A SIMULTANEOUS EQUATIONS MODEL OF CRASH FREQUENCY BY SEVERITY LEVEL FOR FREEWAY SECTIONS

Xin Ye*
Department of Civil and Environmental Engineering
Arizona State University
Room ECG252
Tempe, AZ 85287-5306
Tel: (480) 965-2262; Fax: (480) 965-0557
Email: xin.ye@asu.edu

Ram M. Pendyala
Department of Civil and Environmental Engineering
Arizona State University
Room ECG252
Tempe, AZ 85287-5306
Tel: (480) 727-9164; Fax: (480) 965-0557
Email: ram.pendyala@asu.edu

Venky Shankar
Department of Civil and Environmental Engineering
The Pennsylvania State University
212 Sackett Building
University Park, PA 16802
Tel: (814) 865-9434; Fax: (814) 863-7304
Email: shankarv@engr.psu.edu

Karthik Konduri
Department of Civil and Environmental Engineering
Arizona State University
Room ECG252
Tempe, AZ 85287-5306
Tel: (480) 965-3589; Fax: (480) 965-0557
Email: karthik.konduri@asu.edu

* corresponding author

June 2008
ABSTRACT

This paper presents a simultaneous equations model of crash frequencies by severity level for freeway sections using five-year crash severity frequency data for 275 multilane freeway segments in the State of Washington. Crash severity is a subject of much interest in the context of freeway safety due to higher speeds of travel on freeways and the desire of transportation professionals to implement measures that could potentially reduce crash severity on such facilities. This paper presents a joint Poisson regression model with multivariate normal heterogeneities using the method of Maximum Simulated Likelihood Estimation (MSLE). MSLE serves as a computationally viable alternative to the Bayesian approach that has been adopted in the literature for estimating multivariate simultaneous equations models of crash frequencies. The empirical results presented in this paper suggest the presence of statistically significant error correlations across crash frequencies by severity level. The significant error correlations point to the presence of common unobserved factors related to driver behavior and roadway/traffic/environmental characteristics that influence crash frequencies of different severity levels. In addition, the empirical results show that observed factors generally do not have the same impact on crash frequencies at different levels of severity.

Keywords: crash severity, crash frequency, simultaneous equations model, freeway safety, maximum simulated likelihood estimation, multivariate poisson regression model
1. INTRODUCTION

Transportation professionals are constantly searching for ways to reduce both the occurrence of crashes and the severity of crashes on freeway sections. The severity of crashes on freeways is of much interest due to the higher speeds that could contribute to greater severity levels, conditional on the occurrence of a crash. This paper is aimed at making a contribution in the area of freeway safety modeling by presenting a model of freeway crash frequency by severity level for 275 freeway sections in the State of Washington. Five year freeway section crash counts (by severity level) are available for the 275 freeway sections and a modeling methodology that can simultaneously account for observed and unobserved factors contributing to crash frequencies by severity level is developed and presented in this paper.

Modeling crash frequencies by type of crash (angle, head-on, rear-end, etc.), number of vehicles involved (single-vehicle, two-vehicle, multi-vehicle, etc.), and severity level (property damage only, possible injury, incapacitating injury, etc.) has been the subject of much research in the transportation safety arena. Researchers have employed a variety of single equation count data models including Poisson models, Negative Binomial models, and zero-inflated versions of these count models (to account for the presence of zero crash frequency counts in the data set that may be due to facilities being truly safe or simply due to the limited window of observation for which crash frequency data is collected) to model crash frequency. These single equation models have often been employed to identify factors contributing to total crash frequencies on highway facilities. These methods can also be employed to model crash frequencies by severity level (e.g., modeling the number of fatal crashes as a function of roadway/traffic characteristics, environmental conditions, etc.).

Although modeling crash frequency by severity level using single-equation methods (sets of independent equations) can offer valuable insights into factors affecting crash frequencies, the
fact that such model systems ignore the simultaneity that may be prevalent in the safety phenomenon under investigation is an issue that merits being addressed. There may be a host of unobserved factors related to driver characteristics, vehicular characteristics, and roadway/traffic characteristics that contribute to crash frequencies at various severity levels. In a single equation method, the random error component (and the constant term) may be viewed as capturing the effects of these unobserved factors. However, simultaneity may arise in crash frequency modeling due to the possibility that unobserved factors affecting crash frequency at one severity level may be correlated significantly with unobserved factors affecting crash frequency at another severity level. This possibility calls for the deployment of simultaneous equations modeling methodologies to effectively model crash frequencies at multiple severity levels. This paper aims to make a contribution in this area by presenting a multivariate count data model that is capable of accounting for correlated unobserved factors across equations representing crash frequencies at different severity levels. The correlation is accommodated by allowing for the presence of error covariances across equations, thus contributing to the simultaneity in the phenomenon under investigation.

The need for modeling crash frequencies by severity level in a simultaneous equations framework has been recognized; however, the analytical and computational complexity associated with formulating and estimating such systems has hindered the development of these model systems, particularly in the count model (data) context. This paper makes a methodological contribution by formulating and presenting an n-dimensional multivariate count data model (Poisson regression) that accounts for error correlations through the incorporation of normally distributed heterogeneity terms (more on this in the methodology section of this paper). Model estimation is achieved through the use of maximum simulated likelihood estimation (MSLE) methods that provide robust parameter estimates and test statistics for hypotheses testing. From an
empirical standpoint, the paper makes a contribution to the understanding of crash frequencies at various severity levels for freeways. By presenting a simultaneous equations model system for freeway crash severities, the paper provides key insights into the factors that impact crash frequencies at various severity levels while accounting for the presence of error covariances (common unobserved factors).

Following a brief review of the literature, the paper presents the modeling methodology adopted in this paper. This is followed by a description of the dataset. Model estimation results and key conclusions are presented in the final two sections of the paper.

