here, the δ_i is assumed to come from a univariate Poisson distribution, with parameter λ. According to Equation (B-3) the likelihood contribution by the i^{th} segment is a product of univariate Poisson distributions with rate parameters $\theta_{i1} + \lambda$, $\theta_{i2} + \lambda$, and $\theta_{i3} + \lambda$. Thus, the joint probability mass function of $\mathbf{y}_i \left| \delta_i, \boldsymbol{\beta}, \mathbf{x}\right.$ can be expressed as follows:

$$p\left(\mathbf{y}_i \middle| \delta_i, \boldsymbol{\beta}, \mathbf{x}\right) = \frac{\exp\left(\mathbf{x}_{i1}'\boldsymbol{\beta}_1\right)^{y_{i1}-\delta_i}}{\exp\left(\exp\left(\mathbf{x}_{i1}'\boldsymbol{\beta}_1\right)\right)\left(y_{i1}-\delta_i\right)!} \frac{\exp\left(\mathbf{x}_{i2}'\boldsymbol{\beta}_2\right)^{y_{i2}-\delta_i}}{\exp\left(\exp\left(\mathbf{x}_{i2}'\boldsymbol{\beta}_2\right)\right)\left(y_{i2}-\delta_i\right)!} \frac{\exp\left(\mathbf{x}_{i3}'\boldsymbol{\beta}_3\right)^{y_{i3}-\delta_i}}{\exp\left(\exp\left(\mathbf{x}_{i3}'\boldsymbol{\beta}_3\right)\right)\left(y_{i3}-\delta_i\right)!} \quad \text{(B-5)}$$

which is simply the product of the individual univariate probability mass functions for each y_{i1}, y_{i2}, and y_{i3}. Thus the likelihood function is simply $L\left(\boldsymbol{\beta}, \{\delta_i, i = 1,2,\ldots,n\} \middle| \mathbf{x}, \mathbf{y}\right) = \prod_{i=1}^{n} p\left(\mathbf{y}_i \middle| \delta_i, \boldsymbol{\beta}, \mathbf{x}\right)$. According to Bayes' theorem, the posterior distribution is proportional to the product of the likelihood function and the joint prior of all parameters, so it must be given by $\left\{\prod_{i=1}^{n} p\left(\mathbf{y}_i \middle| \delta_i, \boldsymbol{\beta}, \mathbf{x}\right) p\left(\delta_i \middle| \lambda\right)\right\} p\left(\boldsymbol{\beta}, \lambda\right)$. Therefore, the kernel posterior distribution of the model is obtained as follows:

$$p\left(\boldsymbol{\beta}, \lambda, \boldsymbol{\delta} \middle| \mathbf{x}, \mathbf{y}\right) \propto \prod_{i=1}^{n} \prod_{s=1}^{3} \frac{\exp\left(\mathbf{x}_{is}'\boldsymbol{\beta}_s\right)^{y_{is}-\delta_i}}{\exp\left(\exp\left(\mathbf{x}_{is}'\boldsymbol{\beta}_s\right)\right)\left(y_{is}-\delta_i\right)!} \exp\left(-n\lambda\right) \prod_{i=1}^{n} \frac{\lambda^{\delta_i}}{\delta_i!} p\left(\boldsymbol{\beta}, \lambda\right) \quad \text{(B-6)}$$

where $\delta_i \le \min\left(y_{i1}, y_{i2}, y_{i3}\right)$, $i = 1,2,\ldots,n$. This constraint is caused by the fact that the variables following Poisson distributions take on only nonnegative integers. Simply put, it is assumed that $\boldsymbol{\beta}$ and λ are independent of $\mathbf{x}$. The parameters $\left(\boldsymbol{\beta}, \lambda\right)$ can be assumed to have the following flat (uninformative) prior.

$$p\left(\boldsymbol{\beta}, \lambda\right) \propto \lambda^{-1} \quad \text{(B-7)}$$

Tsionas (2001) and Ma and Kockelman (2006) have calibrated such models using MCMC simulation. This MVP specification only allows for positive correlation between each pair of counts. Specification with a much more general correlation structure was pursued here instead, in Chapter 4.

APPENDIX C ADDITIONAL TABLES AND FIGURES

Table C-1 PDO Crash Frequency UVP Model for the Puget Sound Region in 2002

Variable definition	Variable	Coef.	Std. Err.	P-value
Constant	CONST	-18.44	1.356	0
Horizontal curve length (feet)	*CURV_LGT*	*-8.65E-05*	*9.40E-05*	*0.358*
Degree of curvature (°/100feet)	DEG_CURV	2.96E-02	8.28E-03	0
Vertical curve length (feet)	VCUR_LGT	-3.85E-04	1.42E-04	0.007
Vertical grade (%)	*PCT_GRAD*	*-0.02892*	*0.02925*	*0.323*
Average shoulder width (feet)	*SHLDWID*	*-0.02113*	*0.03672*	*0.565*
Surface width (feet)	SURF_WID	0.04612	0.007516	0
Posted speed limit (miles/hour)	SPD_LIMT	0.1596	0.06044	0.008
Posted speed limit squared (miles^2/hour^2)	SLSQ	-1.83E-03	6.87E-04	0.008
Average annual daily traffic (AADT)	*AADT*	*2.37E-05*	*1.31E-05*	*0.071*
Indicator for minor arterial: 1=yes, 0=otherwise	*MINARTRL*	*-0.1142*	*0.1127*	*0.311*
Indicator for collector: 1=yes, 0=otherwise	*COLLECTOR*	*0.05141*	*0.1253*	*0.682*
Indicator for rolling terrain: 1=yes, 0=otherwise	*ROLLING*	*0.04746*	*0.09782*	*0.628*
Indicator for mountainous terrain: 1=yes, 0=otherwise	*MOUNTAIN*	*0.4308*	*0.3643*	*0.237*
Number of observations				7,773
Log-Likelihood at convergence				-1,764.2
Log-Likelihood at constant				-1,810.8
LRI				0.0257

Note: Smaller, italicized font is used for parameters that are **not** statistically significant at the 5 percent level.

Table C-2 Possible Injury Crash Frequency UVP Model for the Puget Sound Region in 2002

Variable definition	Variable	Coef.	Std. Err.	P-value
Constant	CONS	-21.09	2.371	0
Horizontal curve length (feet)	*CURV_LGT*	*-6.31E-05*	*1.69E-04*	*0.709*
Degree of curvature (°/100feet)	*DEG_CURV*	*0.02726*	*0.01610*	*0.09*
Vertical curve length (feet)	VCUR_LGT	-1.35E-03	2.86E-04	0
Vertical grade (%)	PCT_GRAD	0.1484	0.04756	0.002
Average shoulder width (feet)	*SHLDWID*	*0.04942*	*0.06133*	*0.42*
Surface width (feet)	SURF_WID	0.07502	0.01007	0
Posted speed limit (miles/hour)	*SPD_LIMT*	*0.1302*	*0.1066*	*0.222*
Posted speed limit squared (miles^2/hour^2)	*SLSQ*	*-1.12E-03*	*1.22E-03*	*0.356*
Average annual daily traffic (AADT)	AADT	7.65E-05	2.01E-05	0
Indicator for minor arterial: 1=yes, 0=otherwise	*MINARTRL*	*0.2269*	*0.1896*	*0.231*
Indicator for collector: 1=yes, 0=otherwise	*COLLECTOR*	*0.3033*	*0.2230*	*0.174*
Indicator for rolling terrain: 1=yes, 0=otherwise	ROLLING	-0.3461	0.1761	0.049
Indicator for mountainous terrain: 1=yes, 0=otherwise	*MOUNTAIN*	*-0.9991*	*1.042*	*0.338*
Number of observations				7,773
Log-Likelihood at convergence				-750.1
Log-Likelihood at constant				-811.1
LRI				0.0751

Note: Smaller, italicized font is used for parameters that are **not** statistically significant at the 5 percent level.

Table C-3 Non-disabling Injury Crash Frequency UVP Model for the Puget Sound Region in 2002

Variable definition	Variable	Coef.	Std. Err.	P-value
Constant	CONS	-14.95	2.065276	0
Horizontal curve length (feet)	CURV_LGT	-3.89E-04	1.96E-04	0.047
Degree of curvature (°/100feet)	DEG_CURV	0.03524	0.01120	0.002
Vertical curve length (feet)	*VCUR_LGT*	*-4.36E-04*	*2.40E-04*	*0.069*
Vertical grade (%)	*PCT_GRAD*	*0.06332*	*0.04574*	*0.166*
Average shoulder width (feet)	*SHLDWID*	*7.20E-04*	*0.06283*	*0.991*
Surface width (feet)	*SURF_WID*	*0.005933*	*0.01734*	*0.732*
Posted speed limit (miles/hour)	*SPD_LIMT*	*-6.26E-03*	*0.09347*	*0.947*
Posted speed limit squared (miles^2/hour^2)	*SLSQ*	*-1.87E-05*	*1.08E-03*	*0.986*
Average annual daily traffic (AADT)	*AADT*	*5.24E-06*	*2.30E-05*	*0.82*
Indicator for minor arterial: 1=yes, 0=otherwise	*MINARTRL*	*0.2020*	*0.1867*	*0.279*
Indicator for collector: 1=yes, 0=otherwise	*COLLECTOR*	*0.2334*	*0.2190*	*0.287*
Indicator for rolling terrain: 1=yes, 0=otherwise	*ROLLING*	*-0.05135*	*0.1670*	*0.759*
Indicator for mountainous terrain: 1=yes, 0=otherwise	*MOUNTAIN*	*-1.073*	*1.036*	*0.3*
Number of observations				7,773
Log-Likelihood at convergence				-780.9
Log-Likelihood at constant				-791.9
LRI				0.0139

Note: Smaller, italicized font is used for parameters that are **not** statistically significant at the 5 percent level.

Table C-4 Disabling Injury Crash Frequency UVP Model for the Puget Sound Region in 2002

Variable definition	Variable	Coef.	Std. Err.	P-value
Constant	CONS	-17.94	5.912	0.002
Horizontal curve length (feet)	*CURV_LGT*	*-2.72E-04*	*3.42E-04*	*0.426*
Degree of curvature (°/100feet)	*DEG_CURV*	*0.02508*	*0.03821*	*0.512*
Vertical curve length (feet)	VCUR_LGT	-9.78E-04	4.95E-04	0.048
Vertical grade (%)	*PCT_GRAD*	*0.1472*	*0.08333*	*0.077*
Average shoulder width (feet)	*SHLDWID*	*-0.2000*	*0.1306*	*0.126*
Surface width (feet)	*SURF_WID*	*-0.05071*	*0.06149*	*0.41*
Posted speed limit (miles/hour)	*SPD_LIMT*	*0.04997*	*0.2493*	*0.841*
Posted speed limit squared (miles^2/hour^2)	*SLSQ*	*1.52E-04*	*2.72E-03*	*0.956*
Average annual daily traffic (AADT)	*AADT*	*1.88E-05*	*4.60E-05*	*0.683*
Indicator for minor arterial: 1=yes, 0=otherwise	*MINARTRL*	*0.2597*	*0.3605*	*0.471*
Indicator for collector: 1=yes, 0=otherwise	*COLLECTOR*	*0.4217*	*0.4204*	*0.316*
Indicator for rolling terrain: 1=yes, 0=otherwise	*ROLLING*	*0.01805*	*0.3230*	*0.955*
Indicator for mountainous terrain: 1=yes, 0=otherwise	*MOUNTAIN*	*-12.27*	*483.6*	*0.98*
Number of observations				7,773
Log-Likelihood at convergence				-270.6
Log-Likelihood at constant				-279.7
LRI				0.0328

Note: Smaller, italicized font is used for parameters that are **not** statistically significant at the 5 percent level.

Table C-5 Fatal Crash Frequency UVP Model for the Puget Sound Region in 2002

Variable definition	Variable	Coef.	Std. Err.	P-value
Constant	CONS	-21.87	12.76	0.087
Horizontal curve length (feet)	*CURV_LGT*	*-4.34E-04*	*6.17E-04*	*0.482*
Degree of curvature (°/100feet)	DEG_CURV	0.06714	0.02076	0.001
Vertical curve length (feet)	*VCUR_LGT*	*-3.21E-04*	*8.02E-04*	*0.689*
Vertical grade (%)	*PCT_GRAD*	*0.03037*	*0.1693*	*0.858*
Average shoulder width (feet)	*SHLDWID*	*-0.1415*	*0.2255*	*0.53*
Surface width (feet)	*SURF_WID*	*-0.05304*	*0.1176*	*0.652*
Posted speed limit (miles/hour)	*SPD_LIMT*	*0.1879*	*0.5281*	*0.722*
Posted speed limit squared (miles^2/hour^2)	*SLSQ*	*-1.27E-03*	*5.66E-03*	*0.822*
Average annual daily traffic (AADT)	*AADT*	*-1.4E-05*	*8.57E-05*	*0.869*
Indicator for minor arterial: 1=yes, 0=otherwise	*MINARTRL*	*0.07047*	*0.6246*	*0.91*
Indicator for collector: 1=yes, 0=otherwise	*COLLECTOR*	*0.1256*	*0.7750*	*0.871*
Indicator for rolling terrain: 1=yes, 0=otherwise	*ROLLING*	*-0.6376*	*0.5621*	*0.257*
Indicator for mountainous terrain: 1=yes, 0=otherwise	*MOUNTAIN*	*-13.23*	*782.81*	*0.987*
Number of observations				7,773
Log-Likelihood at convergence				-100.6
Log-Likelihood at constant				-105.0
LRI				0.0417

Note: Smaller, italicized font is used for parameters that are **not** statistically significant at the 5 percent level.

Table C-6 Estimates of MVPLN Variance-Covariance Matrix Terms (σ_{sl})

Variable Pairs		Mean	Std. Err.	The 95% (2.5-97.5%) HDR	
Fatal	Fatal	1.76E-03	3.10E-05	1.71E-03	1.81E-03
Fatal	Disabling	1.42E-04	7.17E-07	1.41E-04	1.43E-04
Fatal	Non-disabling	1.15E-04	2.03E-05	8.13E-05	1.48E-04
Fatal	Possible injury	1.85E-04	2.72E-05	1.40E-04	2.30E-04
Fatal	PDO	4.02E-04	5.16E-05	3.19E-04	4.89E-04
Disabling	Disabling	6.48E-03	1.63E-03	3.79E-03	9.16E-03
Disabling	Non-disabling	6.27E-04	1.06E-04	4.53E-04	8.02E-04
Disabling	Possible injury	9.88E-04	2.53E-04	5.70E-04	1.40E-03
Disabling	PDO	1.82E-03	7.88E-06	1.81E-03	1.83E-03
Non-disabling	Non-disabling	0.02368	6.36E-03	0.01336	0.03428
Non-disabling	Possible injury	2.50E-03	5.65E-04	1.58E-03	3.43E-03
Non-disabling	PDO	7.08E-03	8.17E-04	5.74E-03	8.42E-03
Possible injury	Possible injury	0.04051	1.72E-03	0.03767	0.04336
Possible injury	PDO	0.025233	6.88E-04	0.02408	0.02636
PDO	PDO	0.124587	0.02205	0.08804	0.1608
Number of observations					7,773

Table C-7 Estimates of Variance-Covariance Matrix of $\vec{\boldsymbol{\varepsilon}}_{\mathbf{i}}$

	Fatal	Disabling	Non-Disabling	Possible injury	PDO
Fatal	0.001760	0.0001421	0.0001147	0.0001850	0.0004025
Disabling	0.0001421	0.006478	0.0006268	0.0009882	0.001821
Non-Disabling	0.0001147	0.0006268	0.02368	0.002500	0.007084
possible injury	0.0001850	0.0009882	0.0025	0.04051	0.02523
PDO	0.0004025	0.001821	0.007084	0.02523	0.1246

Table C-8 NHTSA Estimate of Injury Costs (in 2000 dollars)

	PDO	MAIS 0[27]	MAIS 1	MAIS 2	MAIS 3	MAIS 4	MAIS 5	Fatal
Market Cost ($)	2,532	1,962	10,562	66,820	186,097	348,133	1,096,161	977,208
Comprehensive ($)	2,532	1,962	15,017	157,958	314,204	731,580	2,402,997	3,366,388
% Crashes Unreported (by type)	N/A	21.42%	22.74%	15.83%	6.52%	0.67%	0.00%	N/A
Number of persons involved in reported crashes	N/A	2,002,667	3,599,995	366,987	117,694	36,264	9,463	N/A
Number of persons involved in all crashes	N/A	2,548,571	4,659,585	436,007	125,903	36,509	9,463	N/A
Weight (% of persons involved)	N/A	25.62%	46.06%	4.70%	1.51%	0.46%	0.12%	N/A
Cost per injured person ($)	2,532[28]		10,351			232,890	2,402,997	3,366,388

Source: Blincoe et al. (2002)

[27] MAIS denotes the highest (maximum) abbreviated injury severity score (AIS) that corresponds to a crash victim's incurred injuries. It can take on values from 0 (minor injuries) to 5 (fatal injury).
[28] PDO crashes involve no injuries, so the $2,532 crash cost is based on per crash.

