A Multivariate Poisson-Lognormal Regression Model for Prediction of Crash Counts by Severity, using Bayesian Methods

Jianming Ma, Ph.D., The University of Texas at Austin.
6.9 E. Cockrell Jr. Hall, Austin, TX 78712-1076, mjming@mail.utexas.edu

Kara M. Kockelman, Associate Professor & William J. Murray Jr. Fellow of Civil Engineering
The University of Texas at Austin, 6.9 E. Cockrell Jr. Hall, Austin, TX 78712-1076
kkockelm@mail.utexas.edu, Phone: 512-471-0210, FAX: 512-475-8744

Paul Damien, B. M. (Mack) Rankin, Jr. Professor in Business Administration
The University of Texas at Austin, CBA 5.242, Austin, TX 78712
paul.damien@mccombs.utexas.edu, Phone: 512-232-9461, FAX: 512-471-0587

The following paper is a pre-print and the final publication can be found in
Accident Analysis and Prevention, 40 (3):964–975, 2008.
Presented at the 86th Annual Meeting of the Transportation Research Board, January 2007

ABSTRACT
Numerous efforts have been 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, which may neglect shared information in unobserved error terms, reduce efficiency in parameter estimates, and lead to potential biases in sample databases. This paper offers a multivariate Poisson-lognormal (MVPLN) specification that simultaneously models injuries by severity. 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 for crashes on Washington State rural two-lane highways. 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 overdispersion in 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.

KEY WORDS
Bayesian inference, Bayes' theorem, crash severity, Gibbs sampler, highway safety, Metropolis-Hastings algorithm, Markov chain Monte Carlo (MCMC) simulation, multivariate Poisson-lognormal regression

INTRODUCTION

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. In the U.S. traffic crashes bring about more loss of human life (as measured in human-years) than almost any other cause – falling behind only cancer and heart disease (NHTSA, 2005). The situation is of particular interest on rural two-lane roadways, which experience significantly higher fatality rates than urban roads. The annual cost of traffic crashes is estimated to be $231 billion, or $820 per capita in 2000 *(*Blincoe et al., 2000*)*. These costs do not include the cost of delays imposed on other travelers, which also are significant, particularly when crashes occur on busy roadways. Schrank and Lomax (2002) estimate that over half of all traffic delays are due to non-recurring events, such as crashes, costing on the order of $1,000 per peak-period driver per year, particularly in urban areas. Thus, while vehicle and roadway design are improving, and growing congestion may be reducing impact speeds, crashes are becoming more critical in many ways, particularly in societies that continue to motorize.

Given the importance of roadways safety, there has been considerable crash prediction research (see, e.g., Hauer, 1986, 1997, and 2001; Abdel-Aty, and Radwan, 2000; Ulfarsson and Shankar, 2003; Kweon and Kockelman, 2000; Lord and Persaud, 2000; Lord et al., 2005; Ma and Kockelman, 2006; Karlaftis and Rarko, 1998; Shankar et al., 1998; Khattak et al., 2006). Crash frequencies are commonly collected by severity on relatively homogenous roadway segments, supporting the development of crash count models. However, such 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: interdependence due to latent factors is likely to exist across crash rates at different levels of severity for a specific segment of roadway. Recently, Ma and Kockelman (2006) applied a multivariate Poisson (MVP) specification to model crash counts at different levels of severity simultaneously. However, this MVP specification allows only for a common added Poisson error term, resulting in equal positive correlations across crash counts and a very specific data pattern where all counts are equally shifted. In addition, this MVP specification does not allow for overdispersion.

Using a multivariate Poisson-lognormal (MVPLN) specification, as well as Bayesian estimation techniques, this work models correlated traffic crash counts simultaneously at different levels of severity. 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. The data come from Washington State rural two-lane highways in 2002, using the Highway Safety Information System (HSIS) database. The results lend themselves to recommendations for highway safety treatments and general design policies.

This paper is organized as follows: Related research studies are reviewed first. The model's formulation and data sets are then discussed, followed by estimation results, concluding remarks, and future research directions.

LITERATURE REVIEW

Models of crash (or injury) counts can be classified into two major streams: (1) the conventional univariate Poisson and related models, such as the negative binomial (NB); (2) potentially more realistic specifications, like the MVP and MVPLN. The first stream has provided a means for investigating associations between crash frequency and many crucial factors, such as traffic volume, access density, posted speed limit and number of lanes (see, e.g., Miaou et al., 1993; Miaou and Lum, 1993; Miaou 1994, 1996 and 2001;Fridstrøm et al., 1995; Johansson, 1996; Vogt and Bared, 1998; Vogt, 1999; Balkin and Ord, 2001; Zegeer et al., 2002; Pernia, 2004). 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., Lord et al. 2005; Shankar et al., 1997; Garber and Wu, 2001; Lee and Mannering, 2002; Kumara and Chin, 2003; Miaou and Lord, 2003; Rodriguez et al. 2003; Shankar et al. 2003; Noland and Quddus, 2004; Qin et al., 2004).