2. MODELING CRASH FREQUENCY AND SEVERITY

There is a vast body of literature devoted to modeling crash severity outcomes as a function of crash type, driver characteristics, roadway/traffic characteristics, and environmental conditions. These papers have used a variety of discrete choice modeling approaches, most notably the ordered probit and multinomial logit modeling approaches, to model crash severity outcomes. Examples of ordered probit models of injury severity include Quddus et al (2002), Kockelman and Kweon (2002), Zajac and Ivan (2003), and Ma and Kockelman (2006a). Examples of ordered logit models of injury severity include those by Chang and Mannering (1999), Shankar et al (1996), Shankar and Mannering (1996), Khorashadi et al (2005), and Milton et al (2007). The last paper by Milton et al (2007) deploys a mixed logit modeling methodology to account for variations in the effects of explanatory factors on injury severity outcomes. Other methods to estimate injury severity include the logistic regression model (e.g., Ossenbruggen et al, 2001) and Artificial Neural Networks (e.g., Abdelwahab and Abdel-Aty, 2002; Abdel-Aty and Abdelwahab, 2004).
While such models are very useful to model crash severity outcomes at the level of the individual crash, they do not offer insights into crash frequencies. Due to the non-negative nature of crash frequency data, count models have been used extensively to estimate crash frequency. Ivan et al (2000) presented the use of Poisson regression models to estimate single and multi-vehicle crash rates as a function of traffic density, land use, ambient light conditions and time of day. If there are excessive zeros among crash frequencies, it is considered that some of the zeros in the count data are generated by a process that is different from the rest of the counts. This has led to the use of Zero Inflated Regression Models (Lee and Mannering 2002; Lord et al 2005) in place of the traditional Poisson and Negative Binomial regression models. Crash frequency models using a variety of count data modeling approaches have also been presented by Ma and Kockelman (2006a), Milton and Mannering (1998), Shankar et al (1997), Miaou (1994), and Miaou et al (1992). This paper is intended to add to this body of literature by simultaneously modeling crash frequencies by severity for freeway sections by formulating and estimating a multivariate count data model.

The development and application of multivariate frequency models has been attempted in the field of transportation before. Zhao and Kockelman (2001) developed and applied a Multivariate Negative Binomial Regression model to analyze household vehicle ownership by vehicle type. In their model formulation, the same gamma error term is assumed for different types of auto ownership. In that paper, they also present and formulate a univariate Poisson regression model with normal heterogeneity, but they note that convergence was not achieved when model estimation was attempted using simulated maximum likelihood estimation methods. After that initial attempt, Ma and Kockelman (2006b) and Ma et al (2007) continued efforts to develop and estimate Poisson regression model formulations with multivariate normal heterogeneities. In these two papers, the authors adopted a Bayesian approach to estimate the
parameters of the Multivariate Poisson (MVP) Regression Model for jointly modeling crash frequencies at different severity levels (similar to the context of this paper). The analytical and computational complexity associated with using maximum likelihood estimation methods for model estimation is noted in their papers. However, the authors indicate that the maximum simulated likelihood estimation method can be applied for MVP model estimation if a special variance-covariance structure is specified for the error terms.

A significant contribution towards advancing the development of multivariate Poisson regression models for simultaneously modeling crash frequencies, similar to that adopted by Ma et al (2007), is that by Park and Lord (2007). They developed a multivariate Poisson regression – lognormal model for modeling crash frequencies by severity level at intersections using a Bayesian estimation approach. This paper constitutes an attempt to further contribute to this area by presenting a modeling methodology whereby one can estimate the MVP model using simulation-based maximum likelihood estimation methods while accommodating unobserved heterogeneity (overdispersion) and flexible error covariance structures similar to the work by Ma et al (2007) and Park and Lord (2007). Unlike their estimation approach, however, the simulation-based maximum likelihood estimation method employed in this paper uses Halton sequence draws to accurately compute the log-likelihood function (evaluate multidimensional integrals of the Poisson distribution). This method has been used extensively in the travel behavior arena to model a variety of travel behavior choices while incorporating random taste variations (see Bhat, 2003 for a detailed description of this method). Also, the paper by Park and Lord (2007) focuses on crash severity frequencies for intersections while this paper focuses on modeling crash severity frequencies for freeway sections.
3. MODELING METHODOLOGY

This section presents the modeling methodology adopted in this paper. First, the univariate Poisson regression model is presented using two different heterogeneity specifications – the log-gamma heterogeneity (analogous to the Negative Binomial regression model) and normal heterogeneity. Then, the discussion is extended to present the formulation for the multivariate poisson regression model with normal heterogeneity.

3.1. Negative Binomial (NB) Regression Model: Univariate Poisson Regression Model with Log-gamma Heterogeneity

Count data models are most suited to modeling any dependent variable \( y_i \) that constitutes a frequency or “count”. The dependent variable can only take non-negative integer values. In this paper, \( y_i \) represents crash frequency by severity (but the subscript representing severity level is suppressed without loss of generality) for road section \( i \). The expectation of \( y_i \) is assumed to be \( \lambda_i \) and the count data model formulation is as follows:

\[
\ln(\lambda_i) = x_i \beta + \varepsilon_i, \tag{1}
\]

where \( x_i \) is a vector of explanatory variables indicating characteristics for road section \( i \); \( \beta \) is a vector of coefficients associated with \( x_i \). \( \varepsilon_i \) is a random variable representing heterogeneity that accounts for unobserved factors and other random disturbances.

Since \( y_i \) constitutes count data, the probability of \( y_i \) conditional on \( \varepsilon_i \) is given as:

\[
Pr(y_i | \varepsilon_i) = \frac{\exp(-\lambda_i) \lambda_i^{y_i}}{y_i!}, \tag{2}
\]

The Negative Binomial (NB) regression model is formulated based on the assumption that \( \exp(\varepsilon_i) = t_i \sim \Gamma (1/\alpha^2, \alpha^2) \). The corresponding probability density function is:

\[
f(t_i) = \frac{t_i^{1/\alpha^2-1}}{\left(\alpha^2\right)^{1/\alpha^2} \Gamma(1/\alpha^2)} \exp\left(-\frac{t_i}{\alpha^2}\right), t_i > 0, \tag{3}
\]
where $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$.  