Table C-9 Speed Increases Following a 10 mi/h Speed Limit Increase (from 55 - 65 mi/h)

Studies	Change in Observed Speeds (mi/h)
Brown et al. (1990)	2.4
Freedman and Esterlitz (1990)	2.8
Mace and Heckard (1991)	3.5
NHTSA (1989)	1.9
NHTSA (1992)	3.4
Parker (1997)	0.2-2.3
Pfefer, Stenzel, and Lee (1991)	4-5
Kockelman and Bottom (2006)	3.4-4.8
TRB (1998)	4
Average	3.1

Table C-10 PDO Crash Frequency UVNB Model for the Puget Sound Region in 2002

Variables	Initial Model		Final Model	
	Coef.	P-value	Coef.	P-value
Constant	-17.09	0	-17.78	0
Horizontal curve length (feet)	-9.23E-05	0.393		
Degree of curvature (°/100feet)	0.03063	0.004	0.02589	0.012
Vertical curve length (feet)	-3.47E-04	0.031	-4.12E-04	0.005
Vertical grade (%)	-0.02468	0.453		
Average shoulder width (feet)	-0.02390	0.575		
Surface width (feet)	0.03471	0	0.04223	0
Posted speed limit (miles/hour)	0.1241	0.068	0.1517	0.02
Posted speed limit squared ($miles^2/hour^2$)	-1.53E-03	0.050	-1.86E-03	0.012
Average annual daily traffic (AADT)	2.69E-05	0.103		
Indicator for minor arterial: 1=yes, 0=otherwise	-0.1981	0.132		
Indicator for collector: 1=yes, 0=otherwise	8.27E-03	0.955		
Indicator for rolling terrain: 1=yes, 0=otherwise	-0.01940	0.866		
Indicator for mountainous terrain: 1=yes, 0=otherwise	0.3896	0.321		
Number of observations		7,773		7,773
Log-Likelihood at convergence		-1713.7		-1717.3
Log-Likelihood at constant		-1748.3		-1748.3
LRI		0.0198		0.0177

Table C-11 Possible Injury Crash Frequency UVNB Model for the Puget Sound Region in 2002

Variables	Initial Model		Final Model	
	Coef.	P-value	Coef.	P-value
Constant	-16.48	0	-16.15	0
Horizontal curve length (feet)	-2.80E-05	0.885		
Degree of curvature (°/100feet)	0.01928	0.349		
Vertical curve length (feet)	-1.45E-03	0	-1.31E-03	0
Vertical grade (%)	0.1526	0.007	0.1412	0.008
Average shoulder width (feet)	0.06819	0.365		
Surface width (feet)	0.04487	0.002	0.03602	0.004
Posted speed limit (miles/hour)	-0.04121	0.742		
Posted speed limit squared ($miles^2/hour^2$)	6.87E-04	0.631		
Average annual daily traffic (AADT)	9.19E-05	0.002	8.10E-05	0.002
Indicator for minor arterial: 1=yes, 0=otherwise	0.2907	0.209		
Indicator for collector: 1=yes, 0=otherwise	4.86E-01	0.066		
Indicator for rolling terrain: 1=yes, 0=otherwise	-0.3346	0.108	-0.3582	0.066
Indicator for mountainous terrain: 1=yes, 0=otherwise	-0.8846	0.411		
Number of observations	7,773		7,773	
Log-Likelihood at convergence	-719.9		-723.8	
Log-Likelihood at constant	-754.1		-754.1	
LRI	0.0453		0.0402	

Table C-12 Non-disabling Injury Crash Frequency UVNB Model for the Puget Sound Region in 2002

Variables	Initial Model		Final Model	
	Coef.	P-value	Coef.	P-value
Constant	-16.48	0	-16.15	0
Horizontal curve length (feet)	-2.80E-05	0.885		
Degree of curvature (°/100feet)	0.01928	0.349		
Vertical curve length (feet)	-1.45E-03	0	-1.31E-03	0
Vertical grade (%)	0.1526	0.007	0.1412	0.008
Average shoulder width (feet)	0.06819	0.365		
Surface width (feet)	0.04487	0.002	0.03602	0.004
Posted speed limit (miles/hour)	-0.04121	0.742		
Posted speed limit squared (miles^2/hour^2)	6.87E-04	0.631		
Average annual daily traffic (AADT)	9.19E-05	0.002	8.10E-05	0.002
Indicator for minor arterial: 1=yes, 0=otherwise	0.2907	0.209		
Indicator for collector: 1=yes, 0=otherwise	4.86E-01	0.066		
Indicator for rolling terrain: 1=yes, 0=otherwise	-0.3346	0.108	-0.3582	0.066
Indicator for mountainous terrain: 1=yes, 0=otherwise	-0.8846	0.411		
Number of observations	7,773		7,773	
Log-Likelihood at convergence	-779.1		-780.4	
Log-Likelihood at constant	-789.5		-789.5	
LRI	0.0132		0.0115	

Table C-13 Disabling Injury Crash Frequency UVNB Model for the Puget Sound Region in 2002

Variables	Initial Model		Final Model	
	Coef.	P-value	Coef.	P-value
Constant	-17.94	0.002	-18.80	0
Horizontal curve length (feet)	-2.72E-04	0.426		
Degree of curvature (°/100feet)	0.02508	0.512		
Vertical curve length (feet)	-9.78E-04	0.048	-9.68E-04	0.045
Vertical grade (%)	0.1472	0.077	0.1379	0.072
Average shoulder width (feet)	-0.2000	0.126	-0.2108	0.081
Surface width (feet)	-0.05068	0.410		
Posted speed limit (miles/hour)	0.04984	0.842	0.05652	0.030
Posted speed limit squared (miles^2/hour^2)	1.53E-04	0.955		
Average annual daily traffic (AADT)	1.88E-05	0.683		
Indicator for minor arterial: 1=yes, 0=otherwise	0.2598	0.471		
Indicator for collector: 1=yes, 0=otherwise	4.22E-01	0.316		
Indicator for rolling terrain: 1=yes, 0=otherwise	0.01808	0.955		
Indicator for mountainous terrain: 1=yes, 0=otherwise	-15.98	0.996		
Number of observations	7,773		7,773	
Log-Likelihood at convergence	-270.6		-273.1	
Log-Likelihood at constant	-279.7		-279.7	
LRI	0.0328		0.0238	

Table C-14 Fatal Crash Frequency UVNB Model for the Puget Sound Region in 2002

Variables	Initial Model		Final Model	
	Coef.	P-value	Coef.	P-value
Constant	-21.86	0.087	-17.67	0
Horizontal curve length (feet)	-4.34E-04	0.482		
Degree of curvature (°/100feet)	0.06714	0.001	0.05459	0.004
Vertical curve length (feet)	-3.21E-04	0.689		
Vertical grade (%)	0.03037	0.858		
Average shoulder width (feet)	-0.1415	0.530		
Surface width (feet)	-0.05304	0.652		
Posted speed limit (miles/hour)	0.1879	0.722		
Posted speed limit squared ($miles^2/hour^2$)	-1.27E-03	0.822		
Average annual daily traffic (AADT)	-1.42E-05	0.869		
Indicator for minor arterial: 1=yes, 0=otherwise	0.07045	0.910		
Indicator for collector: 1=yes, 0=otherwise	0.1256	0.871		
Indicator for rolling terrain: 1=yes, 0=otherwise	-0.6375	0.257		
Indicator for mountainous terrain: 1=yes, 0=otherwise	-18.56	0.999		
Number of observations	7,773		7,773	
Log-Likelihood at convergence	-100.6		-103.4	
Log-Likelihood at constant	-105.0		-105.0	
LRI	0.0417		0.0143	

Table C-15 Goodness of Fit Measures for Univariate and Multivariate Models

Models	Criteria	Loglikelihood at convergence						K	log(n)	Goodness of fit
		PDO	Possible-injury	Non-disabling	Disabling	Fatal	Total			
UVP	AIC	-1764.2	-750.1	-780.9	-270.6	-100.6	-3666.4	70		7473
	BIC	-1764.2	-750.1	-780.9	-270.6	-100.6	-3666.4	70	3.891	7605
UVNB	AIC	-1717.3	-723.8	-780.4	-273.1	-103.4	-3598	75		7346
	BIC	-1717.3	-723.8	-780.4	-273.1	-103.4	-3598	75	3.891	7488
MVPLN	AIC						-3524	85		7218
	BIC						-3524	85	3.891	7379

Table C-16 Comparisons of Crash Predictions from Univariate and Multivariate Models

		PDO	Possible	Non-disabling	Disabling	Fatal
Observed		981	331	287	83	23
UVP	Prediction	1050	432.6	384.3	120.8	30.44
	Difference	69.24	101.6	97.32	37.77	7.444
	Percentage Difference	7.06%	30.70%	33.91%	45.51%	32.37%
UVNB	Prediction	1039	396.5	345.4	104.8	29.91
	Difference	58	65.5	58.4	21.8	6.91
	Percentage Difference	5.91%	19.79%	20.35%	26.27%	30.04%
MVPLN1[29]	Prediction	1013	358.2	310.1	96.8	27.13
	Difference	32	27.2	23.1	13.8	4.13
	Percentage Difference	3.26%	8.22%	8.05%	16.63%	17.96%
MVPLN2[30]	Prediction	1005	348.3	306.4	97.17	26.52
	Difference	24	17.3	19.4	14.17	3.52
	Percentage Difference	2.45%	5.23%	6.76%	17.07%	15.30%

Note: A total of 13,050 rural two-lane road segments in the Puget Sound region were used for model prediction.

[29] The MVPLN1 prediction is conducted as follows: (1) sampling 1,000 times of severity-specific parameters from a multivariate normal distribution with a mean equal to the posterior mean and correlation correlations shown in Tables C-17 through C-21; (2) sampling 1,000 times observed heterogeneity from a multivariate normal with a mean equal to zero and variance-covariance matrix shown in Table C-7; (3) calculating the expected crash counts for each segment; (4) repeat (1) through (3) for all segments in the sample.

[30] The MVPLN2 prediction is obtained as follows: (1) sampling 7,000 draws of unobserved heterogeneity from a multivariate normal with a mean equal to zero and variance-covariance matrix shown in Table C-7; (2) calculating the expected crash counts for each segment using the 7,000 draws of unobserved heterogeneity and 7,000 draws from the MCMC simulation; (3) repeat (1) and (2) for all segments in the sample.

Table C-17 Correlation Coefficients of Parameters in MVPLN Model for Fatal Crashes

	Constant	Curv_Lgt	Deg_Curv	Vcur_Lgt	Pct_Grad	ShldWid	Surf_Wid	Spd_Limt	SLSQ	AADT	MinArtrl	Collector	Rolling	Mountain
Constant	1													
Curv_Lgt	-0.01343	1												
Deg_Curv	-0.004582	-0.001868	1											
Vcur_Lgt	0.01501	0.01887	0.009172	1										
Pct_Grad	-0.01110	0.01154	0.01526	0.007751	1									
ShldWid	0.02150	0.005651	0.003084	-0.01261	-0.001543	1								
Surf_Wid	0.01420	0.007486	0.02799	0.01032	0.01631	0.01820	1							
Spd_Limt	0.01030	0.02678	0.001745	-0.003702	-0.02799	-0.01632	0.01940	1						
SLSQ	-0.05660	-0.02688	-0.01221	-0.04563	0.01200	-0.01393	-0.01812	-0.04098	1					
AADT	0.02834	0.008708	0.002817	-0.01628	0.02800	0.001875	0.01249	0.03559	-0.03742	1				
MinArtrl	-0.02633	-0.009804	0.03591	0.02043	0.004734	-0.02041	0.005550	0.02815	0.01159	0.02883	1			
Collector	0.02893	-0.008190	0.01668	0.02427	-0.006788	-0.0002803	0.01970	0.01561	0.01242	-0.03088	0.01760	1		
Rolling	0.001899	-0.001838	0.005877	-0.008517	0.01643	-0.01511	-0.01081	-0.03902	0.02561	-0.02529	-0.009289	-0.01299	1	
Mountain	-0.01930	0.008078	0.007216	0.03399	0.003029	0.002417	0.01359	0.005908	0.001364	0.008146	0.01321	0.01703	0.004092	1

Table C-18 Correlation Coefficients of Parameters in MVPLN Model for Disabling-Injury Crashes

	Constant	Curv_Lgt	Deg_Curv	Vcur_Lgt	Pct_Grad	ShldWid	Surf_Wid	Spd_Limt	SLSQ	AADT	MinArtrl	Collector	Rolling	Mountain
Constant	1													
Curv_Lgt	-0.03349	1												
Deg_Curv	-0.01689	0.01520	1											
Vcur_Lgt	0.004343	0.004425	-0.004708	1										
Pct_Grad	-0.04073	-0.01996	-0.004961	0.01902	1									
ShldWid	0.01854	0.005658	0.04584	-0.001640	-0.01518	1								
Surf_Wid	-0.01982	0.01377	-0.003527	0.03538	-0.01076	0.01681	1							
Spd_Limt	-0.003663	-0.01109	0.01264	-0.004282	0.003861	-0.002582	-0.03839	1						
SLSQ	-0.007021	0.01154	0.01966	-0.004871	0.0005451	0.01755	0.0003000	0.006587	1					
AADT	-0.01481	-0.004521	-0.002455	0.03135	0.04420	0.002546	-0.01549	-0.004950	0.01041	1				
MinArtrl	0.001178	-0.01235	-0.02254	0.02785	0.007074	0.006722	-0.001005	0.03102	0.01901	0.0006053	1			
Collector	0.002881	0.02120	0.04166	-0.009675	0.01536	0.001207	0.02519	-0.01007	0.01818	-0.002665	-0.01451	1		
Rolling	-0.01287	0.02988	0.02117	0.0007858	0.003658	0.04673	-0.01073	0.009645	0.02037	0.01710	0.01520	0.01696	1	
Mountain	-0.02789	-0.01183	-0.05177	-0.01542	-0.033728	0.02190	0.006287	0.03243	0.01578	-0.03563	0.009335	-0.02734	-0.004617	1

Table C-19 Correlation Coefficients of Parameters in MVPLN Model for Non-Disabling-Injury Crashes