Due to computational and statistical advances, panel data (in which a cross-section of segments, intersections, etc. is observed over time) have become more amenable to rigorous analysis. In traffic crash analyses, there are a great many unobserved explanatory variables that affect frequencies and severities. Panel data can be used to deal with heterogeneity among individuals. To address the heterogeneity, many recent studies have used (univariate) panel count data models, such as random-effect negative binomial (RENB) and fixed-effect negative binomial (FENB) regression models (Kweon and Kockelman, 2000; Karlaftis and Rarko, 1998; Shankar et al., 1998; Chin and Quddus, 2003).

Such past research endeavors, however, have neglected the role of unobserved factors across different types of counts (e.g., the number of fatalities and the number of debilitating injuries). Recognizing the need for such considerations, Ladron de Guevara and Washington (2004) investigated the simultaneity of fatality and injury crash outcomes. Bijleveld (2005) also examined the correlation structure between crash and injury counts. As expected, he found significant correlations. However, he did not control for any covariates. Multivariate models (of count data), like Ma and Kockelman's MVP (2006) or Li et al's MVZIP (1999), can help correct for this.

This work models correlated traffic crash counts simultaneously at different levels of severity using a MVPLN specification, allowing for a very general correlation structure as well as overdispersion. Such specifications are challenging to estimate. Karlis (2003) developed an EM algorithm for an MVP model, and Ma and Kockelman (2006) used Gibbs sampling, as well as Metropolis-Hastings algorithms, within an MCMC simulation framework.

In recent years, Bayesian methods have found several applications in traffic crash analysis. Christiansen et al. (1992) and MacNab (2003) developed hierarchical Poisson models for crash counts and surveillance data. Miaou and Song (2005) developed a Bayesian multivariate spatial generalized linear mixed model (GLMM) to rank sites for safety improvements using Texas' county-level crash data. And Liu et al. (2005) used a hierarchical Bayesian framework to estimate ZIP regression models and develop safety performance functions (SPFs) for two-lane highways. Pawlovich et al. (2006) employed a Bayesian approach to assess impacts of road design measures on crash frequencies and rates. And Washington and Oh (2006) developed a Bayesian methodology for incorporating expert judgment in ranking countermeasure effectiveness under uncertainty.

Bayesian estimation methods generate a multivariate posterior distribution across all parameters of interest, as opposed to the traditional maximum likelihood estimation approach, which emphasizes and offers only the modal values of parameters (and relies on asymptotic properties to ascertain covariance).

This paper introduces an MVPLN approach to simultaneously model injury counts by severity. A Gibbs sampler and a Metropolis-Hastings (M-H) algorithm are used to estimate the parameters of interest using Bayesian methods. For comparison purposes, a series of independent (univariate) Poisson models for injury counts also are estimated.

MODEL STRUCTURE AND ESTIMATION

Mathematical Formulation

Univariate Poisson regression models cannot account for correlations for different levels of severity; instead, one needs multivariate count data models. For instance, 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, thus introducing correlation. Several such models have been developed (see, e.g., Karlis, 2003; Arbous and Kerrich, 1951; King, 1989; Winkelmann, 2000; Kockelman, 2001; Tsionas, 2001). However, these specifications support only a common unobserved error term 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 (e.g., there is no spatial correlation[1]). The variance-covariance matrix of $\mathbf{y}$ can be expressed as below:

$$Var\left(\mathbf{y}_{nS\times 1}\right)=\begin{bmatrix}\Omega_1 & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \Omega_2 & \cdots & \mathbf{0}\\ & & \vdots & \\ \mathbf{0} & \mathbf{0} & \cdots & \Omega_n\end{bmatrix} \quad (1)$$

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

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 [$i=1,2,\ldots,n$, where n is the number of roadway segments], s denote the severity level [$s=1,2,\ldots,S$, where S is the number of severity levels], and $\boldsymbol{\varepsilon}=\left(\vec{\boldsymbol{\varepsilon}}_1',\vec{\boldsymbol{\varepsilon}}_2',\ldots,\vec{\boldsymbol{\varepsilon}}_n'\right)'$ 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.

$$y_{is}\left|\vec{\boldsymbol{\varepsilon}}_i,\beta_s,x_{is}\right.\sim Poisson\left(\lambda_{is}\right) \quad (3)$$

where $\lambda_{is}=\exp\left(x_{is}'\beta_s+\varepsilon_{is}\right)$. The unobserved heterogeneity terms $\vec{\boldsymbol{\varepsilon}}_i$ are assumed to be uncorrelated with the control (i.e., explanatory) variables.