(4)

The expectation and standard deviation of $t$ are equal to 1 and $\alpha$, respectively. By integrating $t_i$ over its distributional domain, one may obtain the unconditional probability of $y_i$ as:

$$
\Pr(y_i) = \int_{-\infty}^{\infty} \Pr(y_i \mid t_i)f(t_i)dt_i = \frac{\Gamma(1/\alpha^2 + y_i)}{\Gamma(1 + y_i)\Gamma(1/\alpha^2)} r_i^{-y_i} (1 - r_i)^{1/\alpha^2}, 
$$

(5)

where $r_i = \frac{\alpha^2 \exp(x_i \beta)}{\alpha^2 \exp(x_i \beta) + 1}$.  

(6)

Cameron and Trivedi (1986) proposed this unconditional probability function with a closed form solution. This formulation has allowed the NB model to be widely applied for modeling count data in many different areas, including transportation.

It is to be noted that the true heterogeneity in the model is not $t_i$, but $\varepsilon_i$, which accounts for the presence of unobserved variables or factors excluded from the vector $x_i$. Since $\varepsilon_i$ is equal to $\ln(t_i)$, the underlying distributional assumption on $\varepsilon_i$ is the log-gamma distribution and the probability density function can be derived as:

$$
f(\varepsilon_i) = \frac{1}{\Gamma(1/\alpha^2)} \exp\left\{\frac{1}{\alpha^2} [\varepsilon_i - \ln(\alpha^2)] - e^{[\varepsilon_i - \ln(\alpha^2)]}\right\}, \ -\infty < \varepsilon_i < +\infty. 
$$

(7)

Its expectation and standard deviation can be formulated as:

$$
E(\varepsilon_i) = \ln(\alpha^2) + g'(1/\alpha^2) 
$$

(8)

$$
Var(\varepsilon_i) = g''(1/\alpha^2) 
$$

(9)

where $g(x) = \ln[\Gamma(x)]$ 

(10)

$$
g'(x) = \frac{d[\ln[\Gamma(x)]]}{dx} 
$$

(11)

$$
g''(x) = \frac{d^2[\ln[\Gamma(x)]]}{dx^2} 
$$

(12)
Based on Equations (8) and (9), one can develop a plot similar to that shown in Figure 1 to illustrate the variation in expectation and standard deviation of log-gamma heterogeneity as the value of $\alpha$ goes from 0.1 to 1.5. It can be seen that the expectation and standard deviation of log-gamma heterogeneity are inversely related to one another. As the standard deviation of heterogeneity increases, the expectation becomes increasingly negative. This results in the distribution of the log-gamma heterogeneity to be negatively skewed. It is undesirable to have this feature because there is no reason to believe that the random component that accounts for unobserved factors systematically takes more negative values. In vector $x_i$, it is possible to omit or not observe variables with either positive or negative expectation and that take either positive or negative coefficients. In other words, there is no reason to believe that the distribution of $\varepsilon_i$ is skewed in any particular direction. If such a skew is imposed, the NB regression model will tend to incorrectly estimate the variance of the dependent variable. The use of the log-gamma heterogeneity specification may result in imposing such a skew to the random error component and this constitutes a shortcoming of the use of this distributional assumption.

3.2. Univariate Poisson Regression (UVP) Model with Normal Heterogeneity

In view of the shortcoming of using the log-gamma heterogeneity specification in the NB regression model, one may choose to use a normal distribution for representing heterogeneity $\varepsilon_i$. If $\varepsilon_i$ is normally distributed with 0 expectation and standard deviation of $\sigma$, the probability density function is:

$$f(\varepsilon_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{\varepsilon_i^2}{2\sigma^2} \right), -\infty < \varepsilon_i < +\infty.$$ (13)

Unlike the log-gamma distribution, the normal distribution is symmetric and its expectation can be fixed at 0 regardless of the standard deviation of the distribution. Under the
assumption of normality, one can integrate $\varepsilon_i$ over its distributional domain and obtain the unconditional probability of $y_i$ as:

$$\Pr(y_i) = \int_{-\infty}^{\infty} \Pr(y_i | \varepsilon_i) f(\varepsilon_i) d\varepsilon_i$$

$$= \int_{-\infty}^{\infty} \exp[-\exp(x_i\beta + \varepsilon_i)][\exp(x_i\beta + \varepsilon_i)]^{y_i} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\varepsilon_i^2}{2\sigma^2}\right) d\varepsilon_i$$  \hspace{1cm} (14)

The key difference in comparison to the NB regression model presented in Equations (5) and (6) is that the unconditional probability function presented in Equation (14) does not have a closed form solution. One may approximate the unconditional probability by using simulation-based approaches as follows:

$$SP(y_i) = \frac{1}{R} \sum_{r=1}^{R} \exp[-\exp(x_i\beta + \sigma u_{ir})][\exp(x_i\beta + \sigma u_{ir})]^{y_i} \frac{1}{y_i!}$$  \hspace{1cm} (15)

where $SP$ represents the simulated probability function and $u_{ir}$ are random seeds drawn from a standard normal distribution, which can be converted to normal random seeds with standard deviation $\sigma$ by multiplying them with a single factor $\sigma$. The Maximum Simulated Likelihood Estimation (MSLE) method can then be applied to estimate unknown parameters $\beta$ and $\sigma$ with the aid of quasi-random seeds (Bhat, 2003). Ordinary Least Square (OLS) estimators may be used as starting values of $\beta$. As for the initial value of $\sigma$, one may use the standard deviation corresponding to the $\alpha$ value estimated from the NB regression model (see Figure 1) since they both represent the standard deviation of the same heterogeneity term.