	Constant	Curv_Lgt	Deg_Curv	Vcur_Lgt	Pct_Grad	ShldWid	Surf_Wid	Spd_Limt	SLSQ	AADT	MinArtrl	Collector	Rolling	Mountain
Constant	1													
Curv_Lgt	-0.02521	1												
Deg_Curv	0.01449	-0.01642	1											
Vcur_Lgt	-0.01906	0.02595	0.006466	1										
Pct_Grad	-0.01120	-0.01788	0.007416	-0.02578	1									
ShldWid	-0.009898	-0.003142	0.01330	0.007704	0.03767	1								
Surf_Wid	0.002384	0.01636	0.003462	-0.003672	0.009627	-0.003707	1							
Spd_Limt	-0.01632	-0.003669	0.02238	0.002325	0.02308	-0.007268	-0.01876	1						
SLSQ	-0.01560	-0.005109	-0.009308	-0.02010	0.03186	0.003606	0.01743	-0.03537	1					
AADT	0.007003	0.03785	0.02725	0.008597	0.01169	0.01395	-0.04511	0.0007170	0.003543	1				
MinArtrl	-0.01670	-0.04109	-0.007152	-0.02022	0.03477	0.007583	-0.03744	-0.004120	-0.03201	-0.0003203	1			
Collector	0.05215	-0.01832	0.01113	-0.04678	-0.007941	-0.01537	-0.003741	-0.009022	0.02956	0.03073	-0.02753	1		
Rolling	-0.009791	-0.001440	-0.01440	-0.01558	-0.01787	0.0003072	0.03146	0.002079	0.006493	-0.022265	-0.003562	-0.002095	1	
Mountain	-0.02962	-0.005319	-0.003764	0.0006608	-0.03544	0.001765	-0.01478	-0.03723	0.01573	0.02286	-0.01266	0.01828	-0.03760	1

Table C-20 Correlation Coefficients of Parameters in MVPLN Model for Possible-Injury Crashes

	Constant	Curv_Lgt	Deg_Curv	Vcur_Lgt	Pct_Grad	ShldWid	Surf_Wid	Spd_Limt	SLSQ	AADT	MinArtrl	Collector	Rolling	Mountain
Constant	1													
Curv_Lgt	0.03557	1												
Deg_Curv	0.01057	-0.03680	1											
Vcur_Lgt	0.01878	0.03695	0.007838	1										
Pct_Grad	-0.008174	-0.0005354	0.01286	0.001742	1									
ShldWid	0.01215	0.008459	-0.01145	-0.002838	-0.03126	1								
Surf_Wid	-0.01171	-0.01663	0.01290	0.001563	-0.005137	-0.003294	1							
Spd_Limt	0.02200	-0.00597	-0.01977	-0.01067	0.006562	-0.01928	-0.001342	1						
SLSQ	0.002949	-0.02592	-0.004821	0.008377	0.01336	0.01214	0.0008778	0.03568	1					
AADT	0.01462	-0.003448	0.02604	-0.01214	0.01200	0.01591	0.01340	0.01644	-0.008874	1				
MinArtrl	-0.01490	-0.04074	-0.005677	-0.001230	-0.009028	-0.02816	-0.04561	-0.02721	-0.02560	0.004886	1			
Collector	0.02542	0.01497	0.002020	-0.01139	0.01047	0.00744	0.002807	-0.005115	0.007811	-0.005183	0.002620	1		
Rolling	0.03731	0.02735	0.01157	0.02099	-0.01296	0.02397	-0.001563	0.005747	0.006111	-0.01547	-0.003049	0.02495	1	
Mountain	-0.01748	0.003174	-0.01998	0.01875	0.01120	0.001259	0.002732	0.02094	-0.005152	-0.003546	-0.009045	0.01395	0.007382	1

Table C-21 Correlation Coefficients of Parameters in MVPLN Model for PDO Crashes

	Constant	Curv_Lgt	Deg_Curv	Vcur_Lgt	Pct_Grad	ShldWid	Surf_Wid	Spd_Limt	SLSQ	AADT	MinArtrl	Collector	Rolling	Mountain
Constant	1													
Curv_Lgt	-0.01306	1												
Deg_Curv	-0.03180	0.002179	1											
Vcur_Lgt	0.02928	-0.002490	0.0009977	1										
Pct_Grad	0.02483	-0.04760	-0.01990	0.01780	1									
ShldWid	0.01485	0.01152	0.004961	0.004663	0.004494	1								
Surf_Wid	-0.02565	-0.006182	-0.01361	-0.01320	0.01823	-0.0008455	1							
Spd_Limt	-0.01513	0.0008404	0.01508	-0.002541	-0.003405	-0.001476	0.01222	1						
SLSQ	0.01493	-0.02980	-0.02565	-0.03236	-0.03114	-0.02842	0.03393	-0.002728	1					
AADT	-0.02171	0.0005689	-0.01200	0.01066	0.01687	-0.004728	0.005253	-0.002266	-0.01787	1				
MinArtrl	-0.003412	-0.01927	-0.02224	0.02321	-0.009891	0.005623	-0.007263	-0.05035	-0.002551	-0.02611	1			
Collector	0.009972	0.01322	0.008248	-0.006531	-0.002289	-0.004090	-0.009070	0.01058	-0.005172	0.01055	-0.02546	1		
Rolling	0.004693	0.01626	0.008150	-0.01386	0.01198	0.01006	0.006306	0.01681	-0.002518	-0.007526	-0.0002404	-0.02572	1	
Mountain	0.01250	-0.003033	-0.01440	0.003487	-0.007233	-0.003785	0.02125	-0.03292	0.004520	0.02707	0.005514	-0.01307	-0.003973	1