Let $\Lambda_i=diag\left(\vec{\boldsymbol{\lambda}}_i\right)$. This is an S×S matrix, where $\vec{\boldsymbol{\lambda}}_i=\left(\lambda_{i1},\lambda_{i2},\ldots,\lambda_{iS}\right)$ and $\lambda_{is}=\xi_{is}u_{is}$.

Let $\vec{\mathbf{u}}_i=\exp\left(\vec{\boldsymbol{\varepsilon}}_i\right)$, where $\vec{\mathbf{u}}_i=\left(u_{i1},u_{i2},\ldots,u_{iS}\right)'$. Conditioning on β and Σ, the mean and covariance matrix of the marginal distribution of $\vec{\mathbf{y}}_i$ can be obtained as follows:

[1] In reality, spatial correlation may exist and be significant. For example, zoning and design policies create correlation across sites within a city; access management and other policies may simply shift the location of certain crash types. The former leads to positive correlation, the latter to negative.

$$E\left(\vec{\mathbf{y}}_i \left| \beta, x_i, \Sigma\right.\right) = E_{\vec{\mathbf{u}}_i}\left(E_{\vec{\mathbf{y}}_i|\vec{\mathbf{u}}_i}\left(\vec{\mathbf{y}}_i \left| \beta, x_i, \vec{\mathbf{u}}_i, \Sigma\right.\right)\right) = E_{\vec{\mathbf{u}}_i}\left(diag\left(\vec{\boldsymbol{\xi}}_i\right)\vec{\mathbf{u}}_i\right) = \vec{\boldsymbol{\lambda}}_i \tag{4}$$

$$Var\left(\vec{\mathbf{y}}_i \left| \beta, x_i, \Sigma\right.\right) = E_{\vec{\mathbf{u}}_i}\left(Var_{\vec{\mathbf{y}}_i|\vec{\mathbf{u}}_i}\left(\vec{\mathbf{y}}_i \left| \beta, x_i, \vec{\mathbf{u}}_i, \Sigma\right.\right)\right) + Var_{\vec{\mathbf{u}}_i}\left(E_{\vec{\mathbf{y}}_i|\vec{\mathbf{u}}_i}\left(\vec{\mathbf{y}}_i \left| \beta, x_i, \vec{\mathbf{u}}_i, \Sigma\right.\right)\right) \text{ (Greene, 2003)}$$

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

$$= \Lambda_i + \Lambda_i\left[\exp(\Sigma) - \mathbf{11'}\right]\Lambda_i \tag{5}$$

where $\beta = \left(\beta_1, \beta_2, \ldots, \beta_S\right)'$, $x_i = \left(x_{i1}, x_{i2}, \ldots, x_{iS}\right)'$ and $\vec{\boldsymbol{\xi}}_i = \left(\xi_{i1}, \xi_{i2}, \ldots, \xi_{iS}\right)'$. The length of β is $k = k_1 + k_2 + \cdots + k_S$, where k_s is the length of β_s.

From Equation (5), the variance-covariance terms, across counts, can be obtained as follows:

$$\begin{aligned} Cov\left(y_{is}, y_{il}\right) &= 0 + \lambda_{is}\left[\exp\left(\sigma_{sl}\right) - 1\right]\lambda_{il} \\ &= \xi_{is}\exp\left(\sigma_{ss}/2\right)\left[\exp\left(\sigma_{sl}\right) - 1\right]\xi_{il}\exp\left(\sigma_{ll}/2\right), \text{ for } s \neq l \\ Var\left(y_{is}, y_{is}\right) &= \lambda_{is} + \lambda_{is}\left[\exp\left(\sigma_{ss}\right) - 1\right]\lambda_{is} \end{aligned} \tag{6}$$

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

$$\begin{aligned} Corr\left(y_{is}, y_{il}\right) &= \frac{\xi_{is}\left[\exp\left(\sigma_{sl}\right) - 1\right]\xi_{il}}{\sqrt{\xi_{is}\exp\left(-\sigma_{ss}/2\right) + \xi_{is}^2\left[\exp\left(\sigma_{ss}\right) - 1\right]}\sqrt{\xi_{il}\exp\left(-\sigma_{ll}/2\right) + \xi_{il}^2\left[\exp\left(\sigma_{ll}\right) - 1\right]}} \\ &= \frac{\exp\left(\sigma_{sl}\right) - 1}{\sqrt{\xi_{is}^{-1}\exp\left(-\sigma_{ss}/2\right) + \exp\left(\sigma_{ss}\right) - 1}\sqrt{\xi_{il}^{-1}\exp\left(-\sigma_{ll}/2\right) + \exp\left(\sigma_{ll}\right) - 1}} \end{aligned} \tag{7}$$

where $s \neq l$.