3.3. N-Dimensional Multivariate Poisson (MVP) Regression Model

The greatest benefit of using a normal distribution to represent heterogeneity is that one can easily realize an n-dimensional Multivariate Poisson (MVP) regression model, where $n (n \geq 2)$ dependent (count) variables can be jointly modeled. The correlation among the dependent variables can be naturally accommodated into the correlation between their heterogeneities, which turns out to be a
multivariate normal distribution in the case of a simultaneous equations system. The logarithmic expectations for count dependent variables in an n-dimensional Multivariate Poisson Regression Model may be formulated as:

\[
\begin{align*}
\ln(\lambda_{i1}) &= x_{i1}\beta_1 + \epsilon_{i1} = x_{i1}\beta_1 + f_1u_{i1} \\
\ln(\lambda_{i2}) &= x_{i2}\beta_2 + \epsilon_{i2} = x_{i2}\beta_2 + f_2u_{i2} + f_3u_{2i} \\
\ln(\lambda_{i3}) &= x_{i3}\beta_3 + \epsilon_{i3} = x_{i3}\beta_3 + f_4u_{i1} + f_5u_{2i} + f_6u_{3i} \\
&\ldots \\
\ln(\lambda_{ni}) &= x_{ni}\beta_n + \epsilon_{ni} = x_{ni}\beta_n + f_{n(n-1)/2+1}u_{i1} + f_{n(n-1)/2+2}u_{2i} + \ldots \\
&\quad + f_{n(n+1)/2-n}u_{(n-1)i} + f_{n(n+1)/2}u_{ni}
\end{align*}
\]

where \(n\) is the number of dependent (count) variables and \(u_{i1}, u_{2i}, \ldots, u_{ni}\) are \(n\) independent standard normal random variables. \(f_1, f_2, \ldots, f_{n(n+1)/2}\) are parameters that allow the realization of a multivariate normal error structure for the system of equations. The total number of parameter \(f_i\) to be estimated is \(n(n+1)/2\), which is exactly equal to the number of independent elements in the variance-covariance matrix of the error structure. Therefore, all \(f_i\) are uniquely identified. Given all of the \(f_i\) values, error variance-covariance and correlation matrices can be calculated in a straightforward manner. For example, \(\text{Cov}(\epsilon_2, \epsilon_3) = f_2f_4 + f_3f_5\). Conversely, given error variance-covariance or correlation matrices, all of the \(f_i\) values can be obtained by sequentially solving a system of simultaneous equations.

In this particular paper, crash frequencies are modeled jointly for three different severity levels. They are: Property Damage Only, Possible Injury, and Injury/Fatality. Thus, we have three dependent count variables (\(n = 3\)) and the logarithms of expectations \(\lambda_{i1}, \lambda_{i2}, \text{ and } \lambda_{i3}\) for the three count variables are formulated as:

\[
\begin{align*}
\ln(\lambda_{i1}) &= x_{i1}\beta_1 + \epsilon_{i1} = x_{i1}\beta_1 + f_1u_{i1} \\
\ln(\lambda_{i2}) &= x_{i2}\beta_2 + \epsilon_{i2} = x_{i2}\beta_2 + f_2u_{i2} + f_3u_{2i} \\
\ln(\lambda_{i3}) &= x_{i3}\beta_3 + \epsilon_{i3} = x_{i3}\beta_3 + f_4u_{i1} + f_5u_{2i} + f_6u_{3i}
\end{align*}
\]
where $u_{1i}$, $u_{2i}$ and $u_{3i}$ are three independent random variables, which are standard normally distributed and $f_i$ are coefficients to be estimated; $x_i$ are vectors of explanatory variables and $\beta_i$ are associated coefficient vectors. It is preferable to use $(f_1 u_{1i})$, $(f_2 u_{1i} + f_3 u_{2i})$, and $(f_4 u_{1i} + f_5 u_{2i} + f_6 u_{3i})$ to represent trivariate normally distributed heterogeneities $\varepsilon_{1i}, \varepsilon_{2i}$ and $\varepsilon_{3i}$, because this specification avoids difficulties associated with implementing the Cholesky method.

The Cholesky method is usually employed for generating multivariate normal random numbers by calculating $e_{N \times n} A_{n \times n}$, where $e_{N \times n}$ is a matrix with an $n$-column vector of independent univariate normal random seeds and $A_{n \times n}$ is the generalized square root of the variance-covariance matrix $\Sigma$ of the multivariate normal distribution (i.e. $AA^T = \Sigma$). Instead of the Cholesky method, a different simulation method is adopted in the current MVP model. As the equations for the error terms appear triangular (see Equations 16 and 17), the method is referred to as the “triangular simulation” method for brevity.

The probability functions conditional on multivariate normal heterogeneities are given as:

\[
\begin{align*}
\Pr(y_{1i} | u_{1i}) &= \frac{\exp(-\lambda_{1i}) \lambda_{1i}^{y_{1i}}}{y_{1i}!} \\
\Pr(y_{2i} | u_{1i}, u_{2i}) &= \frac{\exp(-\lambda_{2i}) \lambda_{2i}^{y_{2i}}}{y_{2i}!} \\
\Pr(y_{3i} | u_{1i}, u_{2i}, u_{3i}) &= \frac{\exp(-\lambda_{3i}) \lambda_{3i}^{y_{3i}}}{y_{3i}!}
\end{align*}
\] (18)

Then, the unconditional probability can be obtained by integrating conditional probability functions over the distributional domain of random heterogeneities:

\[
\begin{align*}
\Pr(y_{1i}, y_{2i}, y_{3i}) &= \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\Pr(y_{1i} | u_{1i}) \Pr(y_{2i} | u_{1i}, u_{2i}) \Pr(y_{3i} | u_{1i}, u_{2i}, u_{3i})]d\Phi(u_{1i})d\Phi(u_{2i})d\Phi(u_{3i})
\end{align*}
\] (19)
Similar to the UVP-normal heterogeneity model, this expression does not have a neat closed-form solution. The unconditional probability function can be approximated by the simulated probability function given below:

$$SP(y_{1i}, y_{2i}, y_{3i}) = \frac{1}{R} \sum_{r=1}^{R} \Pr(y_{1i} \mid u_{1ir}) \Pr(y_{2i} \mid u_{1ir}, u_{2ir}) \Pr(y_{3i} \mid u_{1ir}, u_{2ir}, u_{3ir})$$