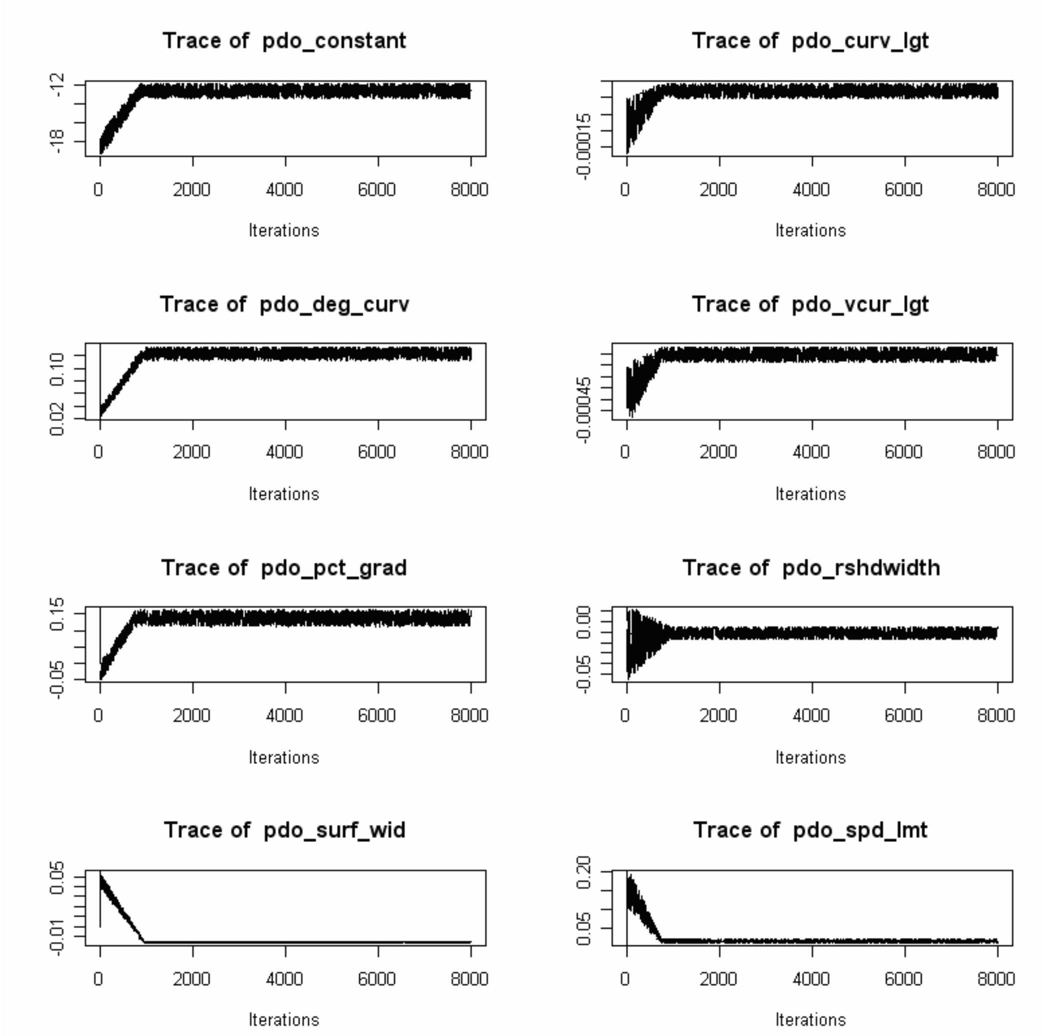


Figure C-1a Trace of Variables of Interest for PDO Crash Frequency

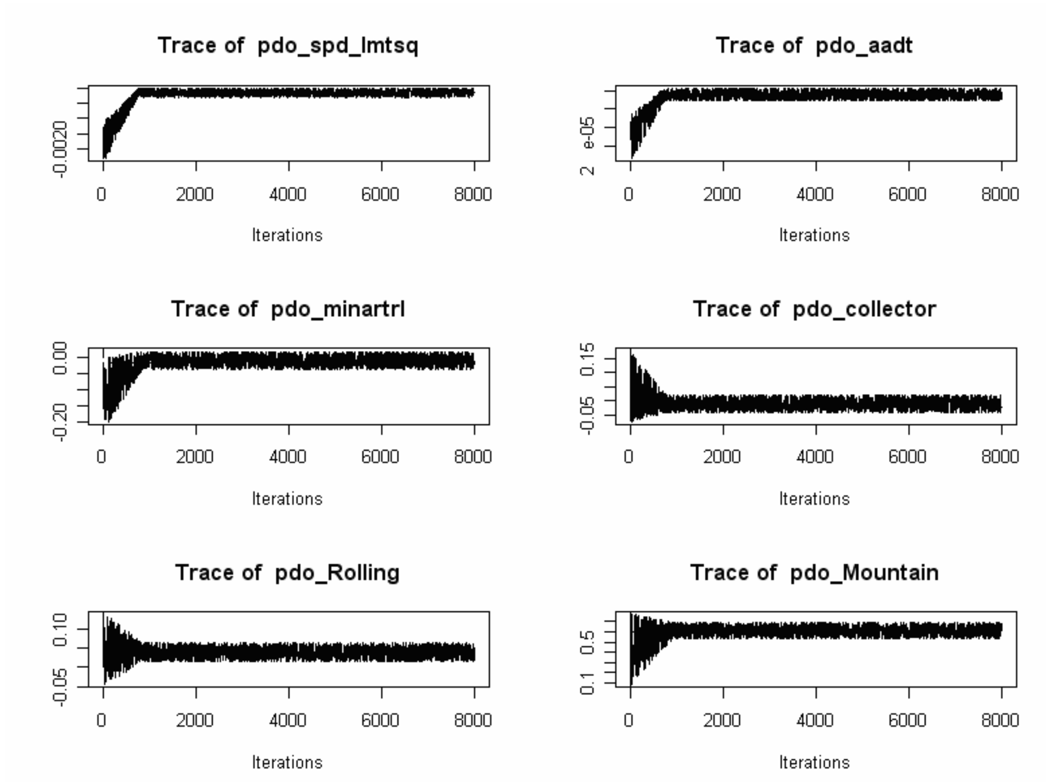


Figure C-1b Trace of Variables of Interest for PDO Crash Frequency

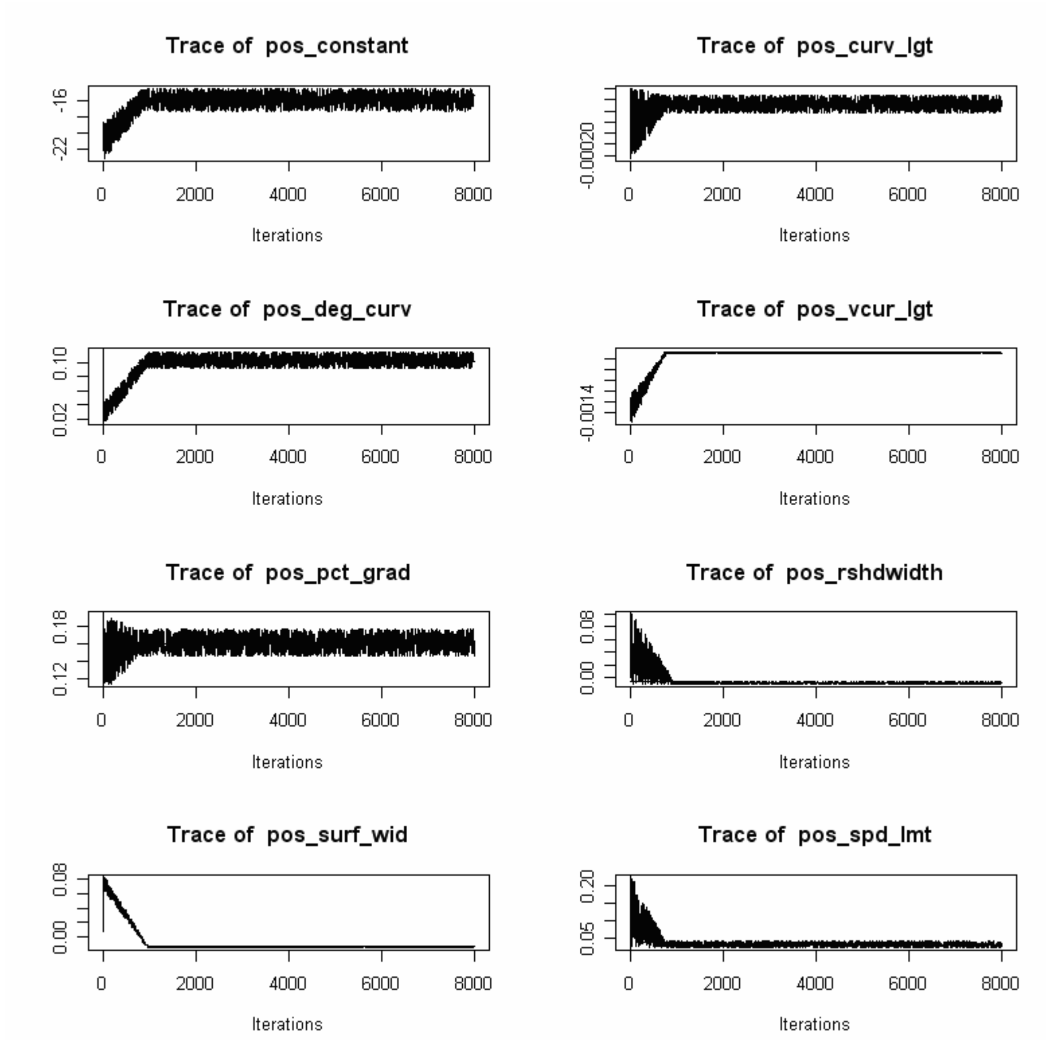


Figure C-2a Trace of Variables of Interest for Possible Injury Crash Frequency

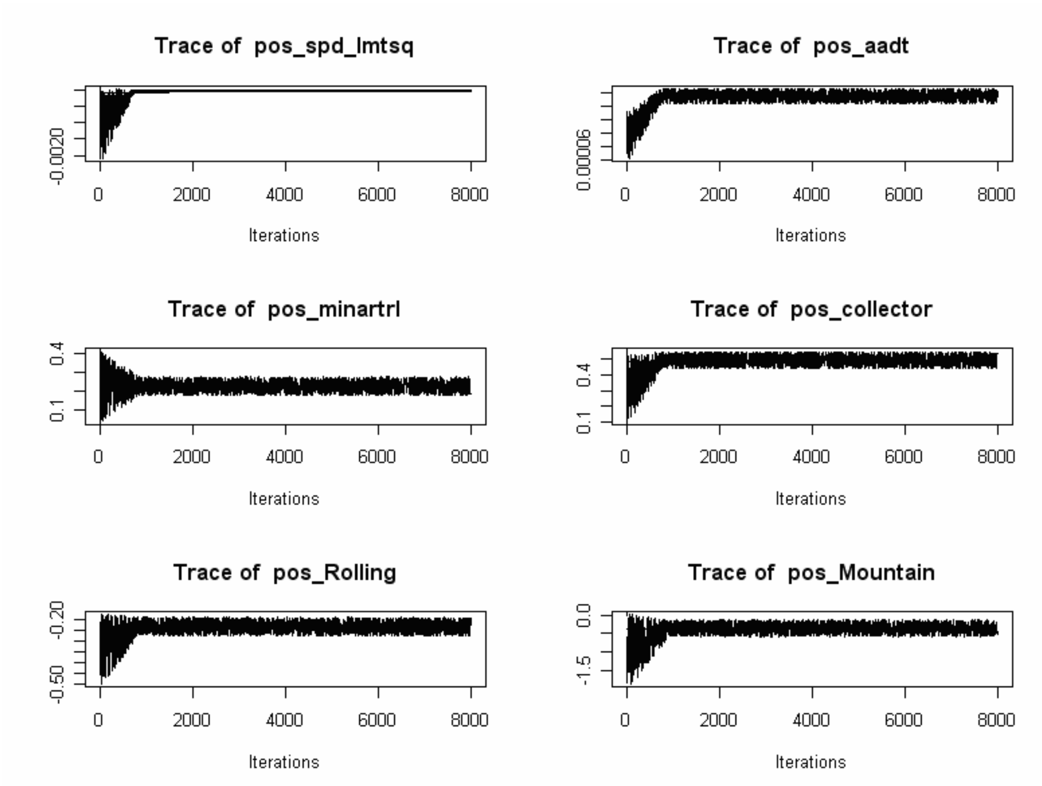


Figure C-2b Trace of Variables of Interest for Possible Injury Crash Frequency

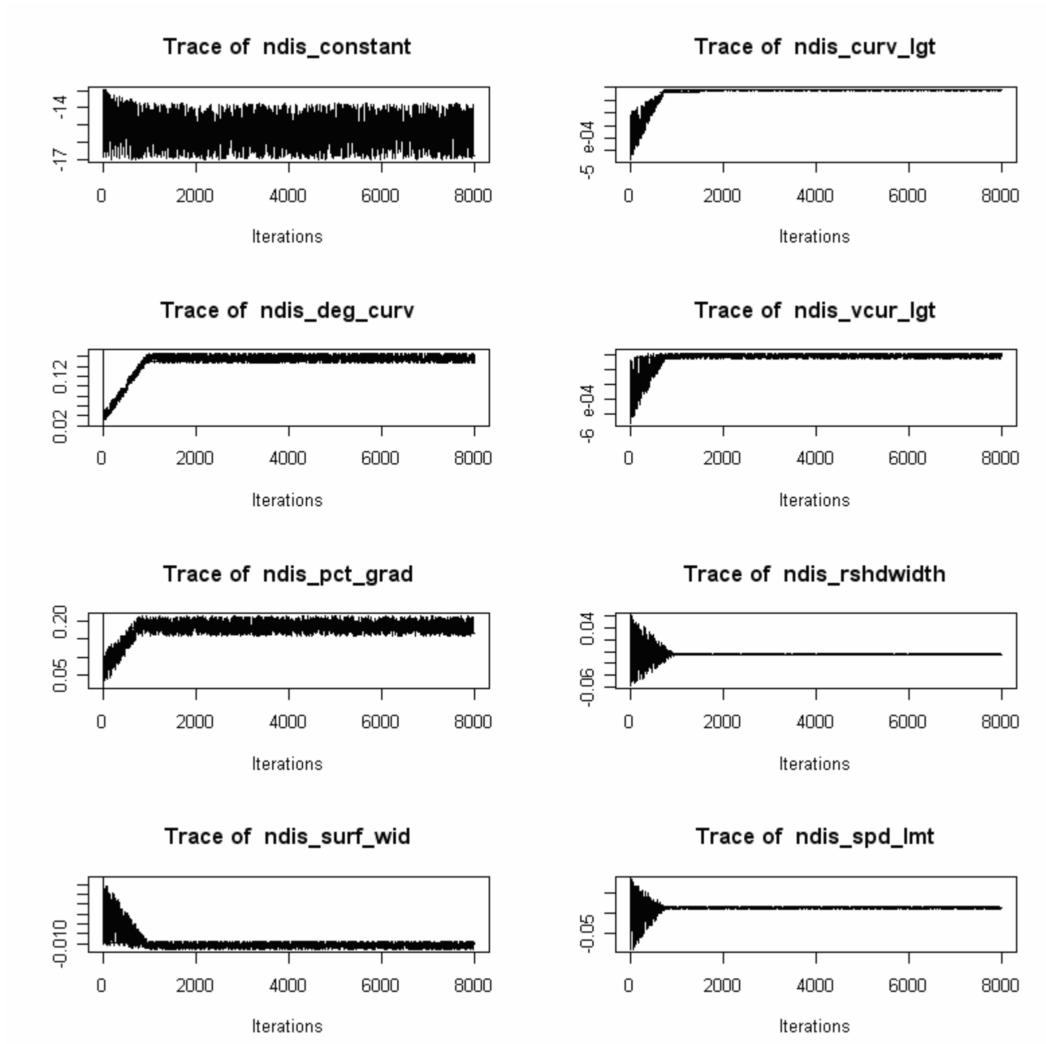


Figure C-3a Trace of Variables of Interest for Non-disabling Injury Crash Frequency

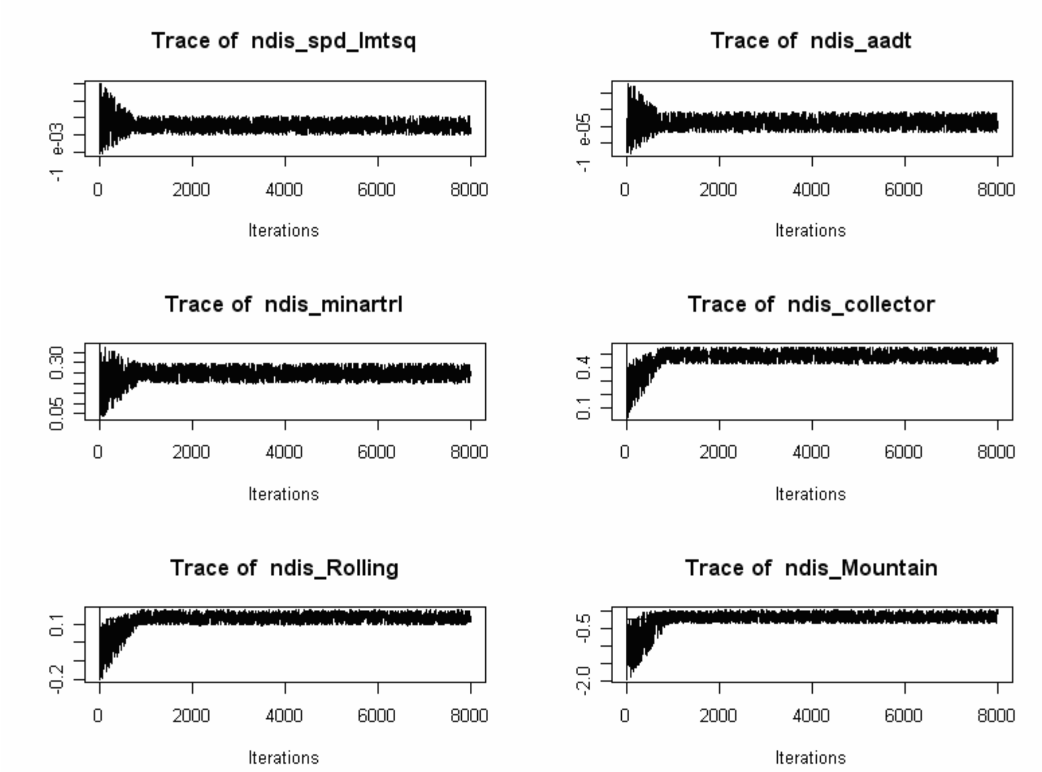


Figure C-3b Trace of Variables of Interest for Non-disabling Injury Crash Frequency