This correlation is unrestricted and can be positive or negative, depending on the sign of σ_{sl}, the (s,l) element of Σ. Moreover, this specification implies overdispersion[2], since $\sigma_{ss} > 0$ for $s = 1, 2, \ldots, S$.

Based on Equation (3), the likelihood of observation i can be represented by the following equation:

$$P\left(\vec{\mathbf{y}}_i \left| \vec{\boldsymbol{\varepsilon}}_i, \beta, x_i\right.\right) = \prod_{s=1}^{S} f_{Poisson}\left(y_{is} \left| \lambda_{is}\right.\right) \tag{8}$$

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 $\vec{\boldsymbol{\varepsilon}}_i$,

$$P\left(\vec{\mathbf{y}}_i \left| \vec{\boldsymbol{\lambda}}_i, \Sigma\right.\right) = \int \prod_{s=1}^{S} f_{Poisson}\left(y_{is} \left| x_{is}, \beta_s, \varepsilon_{is}\right.\right)\phi_S\left[\vec{\boldsymbol{\varepsilon}}_i \left| \mathbf{0}, \Sigma\right.\right] d\vec{\boldsymbol{\varepsilon}}_i \tag{9}$$

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

Estimating Parameters via MCMC

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

In order to illuminate crash rate relationships, the MVPLN model's unknown parameters need to be estimated. Chib et al. (1998) showed how to estimate a posterior distribution of unknown parameters for their models of panel count data[3], 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 is commonly used as a conjugate prior for the inverse of variance-covariance parameters. According to Press (1982), the Wishart and normal distributions are very helpful for multivariate analysis. Suppose that the parameters (β,Σ) independently have the prior distributions:

$$\beta \sim \phi_k\left(\beta_0, V_{\beta_0}\right),\ \Sigma^{-1} \sim f_W\left(\nu_\Sigma, V_\Sigma\right) \tag{10}$$

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. The prior distribution for β_s 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[4], the latent effects $\boldsymbol{\varepsilon}$ can be thought of as ("nuisance") parameters to be estimated. Therefore, the joint posterior density of Σ, $\boldsymbol{\varepsilon}$, and β is written as follows:

$$\pi\left(\Sigma, \boldsymbol{\varepsilon}, \beta \middle| y, X\right) \propto \phi_k\left(\beta_0, V_{\beta_0}\right) f_W\left(\nu_\Sigma, V_\Sigma\right) \prod_{i=1}^{n} \prod_{s=1}^{S} f_{Poisson}\left(y_{is} \middle| x_{is}, \beta_s, \varepsilon_{is}\right) \phi_S\left(\vec{\boldsymbol{\varepsilon}}_i \middle| \mathbf{0}, \Sigma\right) \tag{11}$$

Thanks to this technique, 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 $\pi^p\left(\cdot \middle| \cdot\right)$ denotes the posterior conditional density function.

The draws are sampled sequentially using the most recent values of the conditioning variables at each step.

Gibbs Sampler with Embedded M-H Algorithms

After manipulating the posterior equation (11), the posterior of Σ^{-1} conditional on data and other parameters can be written as

---

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

[4] Data augmentation views unobserved or latent variables as unknown parameters (to be estimated), in order to establish iterative algorithms.

$$\pi\left(\Sigma^{-1}\left|\boldsymbol{\varepsilon}\right.\right) \propto f_W\left(\Sigma^{-1}\left|\nu_\Sigma, V_\Sigma\right.\right)\prod_{i=1}^{n}\phi_S\left(\vec{\boldsymbol{\varepsilon}}_i\left|\mathbf{0},\Sigma\right.\right) \tag{12}$$

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

After manipulating Equation (12), 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,

$$\Sigma^{-1}\left|\varepsilon\right. \sim f_W\left(n+\nu_\Sigma,\left[V_\Sigma^{-1}+\sum_{i=1}^{n}\left(\vec{\boldsymbol{\varepsilon}}_i\vec{\boldsymbol{\varepsilon}}'\right)_i\right]^{-1}\right) \tag{13}$$

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

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

$$\pi\left(\vec{\boldsymbol{\varepsilon}}_i\left|\vec{\mathbf{y}}_i,x_i,\beta,\Sigma\right.\right)=C_i\phi_S\left(\vec{\boldsymbol{\varepsilon}}_i\left|\Sigma\right.\right)\prod_{s=1}^{S}\exp\left(-\lambda_{is}\right)\lambda_{is}^{y_{is}}=C_i\pi^p\left(\vec{\boldsymbol{\varepsilon}}_i\left|\vec{\mathbf{y}}_i,x_i,\beta,\Sigma\right.\right), \tag{14}$$

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, as described below.