(20)

where $u_{1ir}, u_{2ir},$ and $u_{3ir}$ are three independent random seeds drawn from a standard normal distribution. These terms can be transformed as $(f_1u_{1i}), (f_2u_{1i} + f_3u_{2i}),$ and $(f_4u_{1i} + f_5u_{2i} + f_6u_{3i})$ for realizing the trivariate normal heterogeneity that may be prevalent in the simultaneous equations system of crash severity frequencies. Variance, covariance, standard deviation, and correlation terms associated with the error structure can be calculated based on parameter $f_i$ as:

$$\text{Var}(\varepsilon_1) = f_1^2, \text{Var}(\varepsilon_2) = f_2^2 + f_3^2 \text{ and } \text{Var}(\varepsilon_3) = f_4^2 + f_5^2 + f_6^2$$

(21)

$$\text{Cov}(\varepsilon_1, \varepsilon_2) = f_1f_2, \text{Cov}(\varepsilon_1, \varepsilon_3) = f_1f_4 \text{ and } \text{Cov}(\varepsilon_2, \varepsilon_3) = f_2f_4 + f_3f_5$$

(22)

$$\text{Std Dev}(\varepsilon_1) = f_1, \text{Std Dev}(\varepsilon_2) = \sqrt{f_2^2 + f_3^2} \text{ and } \text{Std Dev}(\varepsilon_3) = \sqrt{f_4^2 + f_5^2 + f_6^2}$$

(23)

$$\text{Corr}(\varepsilon_1, \varepsilon_2) = \frac{f_1f_2}{\sqrt{f_1^2(f_2^2 + f_3^2)}}, \text{Corr}(\varepsilon_1, \varepsilon_3) = \frac{f_1f_4}{\sqrt{f_1^2(f_4^2 + f_5^2 + f_6^2)}}$$

(24)

and

$$\text{Corr}(\varepsilon_2, \varepsilon_3) = \frac{f_2f_4 + f_3f_5}{\sqrt{(f_2^2 + f_3^2)(f_4^2 + f_5^2 + f_6^2)}}.$$

As in the case of the univariate model with normal heterogeneity, MSLE method can be applied to estimate unknown parameter $\beta$ and $f_i$ using reasonable starting values. Ordinary Least Squares (OLS) estimators may be used as starting values for $\beta$. The overall standard deviation of error terms $\varepsilon_1, \varepsilon_2$ and $\varepsilon_3$ can be approximated from the NB or UVP models. Then, these estimated standard deviations are taken as starting values for the $f_i$ terms located on the “hypotenuse” of the “right triangle” (i.e. $f_1, f_3$ and $f_5$ as shown in Equation 17). The other $f_i$ (i.e. $f_2, f_4$ and $f_6$) terms start
at zero. 250 random seeds are drawn from the Halton sequence (Bhat, 2001) to accomplish accurate approximation of the log-likelihood function.

The log-likelihood function and its first-order derivative were coded in Gauss (Aptech, 2006) and the default BFGS algorithm provided by the Maxlik module in Gauss was used for maximizing the log-likelihood function. An attempt was made to use the Cholesky method to generate trivariate normal seeds in the MSLE procedure; however, convergence was not achieved using appropriate initial values. It appears that the optimization procedure employing the Cholesky method does not succeed in the absence of highly accurate starting values. In early stages of this research, when the Gauss code was tested using synthetic data sets, it was found that the optimization procedure frequently terminated without achieving convergence because the requirement of the Cholesky method that the variance-covariance matrix be positive-definite was not satisfied during the iterative process. The proposed “triangular simulation” method does not require a positive-definite variance-covariance matrix for calculating the matrix square root and thereby obviates the difficulties of using the Cholesky method.

However, one shortcoming of the triangular simulation method is that the variance-covariance matrix of the error structure needs to be calculated based on more than one estimated parameter. Therefore, it is not straightforward to draw statistical inferences on the elements of the variance-covariance matrix (as they are functions of estimated parameters, $f_i$). To overcome this shortcoming, a simulation-based hypothesis test is applied to approximate statistical inferences for these calculated elements of interest. The idea is that all of the parameter estimators obtained through Maximum Likelihood Estimation (MLE) procedures are asymptotically multivariate normally distributed and their expectations are nothing but the estimated parameters themselves and the variance-covariance matrix is the negative inverse of the Hessian matrix at convergence. To implement the simulation-based hypothesis test, a few columns of random variables that are
subject to the multivariate normal distribution of all the estimators are drawn. Then, the relevant columns of random variables can be selected for calculating the elements of interest and the significance level can be approximated by counting the positive or negative numbers among the calculated elements. For example, suppose $\text{Cov}(\varepsilon_1, \varepsilon_3)$ is of interest. Note that $\text{Cov}(\varepsilon_1, \varepsilon_3) = f_1 f_4$. Then, one needs to pick out two columns of random variables corresponding to $f_1$ and $f_4$ and calculate the product $f_1 f_4$ for each pair of random variables to generate a column of “randomized” $\text{Cov}(\varepsilon_1, \varepsilon_3)$. The percent of positive or negative counts among the random seeds for $\text{Cov}(\varepsilon_1, \varepsilon_3)$ can approximate the significance level of $\text{Cov}(\varepsilon_1, \varepsilon_3)$, i.e., the probability that $\text{Cov}(\varepsilon_1, \varepsilon_3)$ is positive or negative. In addition, one may approximate the expectation and standard deviation of the variances and covariances based on the mean value and standard deviation of these random elements. However, it should be noted that the elements of the variance-covariance matrix are no longer normally distributed and the traditional $t$-statistic, calculated by dividing the mean by the standard deviation, cannot be used for statistical inference.