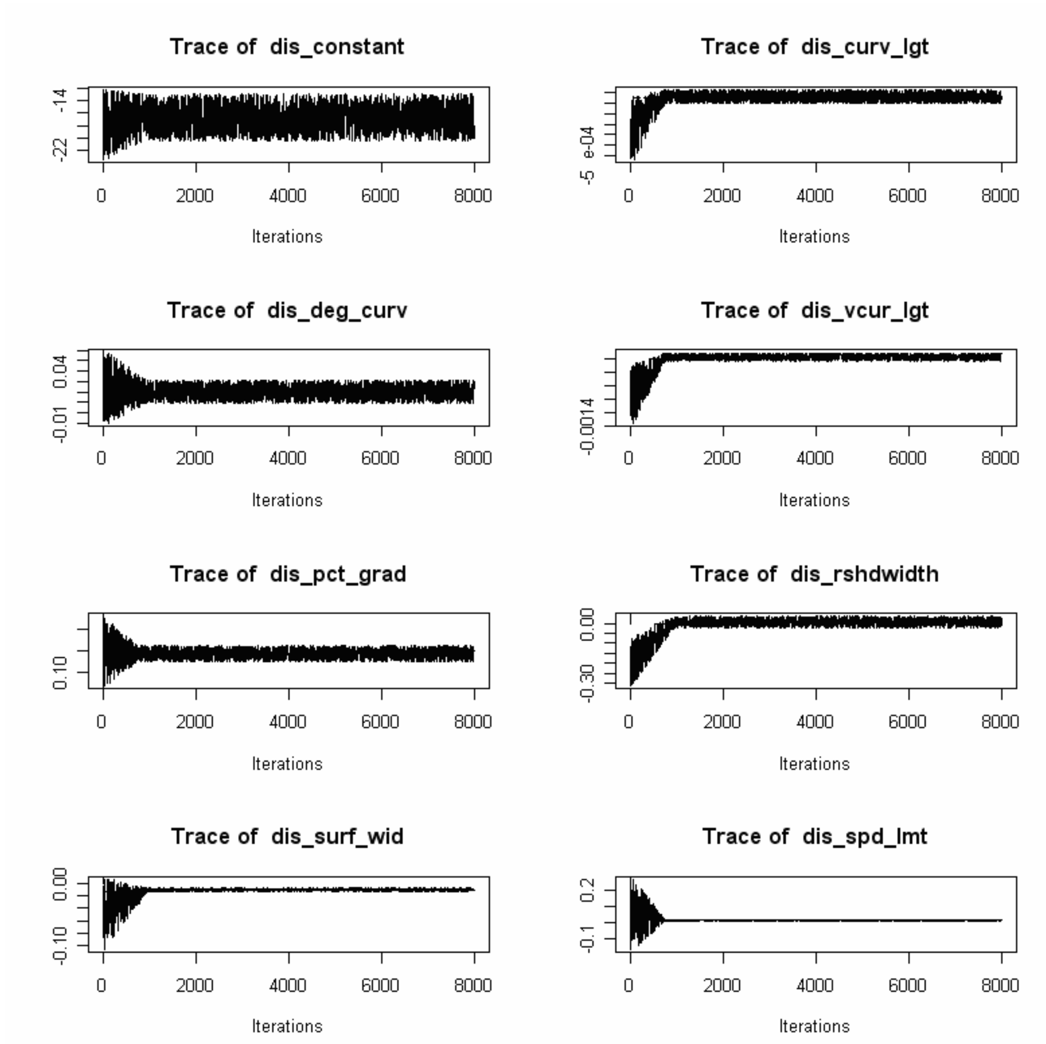


Figure C-4a Trace of Variables of Interest for Disabling Injury Crash Frequency

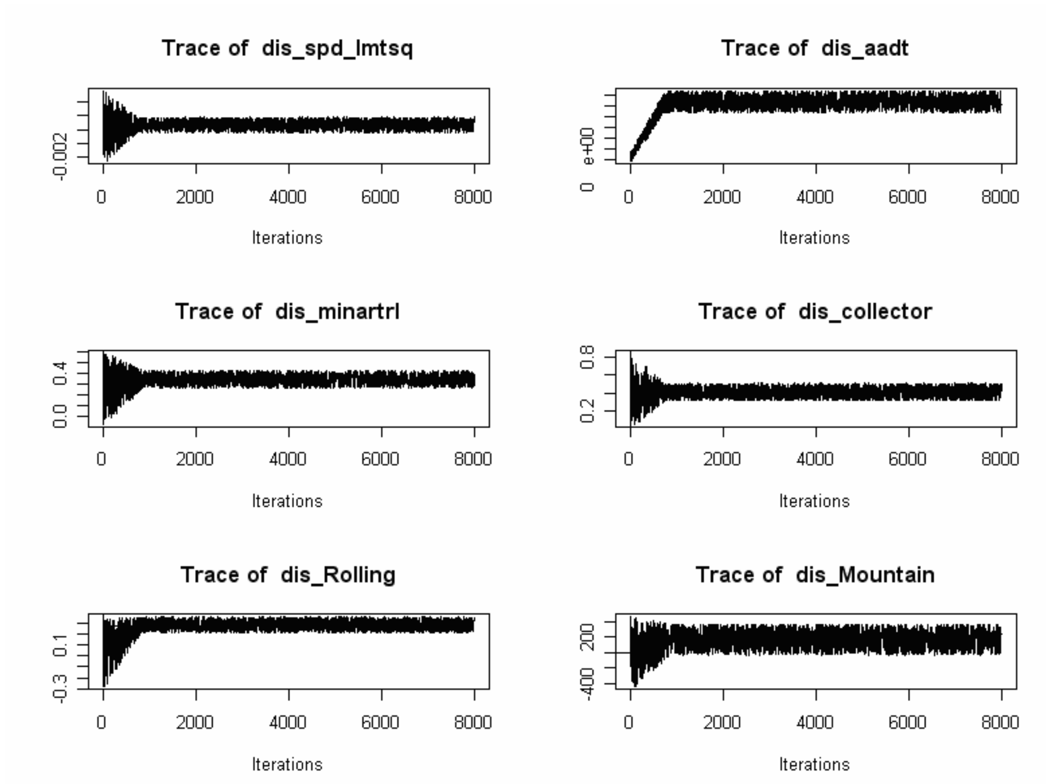


Figure C-4b Trace of Variables of Interest for Disabling Injury Crash Frequency

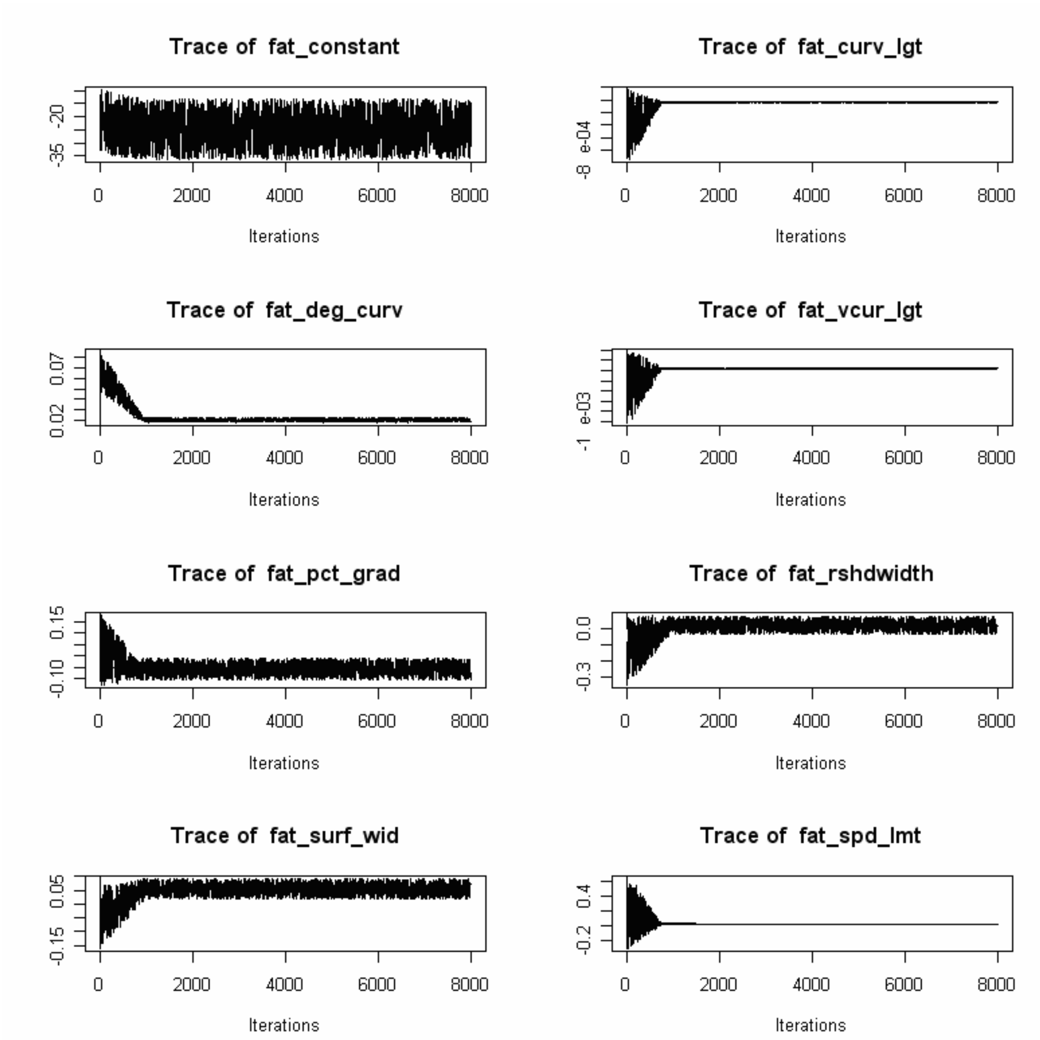


Figure C-5a Trace of Variables of Interest for Fatal Crash Frequency

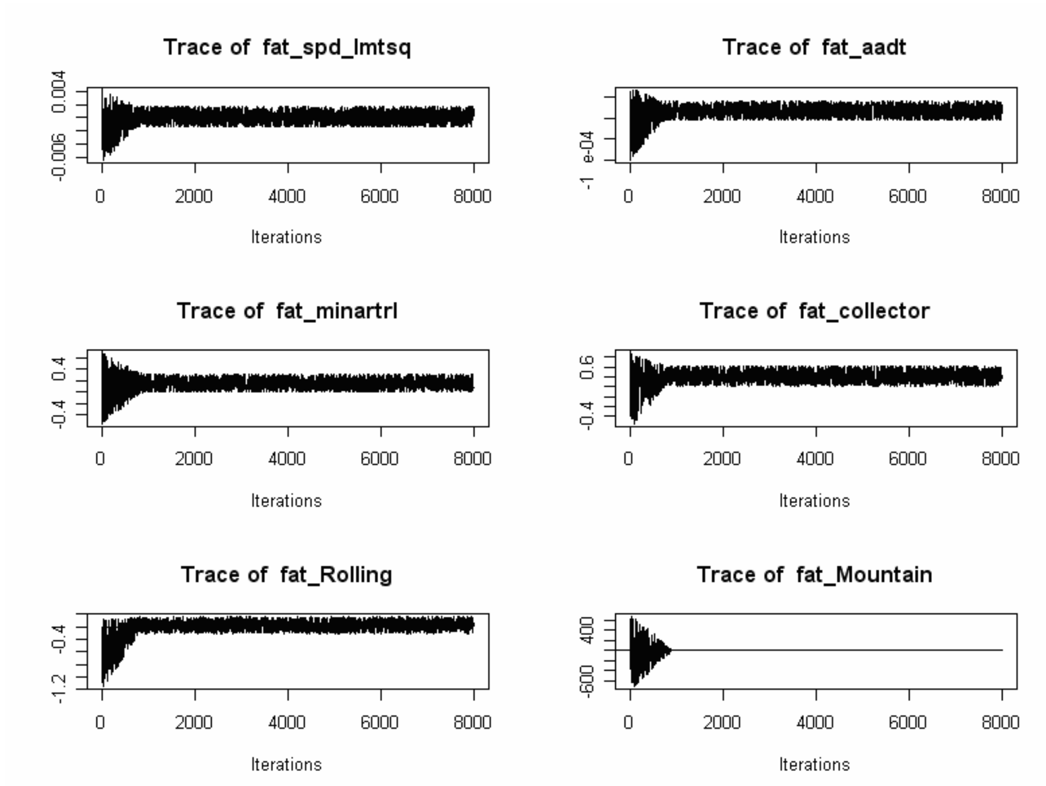


Figure C-5b Trace of Variables of Interest for Fatal Crash Frequency

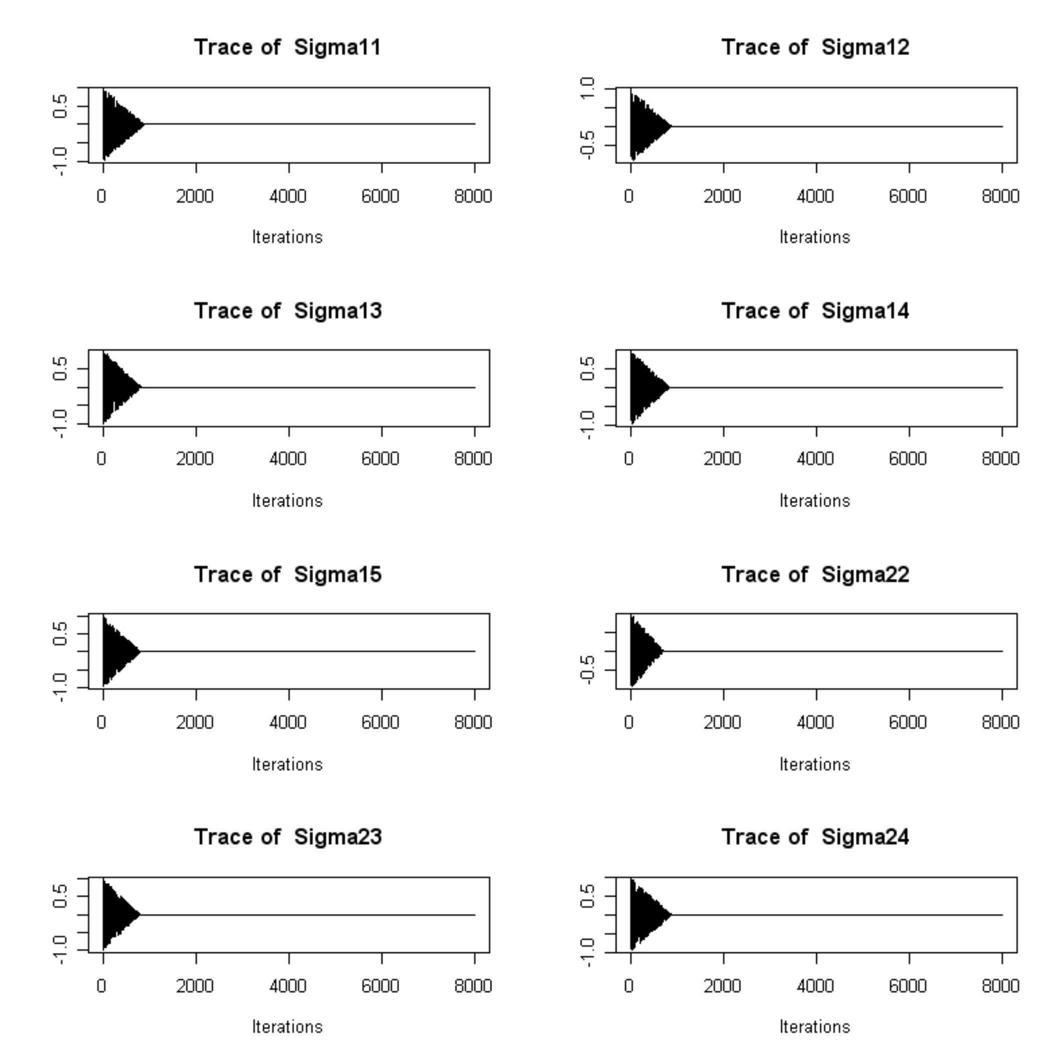


Figure C-6a Trace of Variance-Covariance of ε_{is}

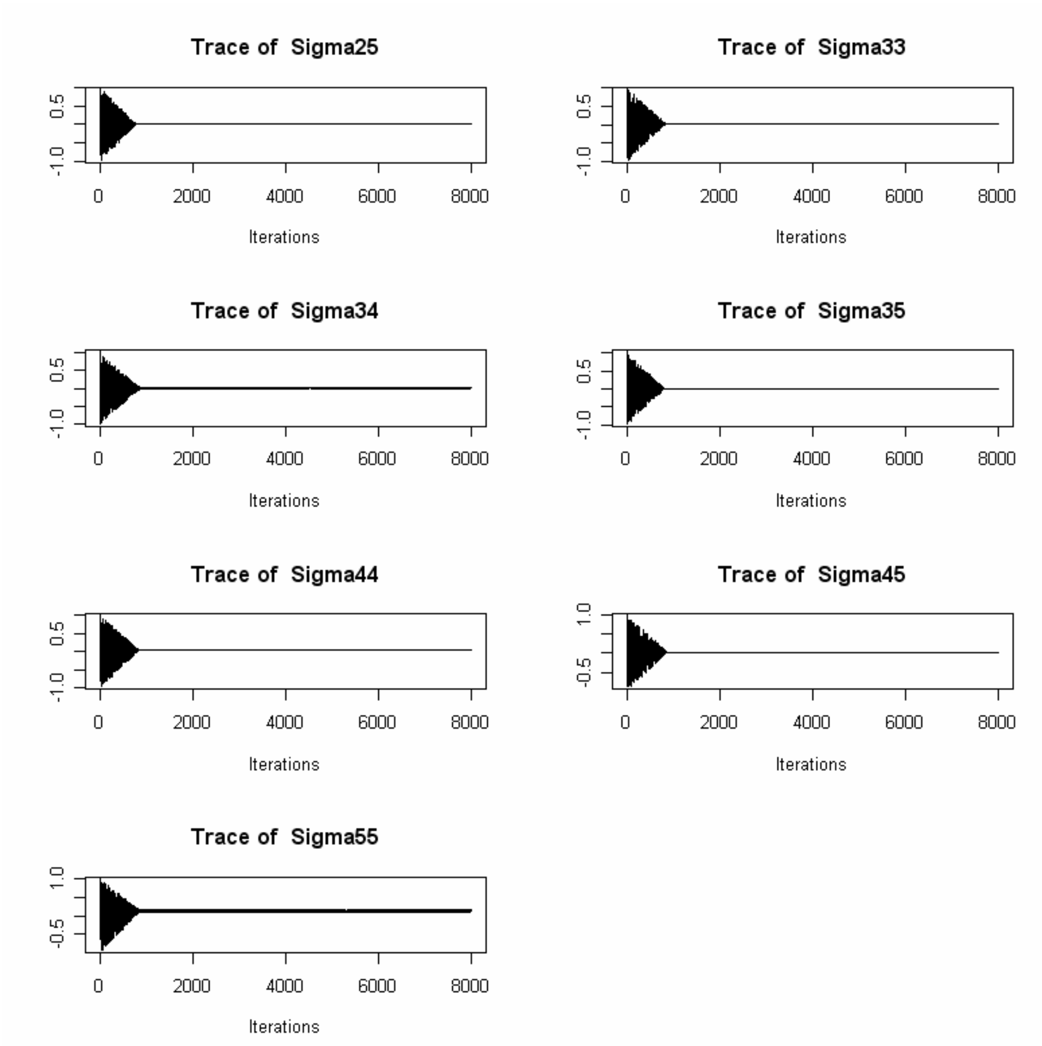


Figure C-6b Trace of Variance-Covariance of ε_{is}

APPENDIX D R CODE

```
inputdata <- function() {
	A <-
read.table(file="C:\\Jianming\\MVPLN\\Data\\wa02roadsegbased_crash_rural2lane_concisedPugetSound_60percentage.da
t",
		na.strings=".", col.names=c("fatal", "disablin", "nondisab", "possible", "noinjry", "seg_lng",
"curv_lgt",
		"deg_curv", "vcur_lgt", "pct_grad", "shldwid", "surf_wid", "spd_limt", "slsq", "aadt", "MinArtrl",
		"Collector", "Rolling", "Mountain", "vmt", "lnvmt"))
	return(A)
}



writedata <- function(x) {
	write.table(x, file = "C:\\Jianming\\MVPLN\\Results\\MVPLN.out", append = TRUE, quote = FALSE,
	sep = ",", eol = "\n", na = "NA", dec = ".", row.names = FALSE, col.names = FALSE, qmethod = "double")
}



NewtonRaphsonEpsi <-function(NRE_xi, NRE_yi, NRE_beta, NRE_epsilon, NRE_SIGMA) {

	eps <- 0.0001
	maxit <- 10

	nvar <- length(NRE_xi)-1

	invSIGMA <- solve(NRE_SIGMA)

	expxb_eps <- matrix(0,5,1)

	for(k in 1: maxit) {
		for(j in 1:5) {
			expxb_eps[j] <- exp(t(as.matrix(NRE_xi[1:nvar]))%*%as.matrix(NRE_beta[,j]) +
NRE_xi[nvar+1] + NRE_epsilon[j])
		}

		grad <- -invSIGMA%*%NRE_epsilon + (t(NRE_yi) - expxb_eps)

		Hess <- -invSIGMA - diag(expxb_eps)

		del <- solve(Hess, grad)
		if (sum(abs(del) < (eps*abs(NRE_epsilon)))==5) {
			return(list(NRE_epsilon=NRE_epsilon, V=solve(-Hess), ier=0))
		}
		else {
			NRE_epsilon <- NRE_epsilon - del
		}
	}
	return(list(NRE_epsilon=NRE_epsilon, V=solve(-Hess), ier=1))
}

NewtonRaphsonBetaj <-function(NRB_x, NRB_yj, NRB_betaj0, NRB_bjSIGMA0, NRB_betaj, NRB_epsilonj, NRB_df)
{
```

```
	eps <- 0.0001
	maxit <- 15
	n_obs <- nrow(NRB_yj)
	nvar <- length(NRB_betaj0)

	for(k in 1: maxit) {
		sumtmp <- matrix(0,nvar,1)
		for(i in 1:n_obs) {
			expxb_eps <- exp(t(as.matrix(NRB_x[i,1:nvar]))%*%NRB_betaj +
as.matrix(NRB_x[i,(nvar+1)]) + NRB_epsilonj[i])
			expxb_eps <- NRB_yj[i] - expxb_eps
			sumtmp <- sumtmp + expxb_eps[1]*as.matrix(NRB_x[i,1:nvar])
		}

		grad <- -solve(NRB_bjSIGMA0)%*%(NRB_betaj - NRB_betaj0) + sumtmp

		sumtmp <- matrix(0, length(NRB_betaj0),length(NRB_betaj0))

		for(i in 1:n_obs) {
			expxb_eps <- exp(t(as.matrix(NRB_x[i,1:nvar]))%*%NRB_betaj +
as.matrix(NRB_x[i,(nvar+1)]) + NRB_epsilonj[i])
			sumtmp <- sumtmp +
expxb_eps[1]*as.matrix(NRB_x[i,1:nvar])%*%t(as.matrix(NRB_x[i,1:nvar]))
		}

		Hess <- -solve(NRB_bjSIGMA0) - sumtmp

		del <- solve(Hess, grad)

		if (sum(abs(del) < (eps*abs(NRB_betaj)))==nvar) {
			return(list(NRB_betaj=NRB_betaj, V=solve(-Hess), ier=0))
		}
		else {
			NRB_betaj <- NRB_betaj - del
		}
	}
	return(list(NRB_betaj=NRB_betaj, V=solve(-Hess), ier=1))
}
```