Following Chib et al. (1998), the multivariate t distribution is used as the proposal density. Let $\hat{\vec{\boldsymbol{\varepsilon}}}_i=\arg\max_{\vec{\boldsymbol{\varepsilon}}_i}\left[\ln\pi^p\left(\vec{\boldsymbol{\varepsilon}}_i\left|\vec{\mathbf{y}}_i,x_i,\beta,\Sigma\right.\right)\right]$ and $V_{\varepsilon_i}=\left(-H_{\varepsilon_i}\right)^{-1}$ be the inverse of the Hessian of $\ln\pi^p\left(\vec{\boldsymbol{\varepsilon}}_i\left|\vec{\mathbf{y}}_i,x_i,\beta,\Sigma\right.\right)$ 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 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 $f_T\left(\vec{\boldsymbol{\varepsilon}}_i\left|\hat{\boldsymbol{\varepsilon}}_i,V_{\varepsilon_i},\nu_\varepsilon\right.\right)$, 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[5] lies between 20 and 45 percent[6]). A proposal value $\vec{\boldsymbol{\varepsilon}}_i^*$ is drawn from $f_T\left(\vec{\boldsymbol{\varepsilon}}_i\left|\hat{\boldsymbol{\varepsilon}}_i,V_{\varepsilon_i},\nu_\varepsilon\right.\right)$, and the chain moves to $\vec{\boldsymbol{\varepsilon}}_i^*$ from the current point $\vec{\boldsymbol{\varepsilon}}_i$ with probability

[5] 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.

[6] 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.

$$\alpha\left(\vec{\boldsymbol{\varepsilon}}_i,\vec{\boldsymbol{\varepsilon}}_i^*\left|\vec{\mathbf{y}}_i,x_i,\beta,\Sigma\right.\right)=\min\left\{\frac{\pi^p\left(\vec{\boldsymbol{\varepsilon}}_i^*\left|\vec{\mathbf{y}}_i,x_i,\beta,\Sigma\right.\right)f_T\left(\vec{\boldsymbol{\varepsilon}}_i\left|\hat{\boldsymbol{\varepsilon}}_i,V_{\varepsilon_i},\nu_\varepsilon\right.\right)}{\pi^p\left(\vec{\boldsymbol{\varepsilon}}_i\left|\vec{\mathbf{y}}_i,x_i,\beta,\Sigma\right.\right)f_T\left(\vec{\boldsymbol{\varepsilon}}_i^*\left|\hat{\boldsymbol{\varepsilon}}_i,V_{\varepsilon_i},\nu_\varepsilon\right.\right)},1\right\} \tag{15}$$

If $\alpha\left(\vec{\boldsymbol{\varepsilon}}_i,\vec{\boldsymbol{\varepsilon}}_i^*\left|\vec{\mathbf{y}}_i,x_i,\beta,\Sigma\right.\right)$ 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.

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

$$\begin{aligned}\pi^p\left(\beta_s\left|y,X,\boldsymbol{\varepsilon},\Sigma\right.\right)&=\pi^p\left(\beta_s\left|y_{\cdot s},X,\varepsilon_{\cdot s},\Sigma\right.\right)\prod_{j=1,j\neq s}^{S}\pi^p\left(\beta_j\left|y_{\cdot j},X,\varepsilon_{\cdot j},\Sigma\right.\right)\\&=C_{-s}\pi^p\left(\beta_s\left|y_{\cdot s},X,\varepsilon_{\cdot s},\Sigma\right.\right)\\&\propto\phi_{k_s}\left(\beta_s\left|\beta_{0s},V_{\beta_{0s}}\right.\right)\prod_{i=1}^{n}\exp\left[-\exp\left(x_{is}'\beta_s+\varepsilon_{is}\right)\right]\left[\exp\left(x_{is}'\beta_s+\varepsilon_{is}\right)\right]^{y_{is}}\\&\propto\phi_{k_s}\left(\beta_s\left|\beta_{0s},V_{\beta_{0s}}\right.\right)p\left(y_{\cdot s}\left|\beta_s,\varepsilon_{\cdot s}\right.\right)\end{aligned} \tag{16}$$

where $C_{-s}=\prod_{j=1,j\neq s}^{S}\pi^p\left(\beta_j\left|y_{\cdot j},X,\varepsilon_{\cdot j},\Sigma\right.\right)$ (which does not involve β_s and thus serves as a constant), and $p\left(y_{\cdot s}\left|X,\beta_s,\varepsilon_{\cdot s}\right.\right)=\prod_{i=1}^{n}\exp\left[-\exp\left(x_{is}'\beta+\varepsilon_{is}\right)\right]\left[\exp\left(x_{is}'\beta+\varepsilon_{is}\right)\right]^{y_{is}}$ 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 the 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 once again serves as the proposal density. Let $\hat{\beta}_s=\arg\max_{\beta_s}\left[\ln\pi^p\left(\beta_s\left|y_{\cdot s},X,\varepsilon_{\cdot s},\beta_{-s},\Sigma\right.\right)\right]$ 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