4. DATA DESCRIPTION

The data set used in this paper is derived from safety (crash frequency) data available for multilane divided highways in the State of Washington. These highways, which are part of the National Highway System, are considered critical routes because of their high economic importance—they are also known for their high speed of travel, significant traffic volumes, and congestion. To obtain roadway segments of sufficient length to allow for use in Washington State safety programming applications, segments along this multilane system are defined by median treatments (safety barriers, cables or landform barriers). A roadway segment’s beginning point was identified where a previous run of a barrier terminated (or began) and ended where the next run of a barrier was encountered (or the current run ended). In all, the data consist of 275 roadway segments of
varying lengths with a mean segment length of roughly 2.4 miles with a standard deviation of about 2.7 miles. Historical crash data were gathered for the 1990–1994 timeframe. For each roadway segment, crashes were sorted by year, and individual crash data reports on the roadway segments were aggregated based on the most severe person-injury in the crash. Thus, crash frequency counts by severity level were obtained for each freeway segment.

The crash data were combined with weather data from the Western Regional Climate Center which included total precipitation (all forms) and snowfall precipitation. These weather data were observed using permanent weather stations and were assigned based on proximity of the station to a roadway segment. Data from the Washington State Department of Transportation databases were used for geometric, pavement, roadside and traffic characteristics associated with roadway segments. Geometric data included number of lanes, width of lanes, shoulder widths, median width, minimum and maximum radii of horizontal curves, central angle of horizontal curves, grade, minimum grade, maximum grade, grade differential, tangent length, number of changes in grade, number of horizontal and vertical curves per mile, presence of interchanges and presence of exit/entrance ramps. Pavement data included roadway pavement type, shoulder pavement type and friction coefficients. Roadside information included slopes, presence of vegetation, ditch information, the presence of crossovers, and information on other fixed objects (such as trees and poles). Traffic operations data included speed limit, average annual daily traffic, average daily traffic per lane, single-unit truck traffic, combination truck traffic, large truck percentages, peak-hour factors and roadway access control.

Information on a total of 22,619 individual crashes was included in this study. Due to the limited number of crashes that resulted in disabling injury and fatality, it was not statistically possible to estimate all five severity-level categories (in other words, it was not possible to statistically differentiate among all five severity categories). Therefore, three severity categories
are considered: property damage only, possible injury, and injury/fatality (with the third category encompassing evident injury, disabling injury, and fatality). With this definition, of the 22,568 individual crashes reported over the 5-year study period, 8,367 resulted in property damage only, 6,988 in possible injury, and 7,264 in injury/fatality as the most severe outcome of the crash. Table 1 provides information on the mean, standard deviation, minimum and maximum of selected variables in the data set.

5. MODEL ESTIMATION RESULTS

This section presents model estimation results for various model forms considered in this paper. First, results are presented for univariate Poisson and Negative Binomial models of total crash frequency (sum of property damage only, possibly injury, and injury/fatality crashes). Then, results are presented for the trivariate Poisson (TVP) regression model of crash frequency for the three severity levels under consideration.

5.1. Estimation Results of UVP and NB models for Total Crash Frequency

The univariate Poisson regression model (UVP with normal heterogeneity) and the Negative Binomial (NB) regression model of total crash frequency are presented on the right hand side of Table 2. The NB and UVP model are estimated using Gauss (Aptech, 2006) employing the Newton-Raphson optimization algorithm (the BFGS algorithm also yields the same solution, but requires more iterations to achieve convergence). As mentioned earlier, 250 random seeds are drawn from the Halton sequence to accurately compute the log-likelihood function.

The value of $\alpha$ is estimated to be 0.6986 in the NB model. This parameter is statistically significant as evidenced by the large t-statistic. The expectation and standard deviation of heterogeneity ($\varepsilon_i$) in the NB model may then be calculated as -0.26 and 0.79 according to Equations (8) and (9). In the UVP model, the estimate of the standard deviation of heterogeneity
(\varepsilon_i) is 0.72 and the expectation is 0. The finding that the estimates of standard deviations of the heterogeneity terms are similar between the UVP and NB models is consistent with expectations as they essentially attempt to capture the same random heterogeneity. To illustrate the difference in heterogeneities between the NB and UVP model, the two distributions are plotted using the estimated parameters. This plot is shown in Figure 2. One plot corresponds to the log-gamma distribution with mean -0.26 and standard deviation 0.79 while the other corresponds to the normal distribution with mean 0 and standard deviation 0.72. It can be seen that the two distributions are very similar to one another and both have a peak at around zero. One obvious difference is that the normal distribution is symmetric and the log-gamma distribution is asymmetric with a slight negative skew. This finding supports the use of the normal distribution to represent heterogeneity in count data models; it approximates very well the traditional log-gamma distribution that is inherently assumed in classical NB models.

It is found that the constant term and a few coefficients associated with explanatory variables in the UVP model are different in magnitude than those in the NB model. This does not necessarily mean that their effects on total crash frequency are substantially different, however. Note that the marginal effect of explanatory variable \(x_i\) on dependent variable \(y\) can be expressed as:

\[
\partial E(y|x)/\partial x_i = \partial E[\exp(x\beta + \varepsilon)]/\partial x_i = \partial \{E[\exp(\varepsilon)]\exp(x\beta)\}/\partial x_i
\]  (25)

In the NB model, \(\varepsilon\) is log-gamma distributed and \(\exp(\varepsilon)\) is gamma distributed with expectation equal to 1. Thus, the marginal effects in the NB model can be derived as:

\[
\partial E(y|x)/\partial x_i = \beta_i \exp(x\beta)
\]  (26)
However, in the UVP model, \( \varepsilon \) is normally distributed and \( \exp(\varepsilon) \) is log-normally distributed with expectation \( \exp(\sigma^2/2) \). Then, the marginal effects in the UVP model can be derived to be:

\[
\frac{\partial E(y|x)}{\partial x_i} = \exp(\sigma^2/2) \beta_i \exp(x\beta)
\]

(27)

As \( \exp(\sigma^2/2) \) must be greater than 1, obtaining the same parameter estimates in the NB and UVP models would result in the marginal effects in the UVP model being scaled up by \( \exp(\sigma^2/2) \). For example, in this study, \( \exp(\sigma^2/2) \) is found to be 1.2931; this implies that the marginal effects in the UVP model would be scaled up by 29.31% relative to the NB model if both models take equal coefficients. In order to neutralize this effect and yield marginal effects that are comparable, the UVP model estimates of \( \beta \) coefficients should be smaller than those in the NB model. However, this is not universally true, suggesting that the two models do offer different marginal effects for several variables included in the model. As for the sign of the coefficients, both UVP and NB models yield coefficients with the same sign. It is also noted that the goodness-of-fit measure for the UVP model is slightly better than that of the NB model (0.7036 versus 0.7026), indicating that the normal distribution is potentially more appropriate for specifying random heterogeneity in the data compared to a log-gamma distribution.