SampleEpsiloni <- function(SE_xi, SE_yi, SE_beta, SE_epsilon, SE_SIGMA, SE_df) {

```
	mu_Sigma <- NewtonRaphsonEpsi(SE_xi, SE_yi, SE_beta, t(SE_epsilon), SE_SIGMA)
	nvar <- length(SE_xi)-1

	mu <- mu_Sigma$NRE_epsilon
	Sigma <- mu_Sigma$V

	Sigma <- Sigma*SE_df/(SE_df-2) #SE_df>2

	epsilon_star <- rmt(1,mu,Sigma,SE_df)
	zeros <- matrix(0, 1,5)

	tmplik <- 0
	for (j in 1:5) {
		tmplik <- tmplik -exp(SE_xi[1:nvar]%*%as.matrix(SE_beta[,j]) + SE_xi[nvar+1] + SE_epsilon[j])
+ SE_yi[j]*(SE_xi[1:nvar]%*%as.matrix(SE_beta[,j]) + SE_xi[nvar+1] + SE_epsilon[j])
	}

	post_old <- log(dmvnorm(SE_epsilon, zeros, SE_SIGMA)) + tmplik
```

```
	tmplik <- 0
	for (j in 1:5) {
		tmplik <- tmplik -exp(SE_xi[1:nvar]%*%as.matrix(SE_beta[,j])+ SE_xi[nvar+1] + epsilon_star[j])
+ SE_yi[j]*(SE_xi[1:nvar]%*%as.matrix(SE_beta[,j])+ SE_xi[nvar+1] + epsilon_star[j])
	}
	post_new <- log(dmvnorm(epsilon_star, zeros, SE_SIGMA)) + tmplik

	prop_old <- log(dmt(SE_epsilon, mu, Sigma, SE_df))

	prop_new <- log(dmt(epsilon_star, mu, Sigma, SE_df))
	tmp <- post_new+prop_old -(post_old+prop_new)
	logalpha <- min(tmp,0)
	u <- log(runif(1,0,1))
	if (u < logalpha) {
		return(epsilon_star)
	}
	else {
		return(SE_epsilon)
	}
}
```

SampleBetaj <- function(SB_x, SB_yj, SB_betaj0, SB_bjSIGMA0, SB_betaj, SB_epsilonj, SB_df) {

```
	mu_Sigma <- NewtonRaphsonBetaj(SB_x, SB_yj, SB_betaj0, SB_bjSIGMA0, SB_betaj, SB_epsilonj, SB_df)
	mu <- mu_Sigma$NRB_betaj
	Sigma <- mu_Sigma$V

	nvar <- length(SB_betaj0)

	Sigma <- Sigma*SB_df/(SB_df-2) #SB_df>2
	beta_star <- rmt(1,mu,Sigma,SB_df)
	beta_star <- t(beta_star)

	tmplik <- 0
	for(i in 1:nrow(SB_yj)) {

		tmplik <- tmplik -exp(t(as.matrix(SB_x[i,1:nvar]))%*%SB_betaj + SB_x[i,(nvar+1)] +
SB_epsilonj[i]) + SB_yj[i]*(t(as.matrix(SB_x[i,1:nvar]))%*%SB_betaj + SB_x[i,(nvar+1)] + SB_epsilonj[i])
	}

	post_old <- log(dmvnorm(t(SB_betaj), t(SB_betaj0), SB_bjSIGMA0)) + tmplik

	tmplik <- 0
	for(i in 1:nrow(SB_yj)) {
		tmplik <- tmplik -exp(t(as.matrix(SB_x[i,1:nvar]))%*%beta_star + SB_x[i,(nvar+1)] +
SB_epsilonj[i]) + SB_yj[i]*(t(as.matrix(SB_x[i,1:nvar]))%*%beta_star + SB_x[i,(nvar+1)] + SB_epsilonj[i])
	}

	post_new <- log(dmvnorm(t(beta_star), t(SB_betaj0), SB_bjSIGMA0)) + tmplik

	prop_old <- log(dmt(t(SB_betaj), mu, Sigma, SB_df))
	prop_new <- log(dmt(t(beta_star), mu, Sigma, SB_df))

	tmp <- post_new+prop_old -(post_old+prop_new)
	logalpha <- min(tmp,0)
	u <- log(runif(1,0,1))
	if (u < logalpha) {
		return(list(SB_betaj=beta_star, accept=1))
	}
	else {
```

```
            return(list(SB_betaj=SB_betaj, accept=0))
        }
}
```

SampleSIGMA <- function(n_obs, SS_epsilon, SS_R0, SS_df0) {

```
        invR0 <- solve(SS_R0)

        sumeps <- matrix(0, 5,5)
        for(i in 1: n_obs) {
            sumeps <- sumeps + as.matrix(SS_epsilon[i,])%*%t(as.matrix(SS_epsilon[i,]))
        }
        nu <- n_obs + SS_df0
        V <- solve(invR0 + sumeps)

        SS_Sigma <- rwishart(nu,V)$IW
        return(SS_Sigma)
}
```

mainfxn <- function(nn=10000,burnin=1000) {

```
        library(mnormt)
        library(bayesm)
        library(mvtnorm)

        numSevrty <- 5
        numVars <- 14

        A <- inputdata()
        n_obs <- nrow(A)

        Y <- as.matrix(A[,1:numSevrty])
        ones <- matrix(1,n_obs,1)

        df <- 10
        beta_df <- matrix(0,numSevrty,1)
        beta_df[1,1] <- 5
        beta_df[2,1] <- 6
        beta_df[3,1] <- 7
        beta_df[4,1] <- 8
        beta_df[5,1] <- 9

        X <- as.matrix(A[,c((numSevrty+2):19, 21)])
        X <- matrix(c(ones,X), nrow(X), ncol(X) +1)

        beta_i00 <- matrix(c(-21.86518, -0.0004342, 0.0671409, -0.0003211, 0.0303661, -0.1414917,
            -0.053037, 0.1878789, -0.0012703, -0.0000142, 0.0704688, 0.1255906, -0.6375601, -13.23141,
            -17.9396, -0.0002722, 0.0250824, -0.0009782, 0.1471798, -0.2000386, -0.0507062, 0.0499684,
            0.0001516, 0.0000188, 0.2597195, 0.4217457, 0.0180495, -12.27068,
            -14.95219, -0.000389, 0.0352405, -0.0004357, 0.0633192, 0.0007201, 0.0059334, -0.0062552,
            -0.0000187, 0.00000524, 0.2019583, 0.2333786, -0.0513502, -1.072978,
            -21.08788, -0.0000631, 0.0272553, -0.001352, 0.1484341, 0.0494227, 0.0750188, 0.1301605,
            -0.001124, 0.0000765, 0.2269426, 0.3033065, -0.3461112, -0.9991414,
            -18.44129, -0.0000865, 0.0296366, -0.0003852, -0.0289177, -0.0211329, 0.0461202, 0.1596419,
            -0.0018342, 0.0000237, -0.1142281, 0.0514133, 0.0474628, 0.4307832), 1, numVars*numSevrty)

        for (i in 1:(numVars*numSevrty)) {
            beta_i00[i] <- rnorm(1, beta_i00[i], abs(beta_i00[i])*0.5)
        }
        beta_i0 <- matrix(beta_i00, numVars,numSevrty)
```

```
	for (i in 1: numSevrty) {
		writedata(c("Initial value for betas'", i))
		writedata(beta_i0[,i])
	}
	beta_smpled <- beta_i0

	beta0 <- matrix(0,numVars,1)

	bSIGMA0 <- 100*diag(numVars)
	df_SIGMA <- 9

	R0 <- diag(numSevrty)

	eps_smpled <- matrix(0,n_obs,numSevrty)
	SIGMA_smpled  <- 1*diag(numSevrty)
	betaacceptflag <- matrix(0,1,numSevrty)

	writedata(date())
	for (ii in 1: nn) {
		for (i in 1: n_obs) {
			eps_smpled[i,]<-SampleEpsiloni(t(as.matrix(X[i,])), t(as.matrix(Y[i,])), beta_smpled,
t(as.matrix(eps_smpled[i,])), SIGMA_smpled, df)
		}

		SIGMA_smpled <- SampleSIGMA(n_obs, eps_smpled, R0, df_SIGMA)

		for(j in 1: 5)		{
			betatmp <- SampleBetaj(X, as.matrix(Y[,j]), as.matrix(beta0), bSIGMA0,
as.matrix(beta_smpled[,j]), as.matrix(eps_smpled[,j]), beta_df[j,1])
			beta_smpled[,j] <- betatmp$SB_betaj
			betaacceptflag[j] <- betatmp$accept
		}

		lentmp <- nrow(SIGMA_smpled)*ncol(SIGMA_smpled) + nrow(beta_smpled)*ncol(beta_smpled)
+ 5
		tmpresults <- matrix(c(SIGMA_smpled,beta_smpled,betaacceptflag),1,lentmp)
		writedata(tmpresults)
		print(c(ii, " iterations have been finished..."),quote=FALSE)
	}
	writedata(date())
}
```

REFERENCES

Abdel-Aty, M.A. (1998). Modeling incident-related routing decisions by using a nested logit structure. *Transportation Research Record 1645*, 103-110.

Abdel-Aty, M. A., Chen, C. and Schott, J. R. (1998). An assessment of the effect of driver age on traffic accident involvement using log-linear models. *Accident Analysis and Prevention 30*(6), 851–861.

Abdel-Aty, M.A., and Radwan, A.E. (1999). Developing crash predictive models for a principal arterial. *Traffic Safety on Two Continents*, 177-194, Malmo, Sweden.

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

Abdel-Aty, M.A., and Pemmanaboina, R. (2005). Assessing crash occurrence on urban freeways using static and dynamic factors. *Advances in Transportation Studies: An International Journal Section A & B 5*, 39-57.

Agent, K.R., and Pigman, J.G. (2001). *Safety impacts of rural road construction*. University of Kentucky, Kentucky Transportation Cabinet, Federal Highway Administration, Report No.: KTC-01-01/SPR219-00-11.

Agent, K.R., Pigman, J.G., and Stamatiadis, N., (2001). *Countermeasures for fatal crashes on two-lane rural roads*. University of Kentucky, Kentucky Transportation Cabinet, Federal Highway Administration, Report No.: KTC-01-11.

Aitchison, J., and Ho, C.H. (1989). The multivariate Poisson-log normal distribution. *Biometrika 76*(4), 643-653.

Albert, J.H., and Pepple, P.A. (1989). A Bayesian approach to some overdispersion models. *The Canadian Journal of Statistics 17*(3), 333-344.

Alfò, M., and Trovato, G. (2004). Semiparametric mixture models for multivariate count data. *The Econometrics Journal 7*(2), 426-454.

Arbous, A.G., and Kerrich, J.E. (1951). Accident statistics and the concept of accident proneness. *Biometrics 7*, 340-431.

Baker, S.G. (1994). The multivariate-Poisson transformation. *The Statistician 43*(4), 495-504.

Banihashemi, M., and Dimaiuta, M. (2005). Maximizing safety improvement benefits in crash prediction models with accident modification factors. *Transportation Research Record 1908,* 9-18.

Baum, H. M., Lund, A. K., and Wells, J. K. (1989). The mortality consequences of raising the speed limit to 65 mph on rural interstates. *American Journal of Public Health 79*(10), 1392-1395.

Baum, H. M., Wells, J. K., and Lund, A. K. (1991). The fatality consequences of the 65 mph speed limits, 1989. *Journal of Safety Research 22*, 171-177.

Bédard, M., Guyatt, G.H., Stones, M.J., and Hirdes, J.P. (2002). The independent contribution of driver, crash and vehicle characteristics to driver fatalities. *Accident Analysis and Prevention 34*(6), 717-727.

Besag, J., and Green, P. (1993). Spatial statistics and Bayesian computation (with discussion). *Journal of the Royal Statistical Society B 55*(1), 25-37.

Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science 10*(1), 3-41.

Bijleveld, F.D. (2005). The covariance between the number of accidents and the number of victims in multivariate analysis of accident related outcomes. *Accident Analysis and Prevention 37*(4), 591-600.

Blincoe, L.J., Seay, A.G., Zaloshnja, E., Miller, T.R., Romano, E.O., Luchter, S., and Spicer, R.S. (2002). *The Economic Impact of Motor Vehicle Crashes, 2000*. NHTSA Technical Report, Report#: HS-809 446.

Booth, J.G., and Hobert, J.P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61*(1), 265-285.

Box, G.E.P., Jenkins, G.M., and Reinsel, G.C. (1994). *Time Series Analysis: Forecasting and Control, Third edition,* Prentice Hall, Upper Saddle River, NJ.

Brown, D.B., Maghsoodloo, S., and McArdle, M.E. (1990). The safety impact of the 65 mph speed limit: A case study using Alabama accident records. *Journal of Safety Research 21*, 125-139.

Bureau of Transportation Statistic (BTS) (2005). *Pocket Guide to Transportation 2005*. Retrieved June 9, 2005, from BTS – Pocket Guide to Transportation – 2005 Web site:

http://www.bts.gov/publications/pocket_guide_to_transportation/2005/pdf/entire.pdf.

Bureau of Transportation Statistics (BTS) (2001). *The Intermodal Transportation Database – Annual Vehicle Miles Traveled*. Retrieved July 31, 2003, http://transtats.bts.gov/.

Burritt, B. E. (1976). Analysis of the relation of accidents and the 88-km/h (55-mph) speed limit on Arizona highways. *Transportation Research Record 609*, 34-35.

Cameron, A.C., and Trivedi, P.K. (1998). *Regression Analysis of Count Data*. Cambridge University Press, New York, USA.

Carlson, P.J., Miles, J.D., and Johnson, P.K. (2005). Findings and implications from daytime high-speed passing maneuvers observed on a rural two-lane, two-way highway. *3rd International Symposium on Highway Geometric Design*. Chicago, Illinois, US.

Chan, K.S., and Ledolter, J. (1995). Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association 90*(429), 242-252.

Chang, G.L., and Paniati, J. F. (1990). Effects of 65-mph speed limit on traffic safety. *Journal of Transportation Engineering 116*(2), 213-226.

Chen, C.S., and Cottrell, B.H. (2005). *Guidelines for Using Centerline Rumble Strips in Virginia*. FHWA/VTRC 05-R30. Virginia Transportation Research Council.

Chib, S., and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician 49*(4), 327-335.

Chib, S., Greenberg, E., and Winkelmann, R. (1998). Posterior simulation and Bayes factors in panel count data models. *Journal of Econometrics 86*(1), 33-54.

Chib, S., and Winkelmann, R. (2001) Markov Chain Monte Carlo Analysis of Correlated Count Data. *Journal of Business & Economic Statistics 19*(4), 428-435.

Chin, H.C.C., and Quddus, M.A. (2003). Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. *Accident Analysis and Prevention 35*(2), 253-259.

Chipman, M.L., Macdonald, S., and Mann, R.E. (2003). Being “at fault” in traffic crashes: does alcohol, cannabis, cocaine, or polydrug abuse make a difference? *Injury Prevention 9*, 343-348.

Christiansen, C.L., Morris, C.N., and Pendleton, O.J. (1992). *A Hierarchical Poisson Model with Beta Adjustments for Traffic Accident Analysis* (Technical Report 103). Austin, TX: Center for Statistical Sciences, The University of Texas at Austin.

Chu, X. (1999). Effects of changes in travel patterns on highway fatalities. *Accident Analysis and Prevention 31*(3), 221-227.

Clarke, D.B., and Sarasua, W. A., 2003. *Fatal crashes on rural secondary highways*. Clemson University, South Carolina Department of Transportation, Federal Highway Administration. Report No.: FHWA-SC-03-04.

Council, F.M., and Stewart, J.R., 2000. Safety effects of the conversion of rural two-lane roadways to four-lane roadways. *ITE Journal*, 70(4), 37-42.

Cowles, M.K., and Carlin, B.P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association 91*(434), 883-904.

Cragg, J.G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica, 39*(5), 829-844.

Damien, P., Wakefield, J., and Walker S. (1999). Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables. *Journal of Royal Statistical Society B 61*(2), 331-344.

Damien, P., and Walker, S.G. (2001). Sampling truncated normal, beta, and gamma densities. *Journal of Computational and Graphical Statistics 10*(2), 206-215.