$$\alpha\left(\beta_s,\beta_s^*\left|y,X,\boldsymbol{\varepsilon},\beta_{-s},\Sigma\right.\right)\min\left\{\frac{\pi^p\left(\beta_s^*\left|y,X,\boldsymbol{\varepsilon},\beta_{-s},\Sigma\right.\right)f_T\left(\beta_s\left|\hat{\beta}_s,V_{\beta_s},\nu_\beta\right.\right)}{\pi^p\left(\beta_s\left|y,X,\boldsymbol{\varepsilon},\beta_{-s},\Sigma\right.\right)f_T\left(\beta_s^*\left|\hat{\beta}_s,V_{\beta_s},\nu_\beta\right.\right)},1\right\} \tag{17}$$

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.

DATA DESCRIPTION

The crash data sets used here were collected from Washington State through the Highway Safety Information System (HSIS). In order to examine traffic crashes patterns on rural two-lane roadways, this research considers crashes in the Puget Sound region. A random sample of 60% of all 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[7] and a total of 510 miles) are available for analysis. This sample contains 16 fatal crashes, 50 disabling-injury crashes, 180 non-disabling-injury crashes, 175 possible-injury crashes and 532 property-damage-only (PDO). Table 1 reports summary statistics for the dependent and independent variables employed in the analysis. A variety of readily available variables are controlled for in the model, including design features, traffic intensity, location information, and roadway functional classification.

MODEL ESTIMATION AND RESULTS

Model Estimation

The MVPLN regression model 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 $I_5=\begin{bmatrix}1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&0&1\end{bmatrix}$. The MLE estimates for the five univariate Poisson models can be found in Ma (2006). A Gibbs sampler and two M-H algorithms were coded in the R language (an open-source statistical computing environment described at http://www.r-project.org/). The prior distributions for the estimation are defined by the hyperparameters $\nu_\Sigma=10$, $V_\Sigma^{-1}=I_5$, $\beta_{0s}=(0,0,\ldots,0)'$, 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 $5\times14=70$ $\beta's$ and each of the $7{,}773\times5=38{,}865$ ε's, respectively. The initial 1,000 draws were discarded as “burn-ins.” An adequate burn-in period eliminates the influence of the starting values. To help ensure chain convergence, the

[7] It is quite possible that very short segments do not faithfully represent the actual location of crashes, since police officers may locate crashes only to the nearest tenth of a mile. Cluster analysis, wherein similar segments/conditions are merged (providing higher crash counts) can address some of this bias in reporting. Ma and Kockelman (2006) conducted such an analysis with Washington State data.

Gibbs sampler and the two M-H algorithms were implemented using two sets of starting values[8] and both converged at the same posterior distribution of parameters. Estimation results are presented in Tables 2 through 6.

Based on the posterior density of Σ, 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 offer somewhat better predictions.

Interpretation of Results

The following discussion of results emphasizes disabling and fatal injuries (Tables 5 and 6), 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., 2000). Tables 2 through 4 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 2 through 6 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 7 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. A 26.6% higher average annual daily traffic level (rising from 3757 to 4757 vehicles) is predicted to increase total crash count by 16.4% — while reducing the total crash *rate* by 5.51%. In this way, the MVPLN model results offer statistically (and practically) significant insights into crash counts' dependence on roadway design.

The magnitudes of the parameter estimates for the MVPLN specification are not directly comparable to those of univariate Poisson models (shown in Ma, 2006) or those of univariate negative binomial (UVNB) models (also shown in Ma, 2006). 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 cases. 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 coefficient signs for remaining control variables are not in agreement across all three models, indicating that specification choice is important to a proper understanding of crash count relationships.

[8] Zeros were used as the starting values for β in the second chain.

Based on the description of the correlation effects earlier in the paper, we should expect the MVPLN specification to yield 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 as shown in Table 8. Note that this is not a theoretical point, but rather an empirical one: in other words, where potential correlation exists, it should be modeled. Like the MVNB approach, our approach allows for overdispersion. 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.

In addition, out-of-sample predictions from both univariate and multivariate models are compared for the different groups. Table 9 suggests that the MVPLN model with MCMC draws predicts better than the univariate models (UVP and UVNB). This is because the MVPLN model addresses the issue of unobserved heterogeneity and allows for correlations among crash counts at all levels of severity.

CONCLUSIONS

Roadway safety is a major concern for the general public -- and its transport 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, almost all 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 neglect 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 an MVPLN regression specification, which allows for a rather general correlation structure as well as overdispersion. With recent advancements in crash modeling and Bayesian statistics, parameter estimation is achieved 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 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 et al., 2003; Pawlovich et al., 1998.) Additionally, the high level of correlation between PDO and disabling crashes may indicate some ambiguity or weakness in severity classification schemes, if one believes that unobserved heterogeneity in omitted variables should generate significant correlation (e.g., in data sets with relatively few control variables available).