As the focus of this research effort is on explaining crash frequencies by severity level, the total crash frequency models are not discussed here in detail. A more detailed discussion of the empirical results is presented next in the context of the trivariate Poisson regression model with multivariate normal heterogeneity.

5.2. Estimation Results of TVP Model for Crash Frequencies at Three Severity Levels

Estimation results for the trivariate Poisson (TVP) regression model with multivariate normal heterogeneity are presented in Table 3. The estimation results provide parameter estimates for
three simultaneous equations included in the model system, one equation for each severity level considered. Table 4 presents the expectation and standard deviation of elements in the error variance-covariance matrix and the error correlation matrix approximated by drawing 100,000 pseudo-random seeds. Among the random seeds drawn for covariances and correlations, no negative seeds are observed. This result indicates that all correlations between each pair of error terms are highly statistically significant (p-value < 0.00001). This finding of significant error correlations will be discussed in greater detail later in this section. Similar model estimation results were obtained when $f_2$, $f_4$ and $f_5$ in the error structure are fixed at zero as shown in the left block of Table 2. In this case, the TVP model reduces to three recursive UVP models which ignore error covariances. Relative to the TVP model, the UVP model exhibits a poorer data fit as reflected in the goodness-of-fit measure (0.5547 vs. 0.5597).

After completing model estimation using the triangular simulation method, an attempt was made to estimate the model using the Cholesky method. Values of $\hat{\beta}$ obtained using the triangular simulation method were used as starting values for $\beta$ and calculated variance and covariance values were used as starting value for parameters of the variance-covariance matrix. Using these superior starting values, the model estimation procedure did indeed converge when using the Cholesky method. This offered an opportunity to examine the performance of the triangular simulation method with regard to overall estimation results, simulated statistical inferences, and associated significance levels.

The last column of Table 4 lists estimation results and statistical inferences obtained using the Cholesky method for variance and covariance values that are specified as single parameters. It is found that the estimates of mean value and standard deviation of elements of the error variance-covariance matrix obtained using the triangular simulation method introduced in this paper are very close to those obtained using the Cholesky method. This comparison validates the estimates
of elements of the error variance-covariance matrix and those of their simulated standard deviations obtained using estimators based on the triangular simulation method. Estimators of $\beta$ are virtually identical between those obtained using the triangular and Cholesky methods; hence, those obtained using the latter method are not presented in the paper. In conclusion, the triangular simulation method in combination with a simulation-based hypothesis test appears to be a robust substitute for the Cholesky method in implementing MSLE procedures in the context of multivariate normal error structures.

The models of crash severity frequencies used virtually the same set of variables as those used in the univariate models of total crash frequency. A few variables turned out to be only marginally significant in some instances, but were retained in light of the intuitively reasonable coefficient value they offered. The values of coefficients are generally found to be quite different across the three severity frequency equations. Such differences cannot be attributed to differences in the values of $\exp(\sigma^2/2)$ across the three equations because Table 4 shows that variance estimates for heterogeneity terms are similar to one another. For example, the variable representing the maximum central angle in the section takes a coefficient value in the model for injury/fatality frequency that is twice that in the possible injury frequency model and the property damage only frequency model. This suggests that the same variable may have differential impacts on the occurrence of crashes of different severity. Thus, the maximum central angle has a greater impact on injury/fatal crash frequency than on property damage only crash frequency. In addition to differential impacts, opposing impacts of variables on crash frequencies at different severity levels are captured in the model. For example, the variable indicating the number of median crossovers has a negative impact on injury/fatal crash frequencies, but a positive impact (reversal of sign) on property damage only and possible injury crash frequencies (although the coefficient for property damage only frequencies is not statistically significant). The univariate models of total crash
frequency are not able to differentiate such opposing effects of an explanatory variable on frequencies of crashes at different levels of severity.

Virtually all other explanatory variables have coefficient estimates with intuitively reasonable values and signs. The logarithm of AADT and the length of freeway section, both of which may be considered to serve as exposure measures, significantly contribute to crash severity frequencies (for all severity levels). The maximum grade and maximum central angle in the section, representing adverse geometric characteristics of the roadway section, also have positive coefficients indicating that they contribute to crash severity frequencies for all three severity levels. As expected, friction factor shows a negative coefficient suggesting that higher friction factors are associated with lower crash frequencies for all severity levels. This is presumably due to the better braking ability that such roadway sections provide.

The number of interchanges and over-crossings in the section appear with statistically significant positive coefficients in all severity frequency models. All of these factors presumably contribute to additional traffic conflicts and driving maneuvers that, in turn, lead to more crashes at all severity levels. The number of ramp entries and exits in the section appears with statistically insignificant negative coefficients in models of property damage only crash frequencies, but with a statistically significant positive coefficient in the frequency model for possible injury and injury/fatal crashes. It appears that the number of ramp entries and exits contributes more to the frequency of severe crashes as opposed to less-severe crashes.

The composition of the traffic volume is found to be influential in all crash frequency equations. It is not particularly clear why the variables representing truck percentages appear with negative coefficients (some are statistically significant). This seemingly conflicting finding merits further investigation.
The average annual snowfall in inches is the only weather-related variable specified in the models and it appears with a statistically insignificant negative coefficient in property damage only crash frequency models, but with statistically insignificant positive coefficients in possible injury and injury/fatality crash frequency models. This finding suggests that snowfall may contribute positively to severe injury crashes and negatively to property damage only crashes; however, the insignificance of the coefficients indicates that snowfall appears to have a minimal impact, if any, on crash frequencies at various severity levels.