Damien, P. (2005). Some Bayesian nonparametric models. *Handbook of Statistics 25*, 279-314.

Daniel, J., and Chien, S.I.J. (2004). Truck safety factors on urban arterials. *Journal of Transportation Engineering 130*(6), 742-752.

Daniel, J., Tsai, C., and Chien, S. (2002). Factors in truck crashes on roadways with intersections. *Transportation Research Record 1818*, 54-59.

Dart, O. K. (1977). Effects of the 88.5-km/h (55-mph) speed limit and its enforcement on traffic speeds and accidents. *Transportation Research Record 643*, 23-32.

Davis, G.A. (2000). Accident reduction factors and causal inference in traffic safety studies: A review. *Accident Analysis and Prevention, 32*(1), 95-109.

Djauhari, M.A. (2002). Stochastic pattern of traffic accidents in Bandung city. *International Association of Traffic & Safety Sciences (IATSS) Research 26*(2), 85-91.

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

Edwards, R.G., and Sokal, A.D. (1988). Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm. *Physical Review D 38*(6), 2009-2012.

Evans, M., and Swartz, T. (1995). Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems. *Statistical Science 10*, 254-272.

Farmer, C.M., Lund, A.K., Trempel, R.E., and Braver, E.R. (1997). Fatal crashes of passenger vehicles before and after adding antilock braking systems. *Accident Analysis and Prevention, 29*(6), 745-757.

Fitzpatrick, K., Parham, A.H., Brewer, M.A., and Miaou, S-P, 2001. *Characteristics of and Potential Treatments of Crashes on Low-Volume, Rural Two-Lane Highways in Texas*. TxDOT Research Report 4048-1.

Fitzpatrick, K., Parham, A.H., and Brewer, M.A., 2002. *Treatments for Crashes on Rural Two-Lane Highways in Texas*. TxDOT Research Report 4048-2.

Fitzpatrick, K., and Brewer, M.A., 2002. *Additional Characteristics of Crashes on Rural Two-Lane Highways*. TxDOT Research Report 4048-3.

Fitzpatrick, K. (2004). *Handbooks for improving safety on two-lane rural highways: what are we saying?* Institute of Transportation Engineers 2004 Annual Meeting and Exhibit, Lake Buena Vista, Florida, US.

Fowles, R., and Loeb, P.D. (1989). Speeding, coordination and the 55 mph limit: Comment. *The American Economic Review 79*(4), 916-921.

Fridstrøm, L., Ifver, J., Ingebrigtsen, S., Kulmala, R., and Thomsen, L.K. (1995). Measuring the contribution of randomness, exposure, weather, and daylight to the various in road accident counts. *Accident Analysis and Prevention 27*(1), 1-20.

Freedman, M., and Esterlitz, J. R. (1990). Effect of the 65 mph Speed Limit on Speeds in Three States. *Transportation Research Record 1281*,52-61.

Furness, S., Connor, J., Robinson, E., Norton, R., Ameratunga, S., and Jackson, R. (2003). Car colour and risk of car crash injury: Population based case control study. *British Medical Journal 327*, 1455-1456.

Garbacz, C. (1990). How effective is automobile safety regulation? *Applied Economics 22*(12), 1705-1714.

Garber, N.J., and Ehrhart A.A. (2000). Effects of speed, flow, and geometric characteristics on crash frequency for two-lane highways. *Transportation Research Record 1717*, 76–83.

Garber, N.J., and Wu, L. (2001). *Stochastic Models Relating Crash Probabilities with Geometric and Corresponding Traffic Characteristics Data* (UVACTS-5-15-74). Charlottesville, VA: Center for Transportation Studies, University of Virginia.

Gattis, J.L., Bhave, R.A., and Duncan, L.K. (2003). *Incremental Improvements to Rural Two-Lane Highways*. MBTC 1098, Final Report. University of Arkansas, Fayetteville, Mack-Blackwell Transportation Center. US.

Gelfand, A.E., and Smith, F.M. (1990), Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association 85*(410), 398-409.

Gelfand, A.E., Hills, S.E., Racine-Poon, A., and Smith, A.F.M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association 85*(412), 972-985.

Gelfand, A.E., Sahu, S.K., and Carlin, B.P. (1995). Efficient parametrisations for normal linear mixed models. *Biometrika 82*(3), 479-488.

Gelman A., and Rubin D.B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science 7*(4), 457-511.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (2004). Bayesian *Data Analysis 2nd Edition*. Chapman & Hall/CRC, Boca Raton, Florida, USA.

Geman, S., and Geman D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence 6*, 721-741.

Geweke, J. (1991). Efficient simulation from the multivariate normal and student t-distributions subject to linear constraints. *Computer Sciences and Statistics Proceedings of the 23rd Symposium on the Interface*, 571-578.

Gilks, W.R., Richardson, S., and Speigelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.

Golob, T.F., and Recker, W.W., 1987. Analysis of truck-involved freeway accidents using log-linear modeling. *Journal of Safety Research 18*(3), 121-136.

Golob, T.F., and Recker, W.W. (2002). *Relationships among Urban Freeway Accidents, Traffic Flow, Weather and Lighting Conditions.* Institute of Transportation Studies Report UCI-ITS-WP-02-1, University of California, Irvine: CA.

Greene, W.H. (2003). *Econometric Analysis, Fifth Edition*. Prentice Hall, New Jersey.

Gurmu, S., Rilstone, P., and Stern, S. (1999). Semiparametric estimation of count regression models. *Journal of Econometrics 88*(1), 123-150.

Hallmark, S.L., and Mueller, K. (2004). *Impact of Left-turn Phasing on Older and Younger Drivers at High-speed Signalized Intersections*. CTRE Project 03-149 Final Report. Center for Transportation Research and Education, Iowa State University, Ames, IA, USA.

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika 57*(1), 97-109.

Hauer, E. (1986). On the estimation of the expected number of accidents. *Accident Analysis and Prevention 18*(1), 1-12.

Hauer, E. (1997). *Observational Before-After Studies in Road Safety*. Pergamon, Oxford.

Hauer, E. (2001). Overdispersion in modeling accidents on road sections and in empirical Bayes estimation. *Accident Analysis and Prevention 33*(6), 799-808.

Heibron, D.C. (1994). Zero-altered and other regression models for count data with added zeros. *Biometrical Journal 36*(5), 531-547.

Higdon, D. (1998). Auxiliary variable methods for Markov chain Monte Carlo with applications. *Journal of American Statistical Association 93*(442), 585-595.

Houston, D. J. (1999). Implications of the 65 mph speed limits for traffic safety. *Evaluation Review 23*(3), 304-315.

Huang, H. F., Schneider, R.J., Zegeer, C.V., Khattak, A.J., Amerlynck, V.J., and Lacy, J.K., (2001). *Identification of severe crash factors and countermeasures in North Carolina*. University of North Carolina, North Carolina Department of Transportation, *Federal Highway Administration. Report No.: FHWA/NC/2001-003.*

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

Ivan, J.N., and Wang, C., (2000). Explaining two-lane highway crash rates using land use and hourly exposure. *Accident Analysis and Prevention* 32(6), 487-795.

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

Jaarsma, C.F., Botma, H., and Beunen, R. (2005). Passing bays for slow moving vehicles on rural two-lane roads. *Transportation Reviews 25*(4), 491-509.

Johansson, P. (1996). Speed limitation and motorway casualties: A time-series count data regression approach. *Accident Analysis and Prevention 28*(1), 73-87.

Johnson, N.L., Kotz, S., and Balakrishnan, N. (1997). *Discrete Multivariate Distributions*. John Wiley & Sons, Inc. New York, NY.

Karlis, D. (2003). An EM algorithm for multivariate Poisson distribution and related models. *Journal of Applied Statistics 30*(1), 63-77.

Khattak, A.J., and Knapp, K.K. (2001). Snow event effects on interstate highway crashes. *Journal of Cold Regions Engineering 15*(4), 219-229.

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

Khattak, A.J., Naik, B., and Kannan, V. (2004). *Safety Evaluation of Left-turn Lane Line Width at Intersections with Opposing Left-turn Lanes*. University of Nebraska, Lincoln, Mid-America Transportation Center.

Kim, D.G., and Washington, S. (2006). The significance of endogeneity problems in crash models: An examination of left-turn lanes in intersection crash models. *Proceedings of the Transportation Research Board 85th Annual Meeting*, January 2006, Washington D.C, USA.

Kindler, C., Wood, R.M., Harwood, D.W., Neuman, T.R., Potts, I.B., and Antonucci, N. (2003). Intersection diagnostic review module: expert system for geometric design review of intersections on rural two-lane highways. *Transportation Research Record 1851*, 3-12.

King, G. (1989). A seemingly unrelated Poisson regression model. *Sociological Methods and Research 17*(3), 235-255.

Kockelman, K. (2001). A model for time- and budget-constrained activity demand analysis. *Transportation Research 35B* (3), 255-269.

Kockelman, K., Heard, L., Kweon, Y.J., and Rioux, T. (2002). Sidewalk Cross-Slope Design: Analysis of Accessibility for Persons with Disabilities. *Transportation Research Record 1818*, 108-118.

Koceklman, K., and Bottom, J. (2006). *Safety Impacts and Other Implications of Raised Speed Limits on High-Speed Roads*. National Cooperative Highway Research Program Project #17-23, Report Draft, Transportation Research Board. Washington, D.C.

Koorey, G. (2005). The use of road geometry data for highway applications. Transportation Research Board 3rd International Symposium on Highway Geometric Design. Chicago, Illinois, USA.

Kumara, S.P., and Chin, H.C. (2003). Modeling accident occurrence at signalized T intersections with special emphasis on excess zeros. *Traffic Injury Prevention 4*(1), 53-57.

Kumara, S.P., and Chin, H.C. (2004). Study of fatal traffic accidents in Asia Pacific countries. *Transportation Research Record 1897*, 43-47.

Kumara, S.P., and Chin, H.C. (2005). Application of Poisson underreporting model to examine crash frequencies at signalized three-legged intersections. *Transportation Research Record 1908*, 46-53.

Kumara, S.P., Chin, H.C., and Weerakoon, WMSB. (2003). Identification of accident causal factors and prediction of hazardousness of intersection approaches. *Transportation Research Record 1840*, 116-122.

Kweon, Y.J., and Kockelman, K.M. (2003). Overall injury risk to different drivers: combining exposure, frequency, and severity models. *Accident Analysis and Prevention 35*(4), 441-450.

Kweon, Y.J., and Kockelman, K.M. (2004). Spatially disaggregate panel models of crash and injury counts: The effect of speed limits and design. *Proceedings of the Transportation Research Board 83rd Annual Meeting*, January 2004, Washington D.C, USA.

Kweon, Y.J. and Kockelman, K.M. (2005). The safety effects of speed limit changes: Use of panel models, including speed, use and design variables. *Transportation Research Record 1908*, 195-204.

Laberge-Nadeau, C., Dionne, G., Maag, U., Desjardins, D., Vanasse, C., and Ekoe, J.M. (1996). Medical conditions and the severity of commercial motor vehicle drivers' road accidents. *Accident Analysis and Prevention 28*(1), 43-51.

Lacy, J.K., 2002. *Southeast regional fatal study, a causal chain analysis in North Carolina*. University of North Carolina, North Carolina Department of Transportation, Federal Highway Administration. Report No.: FHWA/NC/2002-007.

Ladron de Guevara, F., and Washington, S. (2004). Forecasting crashes at the planning level. A simultaneous negative binomial crash model applied in Tucson, Arizona. *Transportation Research Record 1897*, 191-199.

Lamm, R., Choueiri, E.M., and Mailaender, T. (1991). Traffic safety on two-continents – A ten year analysis of human and vehicular involvements. *Proceedings of Strategic Highway Research Program (SHRP) and Traffic Safety on Two Continents*, Gothenburg, Sweden, September 18-20, 1991, Swedish Road and Traffic Research Institute, 121-136.

Lamm R., Psarianos, B., and Cafiso, S. (2002). Safety Evaluation Process for Two-Lane Rural Roads: A 10-Year Review. *Transportation Research Record 1796*, 51-59.

Land, K.C., McCall, P.L., and Nagin, D.S. (1996). A comparison of Poisson, negative binomial, semiparametric mixed Poisson regression models, with empirical applications to criminal careers data. *Sociological Methods & Research 24*(4), 387-442.

Lave, C., and Elias, P. (1994). Did the 65 mph speed limit save lives? *Accident Analysis and Prevention 26*(1), 49-62.

Lave, C., and Elias, P. (1997). Resource allocation in public policy: The effects of the 65-mph speed limit. *Economic Inquiry 35*, 614-620.

Lee, J., and Mannering, F.L. (1999). *Analysis of Roadside Accident Frequency and Severity and Roadside Safety Management* (WA-RD 475.1). Seattle, WA: Washington State Transportation Center, University of Washington.

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

Lee, P.M. (2004). *Bayesian Statistics: An Introduction*. New York, NY: Oxford University Press Inc.

Levy, D.T., and Asch, P. (1989). Speeding, coordination and the 55 MPH limit: Comment. *The American Economic Review 79*(4), 913-915.

Li, C.C., Lu, J.C., Park, J., Kim, K., Brinkley, P.A., and Peterson, J.P. (1999). Multivariate zero-inflated Poisson models and their applications. *Technometrics, 41*(1), 29-38.

Liu, J., Ravishanker, N., Ivan, J.N., and Qin, X. (2005). Hierarchical Bayesian estimation of safety performance functions for two-lane highways using Markov chain Monte Carlo modeling. *Journal of Transportation Engineering, 131*(5), 345-351.

Loeb, P. D. (1995). The effectiveness of seat-belt legislation in reducing injury rates in Texas. *American Economic Review Papers and Proceedings 85*(2), 81-84.

Loeb, P. D. (2001). The effectiveness of seat belt legislation in reducing driver-involved injury rates in Maryland. *Transportation Research Part E 37*(4), 297-310.

Lord, D., and Persaud, B.N. (2000). Accident prediction models with and without trend: application of the generalized estimating equations procedure. *Transportation Research Record 1717*, 102-108.

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

Ma, J., and Kockelman, K. (2006). Bayesian multivariate Poisson regression for models of injury count, by severity. *Transportation Research Record* (forthcoming).

Ma, J., and Kockelman, K. (2004). Anticipating injury and death: Controlling for new variables on southern California highways. *Proceedings of the Transportation Research Board 83rd Annual Meeting*, January 2004, Washington D.C, USA.

Mace, D. J., and Heckard, R. (1991). *Effect of the 65 mph Speed Limit on Travel Speeds and Related Crashes.* Publication DOT HS 807 764, NHTSA, U.S. DOT.

MacNab, Y.C. (2003). A Bayesian hierarchical model for accident and injury surveillance. *Accident Analysis and Prevention 35*(1), 91-102.

Mahamunulu, D.M. (1967). A note on regression in the multivariate Poisson distribution. *Journal of the American Statistical Association 62*(317), 251-258.

Mathworld (2006). Newton's Method. Retrieved February 24, 2006, http://mathworld.wolfram.com/NewtonsMethod.html.

Meliker, J.R., Maio, R.F., Zimmerman, M.A., Kim, Y.M., Smith, S.C., and Wilson, M.L. (2004). Spatial analysis of alcohol-related motor vehicle crash injuries in southeastern Michigan. *Accident Analysis and Prevention 36*(6), 1129-1135.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics 21*(6), 1087-1092.

Miaou, S.P., Hu, P.S., Wright, T., Rathi, A.K., and Davis, S.C. (1992). Relationship between truck accidents and highway geometric design: a Poisson regression approach. *Transportation Research Record 1376*, 10-18.

Miaou, S.P., and Lum, H. (1993). Modeling vehicle accidents and highway geometric design relationships. *Accident Analysis and Prevention 25*(6), 689-709.

Miaou, S.P. (1994). The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. *Accident Analysis and Prevention 26*(4), 471-482.

Miaou, S.P. (1996). *Measuring the Goodness-of-fit of Accident Prediction Models* (FHWA-RD-96-040). Oak Ridge, TN: Oak Ridge National Laboratory.