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 subsequent data analysis. Semi-parametric and nonparametric regression analysis relaxes these assumptions[9] (see, e.g., Gurmu et al., 1999; Wooldridge, 1999; Alfò and Trovato, 2004). For example, Gurmu et al. (1999) developed a semiparametric approach to investigate overdispersed count data using a Laguerre series expansion of an unknown density function for unobserved heterogeneity.

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 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. Such interactions may exist (such as combinations of horizontal and vertical curvature on the same segment), and these should be examined in the future endeavors of this type.

ACKNOWLEDGEMENTS

The authors thank the Texas Department of Transportation (TxDOT) for funding this research under contract number 0-4965. The authors also are grateful to the FHWA's Yusuf Mohamedshah for provision of the crash data sets, Dr. Miaou for offering useful discussions related to methods of analysis, and Ms. Annette Perrone for editorial assistance.

---

[9] 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.

REFERENCES

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

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

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

Balkin, S., and Ord, J.K. (2001). Assessing the impact of speed-limit increases on fatal interstate crashes. *Journal of Transportation and Statistics 4*(1), pp.1-26.

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), pp.591-600.

Blincoe, L., Seay, A., Zaloshnja, E., Miller, T., Romano, E., Luchter, S., and Spicer R. (2002). *The Economic Impact of Motor Vehicle Crashes, 2000*. U.S. DOT, Washington, D.C.

Chib, S., and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician 49*(4), pp.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), pp.33-54.

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), pp.253-259.

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.

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

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), pp.1-20.

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

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.

Greene, W.H. (2003). *Econometric Analysis 5th Edition*, by Pearson Education, Inc., Upper Saddle River, New Jersey.

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

Hauer, E. (1986). On the estimation of the expected number of accidents. *Accident Analysis and Prevention, 18*(1), pp.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), pp.799-808.

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

Karlaftis, M.G. and Rarko, A.P. (1998). Heterogeneity considerations in accident modeling. *Accident Analysis and Prevention 30*(4), pp.425-433.

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

Khattak, A., Zhang, L., Hochstein, J.L. and Tee, S.C. (2006). Crash analysis of expressway intersections in Nebraska. *Proceedings of the Transportation Research Board's 85th Annual Meeting*, Washington DC.

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. *Accident Analysis and Prevention*, Forthcoming.

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

Kockelman, K. (2001). A Model for Time- and Budget-Constrained Activity Demand Analysis. *Transportation Research B 35*(3), pp.255-269.

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), pp.53-57.

Kweon, Y.J., and Kockelman, K. (2005). The safety effects of speed limit changes: use of panel models, including speed, use, and design variables. *Transportation Research Record 1908*, pp.148-158.

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*, pp.191-199.

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), pp.149-161.

Li, C.C., Lu, J.C., Park, J., Kim, K., Brinkley, P.A. & Peterson, J.P. (1999). Multivariate zero-inflated Poisson models and their applications. *Technometrics, 41*(1), pp.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), pp.345-351.

Lord, D., Persaud, B.N. (2000). Accident prediction models with and without trend: application of the generalized estimating equations procedure. *Transportation Research Record 1717*, pp.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), pp.35-46.

Ma, J. (2006). *Bayesian Multivariate Poisson-Lognormal Regression for Crash Prediction on Rural Two-Lane Highways*. Ph.D. Dissertation, The University of Texas at Austin.

Ma, J. and Kockelman, K.M. (2006). Crash modeling using clustered data from Washington State: Prediction of optimal speed limits. *Proceedings of the IEEE Intelligent Transportation Systems Conference*, Toronto Canada.

Ma, J., and Kockelman, K. (2006). Bayesian multivariate Poisson regression for models of injury count, by severity. *Transportation Research Record* 1950, pp.24-34.

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

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), pp.1129-1135.

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), pp.471-482.

Miaou, S.P. (1996). *Measuring the Goodness-of-fit of Accident Prediction Models.* Publication FHWA-RD-96-040. FHWA, U.S. DOT.

Miaou, S.P. (2001). *Estimating Roadside Encroachment Rates with the Combined Strengths of Accident- and Encroachment-Based Approaches.* Publication FHWA-RD-01-124. FHWA, U.S. DOT.

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*, pp.31-40.

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

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), pp.699-720.

Miaou, S.P., Hu, P.S., Wright, T., Davis, S.C., and Rathi, A.K. (1993). *Development of Relationship Between Truck Accidents and Geometric Design: Phase I.* Publication FHWA-RD-91-124. FHWA, U.S. DOT.

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), pp.33-57.

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), pp.253-270.