Miaou, S.P. (2001). *Estimating Roadside Encroachment Rates with the Combined Strengths of Accident- and Encroachment-Based Approaches* (FHWA-RD-01-124). Oak Ridge, TN: Oak Ridge National Laboratory.

Miaou, S.P., and Lord, D. (2003). Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and Bayes versus empirical Bayes. *Transportation Research Record 1840*, 31-40

Miaou, S.P., Song, J.J., and Mallick, B.K. (2003). Roadway traffic crash mapping: A space-time modeling approach. *Journal of Transportation and Statistics 6*(1), 33-57.

Miaou, S.P., and Song, J.J. (2005). Bayesian ranking of sites for engineering safety improvements: Decision parameter, treatability concept, statistical criterion, and spatial dependence. *Accident Analysis and Prevention 37*(4), 699-720.

Miaou, S.P., Bligh, R.P., and Lord, D. (2005). Developing guidelines for median barrier installation: benefit-cost analysis with Texas data. *Transportation Research Record 1904*, 3-19.

Miller, T.R., (1992). Benefit-cost analysis of lane marking. *Transportation Research Record 1334*, 38-45.

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

Mira, A., and Tierney, L. (1997). *On the Use of Auxiliary Variables in Markov Chain Monte Carlo Sampling*, Technical Report, University of Minnesota, School of Statistics.

Mrozek, J.R., and Taylor, L.O. (2002). What Determines the Value of Life? A Meta-Analysis. *Journal of Policy Analysis and Management 21*(2), 253-270.

Najjar, Y.M., Russell, E. R., Stokes, R.W., and Abu, Lebdeh, G. (2002). New speed limits on Kansas highways: impact on crashes and fatalities. *Journal of the Transportation Research Forum 56*(4), 119-147.

National Highway Traffic Safety Administration (NHTSA). (2004). *State and National Fatality and Injury Data*. Retrieved October 23, 2005, http://www.nhtsa.dot.gov/people/Crash/crashstatistics/index.htm

Nedelman, J., and Wallenius, T. (1986). Bernoulli trials, Poisson trials, surprising variances, and Jensen's inequality. *The American Statistician 40*(4), 286-289.

Newstead S.V., Cameron M.H., and Leggett L.M. (2001). The crash reduction effectiveness of a network-wide traffic police deployment system. *Accident Analysis and Prevention 33*(3), 393-406.

NHTSA. (1989). *Report to Congress on the Effects of the 65 mph Speed Limit During 1987*. U.S. DOT.

NHTSA. (1992). *Report to Congress on the Effects of the 65 mph Speed Limit Through 1990*. U.S. DOT.

NHTSA (2005a). Fatality Analysis Reporting System (FARS) Web-Based Encyclopedia. Retrieved January 29, 2006, http://www-fars.nhtsa.dot.gov/.

NHTSA (2005b). *Traffic Safety Facts: Research Notes*. National Highway Traffic Safety Administration, Report* DOT HS 809 831. January. Washington, D.C.

NHTSA (2005c). *Traffic Safety Facts 2003*. National Highway Traffic Safety Administration, Report* DOT HS 809 775. January. Washington, D.C.

NHTSA (2005d). Safety – Your Most Important Call. Retrieved January 30, 2006, http://www-nrd.nhtsa.dot.gov/PDF/ nrd-13/WheelerT_ppt.pdf

Noland, R.B. (2003). Traffic fatalities and injuries: the effect of changes in infrastructure and other trends. *Accident Analysis and Prevention 35*(4), pp. 599-611.

O'Donnell, C.J., and Connor, D.H. (1996). Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice. *Accident Analysis and Prevention 28*(6), 739-753.

Ogden, K.W. (1997). The effects of paved shoulders on accidents on rural highways. *Accident Analysis and Prevention 29*(3), 353-362.

Oh, J., Washington, S., and Choi, K. (2004). Development of accident prediction models for rural highway intersections. *Transportation Research Record 1897*, 18-27.

Ossenbruggen, P.J., Pendharkar, J., and Ivan, J. (2001). Roadway safety in rural and small urbanized areas. *Accident Analysis and Prevention 33*(4), 485-498.

Parker, M. R. (1997). *Effects of Raising and Lowering Speed Limits on Selected Roadway Sections*. Publication FHWA-RD-92-084. FHWA, U.S.DOT.

Pawlovich, M.D., Souleyrette, R.R., and Strauss, T. (1998). A methodology for studying crash dependence on demographic and socioeconomic data. Conference: *Crossroads 2000,* Iowa State University and Iowa Department of Transportation, 209-215.

Pernia, J., Lu, J.J., and Peng, H. (2004). *Safety Issues Related to Two-way Left-turn Lanes* (Final Technical Report). Tampa, FL: Department of Civil and Environmental Engineering, University of South Florida.

Persaud, B.N., Retting, R.A., and Lyon, C. (2003). *Crash Reduction Following Installation of Centerline Rumble Strips on Rural Two-lane Roads*. Insurance Institute for Highway Safety, Arlington, VA USA.

Pfefer, F. M., Stenzel, W.W., and Lee, B.D. (1991). Safety Impact of the 65-mph Speed Limit: A Time Series Analysis. *Transportation Research Record 1318*, 22-33.

Plassmann, F., and Tideman, T.N. (2001). Does the right to carry concealed handguns deter countable crimes? Only a count analysis can say. *Journal of Law and Economics 44*(2-2), 771-798.

Press, S.J. (1982). *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference* 2^{nd} *Edition*. Robert E. Krieger Publishing Company, Malabar, Florida, USA.

Prinsloo, B., and Goudanas, C. (2003). Development of a crash prediction model for rural roads in NSW. *Proceedings of the 21st ARRB and 11th REAAA Conference*, Cairns, Australia.

Qin, X., Ivan, J.N., and Ravishanker, N. (2004). Selecting exposure measures in crash rate prediction for two-lane highway segments. *Accident Analysis and Prevention 36*(2), 183-191.

Qin, Z.S., Damien, P., and Walker, S. (2003). Scale mixture models with applications to Bayesian inference. *The Monte Carlo Method in the Physical Sciences*, 394-395.

Quddus, M.A., Chin, H.C., and Wang, J. (2001). Motorcycle crash prediction model for signalized intersections. *Seventh International Conference on Urban Transport and the Environment in the 21st Century*, 609-617, Lemnos, Greece.

Quddus, M.A., Noland, R.B., and Chin, H.C. (2002). An analysis of motorcycle injury and vehicle damage severity using ordered probit models. *Journal of Safety Research 33*(4)*, 445-462.*

Ranck, F. (2003). Applying safety and operational effects of highway design features to two-lane rural highways. *The Transportation Factor 2003. Annual Conference and Exhibition of the Transportation Association of Canada*. St. John's, Newfoundland and Labrador.

Roberts, G.O., and Smith, F.M. (1994). Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. *Stochastic Processes and their Applications 49*(2), 207-216.

Rodriguez, D.A., Rocha, M., Khattak, A.J., and Belzer, M.H. (2003). Effects of truck driver wages and working conditions on highway safety: Case study. *Transportation Research Record 1833*, 95-102.

Rodriguez-Yam, G.A. (2003). *Estimation for State-Space Models and Bayesian Regression Analysis with Parameter Constraints*. Ph.D. Dissertation, Colorado State University.

Schrank, D., and Lomax, T. (2002). *The 2002 Urban Mobility Report*. Texas Transportation Institute, The Texas A&M University System.

Schurr, K.S., McCoy, P.T., Pesti, G., Huff, R., and Spargo, B. (2004). *Appropriate Design Speed for Horizontal Curves Approaching a Stop*. University of Nebraska, Lincoln, Mid-America Transportation Center.

Scuffham, P.A., and Langley, J.D. (1997). Trends in cycle injury in New Zealand under voluntary helmet use. *Accident Analysis and Prevention 29*(1), 1-12.

Shankar, V., Milton, J. and Mannering F. (1997). Modeling accident frequencies as zero-altered probability processes: An empirical inquiry. *Accident Analysis and Prevention 29*(6), 829-837.

Shankar, V., Chayanan, S., Sittikariya, S., Shyu, M.B., Juvva, N.K. and Milton, J.C. (2003). The marginal impacts of design, traffic, weather and related interactions on roadside crashes. *Proceedings of the Transportation Research Board 83rd Annual Meeting*, January 2004, Washington D.C, USA.

Shankar V.N., Ulfarsson, G.F., Pendyala, R.M., and Nebergall, M.B. (2003). Modeling crashes involving pedestrians and motorized traffic. *Safety Science 41*(7), 557-640.

Shankar, V., Milton, J., and Mannering, F.L. (1997). Modeling accident frequency as zero-altered probability processes: an empirical inquiry. *Accident Analysis and Prevention 29*(6), 829-837.

Smith, A.F.M., and Roberts, G.O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B 55*(1), 3-23.

Snyder, D. (1989). Speeding, coordination and the 55 mph limit: Comment. *The American Economic Review 79*(4), 922-925.

Solomon, D. (1964). *Accidents on Main Rural Highways Related to Speed, Driver, and Vehicle*. Federal Highway Administration, Washington, D.C., July.

Streff, F.M., and Kostyniuk, L.P. (1997). *Michigan fatal-traffic-crash patterns: 1993-1995*. University of Michigan Transportation Research Institute. Report No.: HS-042414.

Swendsen, R.H., and Wang, J.S. (1987). Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters 58*(2), 86-88.

Tanner, M.A., and Wong, W.H. (1987). The calculation of posterior distribution by data augmentation. *Journal of the American Statistical Association 82*(398), 520-540.

Terza, J.V., and Wilson, P. (1990). Analyzing frequencies of several types of events: a mixed multinomial-Poisson approach. *The Review of Economics and Statistics 72*(10), 108-115.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics 22*(4), 1701-1728.

TRB. (1998). *Managing Speed – Review of Current Practice for Setting and Enforcing Speed Limits*. TRB Special Report 254, National Academy Press, Washington, D.C.

Tsionas, E.G. (1999). Bayesian analysis of the multivariate Poisson distribution. *Communications in Statistics – Theory and Methods 28*, 431-451.

Tsionas, E.G. (2001). Bayesian multivariate Poisson regression. *Communications in Statistics – Theory and Methods 30*(2), 243-255.

Ulfarsson, G.F., and Shankar, V.N. (2003). Accident count model based on multiyear cross-sectional roadway data with serial correlation. *Transportation Research Record 1840*, 193-197.

U.S. Department of Transportation. (1997). *Departmental Guidance for the Valuation of Travel Time in Economic Analysis*. ostpxweb.dot.gov/policy/Data/VOT97guid.pdf.

U.S. Department of Transportation. (2003). *Revised Departmental Guidance: Valuation of Travel Time in Economic Analysis*. ostpxweb.dot.gov/policy/Data/VOTrevision1_2-11-03.pdf.

US Department of Transportation, Bureau of Statistics. (2004). Table 1-5: *U.S. Public Road and Street Mileage by Functional System*. Retrieved January 31, 2006, http://www.bts.gov/publications/national_transportation_statistics/2004/html/table_01_05.html

US Department of Transportation, Bureau of Statistics. (2005). Table 1-33: *Roadway Vehicle-Miles Traveled (VMT) and VMT per Lane-Mile by Functional Class*. Retrieved March 12, 2006, http://www.bts.gov/publications/national_transportation_statistics/2005/html/table_01_33.html

United States General Accounting Office (USGAO). (2004). *Highway Safety – Federal and State Efforts to Address Rural Road Safety Challenges*. Report to Congressional Committees, GAO-04-663.

Venugopal, S., and Tarko, A. (2000). Safety models for rural freeway work zones. *Transportation Research Record 1715*, 1-9.

Viner, J.G. (1995). Rollovers on sideslopes and ditches. *Accident Analysis and Prevention 27*(4), 483-491.

Vogt, A. (1998). Accident models for two-lane rural segments and intersections. *Transportation Research Record 1635*, 18-29.

Vogt, A. (1999). *Crash Models for Rural Intersections: Four-lane by Two-lane Stop-controlled and Two-lane by Two-lane Signalized* (FHWA-RD-99-

128). McLean, VA: Turner Fairbank Highway Research Center, Federal Highway Administration.

Vogt, A., and Bared, J.G. (1998). *Accident Models for Two-lane Rural Roads: Segments and Intersections* (FHWA-RD-98-133). McLean, VA: Turner Fairbank Highway Research Center, Federal Highway Administration.

West, L.B., and Dunn, J.W. (1971). Accidents, speed deviations, and speed limits. *Traffic Engineering 41*(10), 52-55.

Whitfield, R.A., and Whitfield, A.K. (2004). Improving surveillance for injuries associated with potential motor vehicle safety defects. *Injury Prevention 10*(2), 88-92.

Winkelmann, R. (2000). Seemingly unrelated negative binomial regression. *Oxford Bulletin of Economics and Statistics 62*(4), 553-560.

Winkelmann, R. (2003). *Econometric Analysis of Count Data 4th Edition*. Springer – Verlag Berlin Heidelberg New York.

Wong, S. H., and Wu, Y. (1998). The efficacy of safety policies on traffic fatalities in Singapore. *Accident Analysis and Prevention 30*(6), 745-754.

Wooldridge, J.M. (1999). Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics 90*(1), 77-97.

Wooldridge, M.D., Fitzpatrick, K., Harwood, D.W., Potts, I.B. Elefteriadou, L., and Torbic, D.J. (2003). *Geometric Design Consistency on High-speed Rural Two-Lane Roadways*. NCHRP Report 502, Transportation Research Board.

World Heath Organization (WHO) (2002a). *Estimates of numbers of deaths by sex, cause and WHO Region for 2002*. Retrieved September 21, 2005, http://www3.who.int/whosis/menu.cfm?path=burden,burden_estimates,burden_estimates_2002N,burden_estimates_2002N_2002Rev,burden_estimates_2002N_2002Rev_Region&language=english.

World Heath Organization (WHO) (2002b). *Estimates of Years of Life Lost due to premature mortality by sex, cause and WHO Region for 2002*. Retrieved September 21, 2005, http://www3.who.int/whosis/menu.cfm?path=burden,burden_estimates,burden_estimates_2002N,burden_estimates_2002N_2002Rev,burden_estimates_2002N_2002Rev_Region&language=english.

Zegeer, C.V., Deen, R.C., and Mayes, J.G. (1981). Effect of Lane and Shoulder Widths on Accident Reduction on Rural, Two-lane Roads. *Transportation Research Record 806*, 33-43.

Zegeer, C.V., Stewart, J.R., Huang, H.H., and Lagerwey, P.A. (2001). Safety Effects of Marked vs. Unmarked Crosswalks at Uncontrolled Locations: Analysis of Pedestrian Crashes in 30 Cities (with discussion and closure). *Transportation Research Record 1773*, 56-68.

VITA

Jianming Ma was born in Xinjiang, China on July 16, 1972, the son of Guiying Su and Zhigang Ma. Jianming Ma earned his Bachelor's degree from the China University of Geosciences in 1995 and his Master's degree in Transportation Engineering from the Beijing University of Technology (BJUT) in 1998. After graduation, he worked at the Transportation Research Center of BJUT as a research faculty and conducted several research projects related to traffic engineering, transportation planning, computer simulation, non-motorized transport, and intelligent transportation systems (ITS).

He started his Ph.D. study at the University of Texas at Austin in fall of 2002 as a graduate student research assistant. During the three-and-half-year graduate study at the university, he has developed a keen interest in the analysis of traffic crash frequency and severity using advanced statistical techniques. He has been a key contributor to three research projects: "Impacts and Other Implications of Raised Speed Limits on High-Speed Roads" funded by the National Cooperative Highway Research Program (NCHRP), "Public Perceptions of Toll Roads," and "Rural Two-Lane Roadway Crash Analysis" sponsored by the Texas Department of Transportation (TxDOT). He also is an author of several published papers on such topics.

Permanent Address: 64 Tuan 9 Lian, Huocheng, Xinjiang, China 835214

This dissertation was typed by the author.