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

Noland, R.B., *and* Quddus, M.A. (2004). A spatially disaggregate analysis of road casualties in England. *Accident Analysis and Prevention 36*(6), pp.973-984.

Pawlovich, M.D., Li, W., Carriquiry, A. and Welch, T.M. (2006). Iowa's experience with road diet measures: Use of Bayesian approach to assess impacts on crash frequencies and crash rates. *Transportation Research Record 1953*, pp.163-171.

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, pp.209-215.

Pernia, J., Lu, J.J., and Peng, H. (2004). *Safety Issues Related to Two-way Left-turn Lanes*. Tampa, FL: Department of Civil and Environmental Engineering, University of South Florida.

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), pp.771-798.

Press, S.J. (1982). *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference 2nd Edition*. Robert E. Krieger Publishing Company, Malabar, Florida, USA.

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), pp.183-191.

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*, pp.95-102.

Schrank, D., and Lomax, T. (2002). *The 2002 Urban Mobility Report*. Texas Transportation Institute, The Texas A&M University System.

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), pp.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), pp.829-837.

Shankar, V.N., Albin, R.B., Milton, J.C. and Mannering, F.L. (1998). Evaluating median crossover likelihoods with clustered accident counts: An empirical inquiry using the random effects negative binomial model. *Transportation Research Record 1635*, pp.44-48.

Tsionas E.G. (2001). Bayesian multivariate Poisson regression. *Communications in Statistics – Theory and Methods 30*(2), pp.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*, pp.193-197.

Vogt, A. (1999). *Crash Models for Rural Intersections: Four-lane by Two-lane Stop-controlled and Two-lane by Two-lane Signalized.* Publication FHWA-RD-99-128. FHWA, U.S. DOT.

Vogt, A., and Bared, J.G. (1998). *Accident Models for Two-lane Rural Roads: Segments and Intersections.* Publication FHWA-RD-98-133. FHWA, U.S. DOT.

Washington, S. and Oh, J. (2006). Bayesian methodology incorporating expert judgment for ranking countermeasure effectiveness under uncertainty: Example applied to at grade railroad crossings in Korea. *Accident Analysis and Prevention, 38*(2), pp.234-247.

Winkelmann, R. (2000). Seemingly unrelated negative binomial regression. *Oxford Bulletin of Economics and Statistics 62*(4), pp.553-560.

Wooldridge, J.M. (1999). Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics 90*(1), pp.77-97.

Zegeer, C.V., Stewart, J.R., Huang, H.H., and Lagerwey, P.A. (2002). *Safety Effects of Marked Vs. Unmarked Crosswalks at Uncontrolled Locations: Executive Summary and Recommended Guidelines.* Publication FHWA-RD-01-075. FHWA, U.S. DOT.

List of Tables

Table 1 Summary Statistics of Variables

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 on each side (feet)	2.087	1.298	.00	16.50
Surface width (feet)[10]	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

[10] Surface width does not include the width of shoulders (paved or unpaved).

Table 2 PDO Crash Frequency MVPLN Model Results

Variable definition	Mean	Std. Err.	The 95% (2.5-97.5%) sample-based credible sets	
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) sample-based credible sets.

Table 3 Possible-Injury Crash Frequency MVPLN Model Results

Variable definition	Mean	Std. Err.	The 95% (2.5-97.5%) sample-based credible sets	
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) sample-based credible sets.

Table 4 Non-disabling Injury Crash Frequency MVPLN Model Results

Variable definition	Mean	Std. Err.	The 95% (2.5-97.5%) sample-based credible sets	
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) sample-based credible sets.

Table 5 Disabling Injury Crash Frequency MVPLN Model Results

Variable definition	Mean	Std. Err.	The 95% (2.5-97.5%) sample-based credible sets	
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) **sample-based credible sets**.

Table 6 Fatal Crash Frequency MVPLN Model Results

Variable definition	Mean	Std. Err.	The 95% (2.5-97.5%) sample-based credible sets	
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) **sample-based credible sets**.

Table 7 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%

Table 8 Correlation-Coefficients of $\vec{\varepsilon}_i$

	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

Table 9 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[11]	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[12]	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.

[11] The MVPLN1 predictions were computed as follows: (1) 1,000 samples of all severity-specific parameters were taken from a multivariate normal distribution with the posterior distribution's mean and correlation correlations; (2) 1,000 samples of nuisance parameters (error terms) were drawn from a multivariate normal with zero and correlation coefficients shown in Table 8; (3) expected crash counts for each segment were calculated, for all 1,000 samples.

[12] The MVPLN2 predictions were obtained as follows: (1) 7,000 samples of nuisance parameters (error terms) were drawn from a multivariate normal with zero mean and correlation coefficients shown in Table 8; (2) 7,000 expected crash counts were computed for all segments using these 7,000 draws along with the 7,000 draws from the MCMC simulation.