Returning to the point made earlier that all error correlations (presented in Table 4) were found to be highly statistically significant, it is clear that a simultaneous equations modeling methodology that accommodates cross-equation error covariances is appropriate for modeling safety phenomena such as that considered in this paper. It is possible that there are driver, vehicular, and roadway/traffic characteristics that are unobserved and simultaneously contributing to crash frequencies of various severity levels. The correlated unobserved factors contributing to different crash frequencies clearly call for the modeling of crash frequencies jointly using simultaneous equation systems. Such systems are also able to capture the differential impacts of variables on crash frequencies at various severity levels. While such differential impacts could potentially be captured in independent (non-joint) single-equation model systems, the estimates of such impacts are likely to be inaccurate if one were to ignore the presence of correlated error terms (i.e., the presence of correlated unobserved factors).

6. CONCLUSIONS

This paper focuses on modeling crash frequencies for freeway sections and identifying the factors and their relative contribution to crash frequencies by severity level. Freeway safety is a topic of much interest to transportation professionals for several reasons. First, as speeds on freeways are
higher than on other roadways, the occurrence of a crash on a freeway section is likely to result in a more severe outcome than on other roadways. Second, freeways are of vital importance for meeting travel needs of businesses and people and the occurrence of crashes on freeways results in severe economic losses through delays in travel time and costs associated with emergency response services, not to mention the enormous personal toll and cost associated with severe crashes. Enhancing safety on the nation’s freeways could play a significant part in reducing fatalities and saving precious public and private resources.

Research in the transportation safety arena has focused on the development of models of crash frequency using a variety of count data modeling methodologies including Poisson regression, Negative Binomial regression, and zero-inflated versions of these models. However, most research efforts in the past have treated crash frequency models as single equation models where a single crash frequency (either total or of a certain type or severity) is modeled as a function of roadway characteristics, traffic characteristics, roadway conditions, and environmental conditions. While such single equation systems offer very useful insights, there is the potential to significantly enhance such models by simultaneously modeling crash frequencies of different severity levels while explicitly recognizing that there may be common unobserved factors affecting crash frequencies at various severity levels. To this end, this paper presents a simultaneous equations model for several count dependent variables while explicitly recognizing the multivariate nature of the error structure across equations. The unobserved heterogeneity terms (or random error components) in the different equations are potentially correlated with one another leading to the simultaneity in the phenomenon under study and the multivariate nature of the error structure.

In this paper, a simultaneous multivariate Poisson (MVP) regression model that accounts for correlated error structures is formulated, developed, and estimated to analyze crash frequencies
for three different severity levels, i.e., property damage only, possible injury, and injury/fatal. Five-year crash frequency data for 275 freeway sections in the State of Washington are used in this study. The study involved formulating the MVP model and adopting a multivariate normal heterogeneity specification that allows the accommodation of correlated error structures while facilitating efficient model estimation using maximum simulated likelihood estimation (MSLE) procedures. In addition, the study involved the development of a simulation method (referred to as the triangular simulation method) to compute error variances and covariances in a computationally practical way. Comparisons between the results obtained using the triangular simulation method and the traditional Cholesky method suggested that the method presented in this paper offers robust parameter estimates.

The model estimation results are intuitively reasonable and offer valuable insights into the factors that affect crash frequencies at various severity levels. The key finding in this paper is that all error correlations are highly statistically significant strongly supporting the notion that crash frequencies (by severity) should be modeled in a rigorous simultaneous equations modeling framework such as that presented in this paper. Further research should explore the sensitivity of the findings to model specification and geographical contexts. In addition, data sets that separate out severe injury and fatal crash frequencies may offer richer insights into the differential impacts of various factors on crash frequencies at different severity levels.
REFERENCES


Table 1
Description of Dependent and Independent Variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total frequency of crashes</td>
<td>0</td>
<td>182</td>
<td>16.45</td>
<td>21.87</td>
</tr>
<tr>
<td>Frequency of property-damage-only crashes</td>
<td>0</td>
<td>113</td>
<td>6.09</td>
<td>10.01</td>
</tr>
<tr>
<td>Frequency of possible injury crashes</td>
<td>0</td>
<td>96</td>
<td>5.08</td>
<td>8.49</td>
</tr>
<tr>
<td>Frequency of injury crashes</td>
<td>0</td>
<td>93</td>
<td>5.28</td>
<td>8.67</td>
</tr>
<tr>
<td>Logarithm of AADT</td>
<td>8.12</td>
<td>12.06</td>
<td>10.13</td>
<td>0.88</td>
</tr>
<tr>
<td>Length of section (miles)</td>
<td>0.50</td>
<td>19.30</td>
<td>2.43</td>
<td>2.69</td>
</tr>
<tr>
<td>Maximum grade in section</td>
<td>-5.50</td>
<td>6.72</td>
<td>-0.22</td>
<td>3.07</td>
</tr>
<tr>
<td>Maximum central angle in section</td>
<td>0.00</td>
<td>111.49</td>
<td>30.29</td>
<td>23.88</td>
</tr>
<tr>
<td>Single-unit trucks’ percent in volume</td>
<td>1.90</td>
<td>10.00</td>
<td>4.20</td>
<td>1.21</td>
</tr>
<tr>
<td>Double-unit trucks’ percent in volume</td>
<td>0.55</td>
<td>17.80</td>
<td>7.76</td>
<td>4.62</td>
</tr>
<tr>
<td>Friction factor (scaled 0 to 100)</td>
<td>20.00</td>
<td>61.50</td>
<td>46.82</td>
<td>5.63</td>
</tr>
<tr>
<td>Number of grade breaks in section</td>
<td>0.00</td>
<td>28.00</td>
<td>3.87</td>
<td>4.09</td>
</tr>
<tr>
<td>Number of interchanges in section</td>
<td>0.00</td>
<td>4.00</td>
<td>0.85</td>
<td>0.83</td>
</tr>
<tr>
<td>Number of over-crossings in section</td>
<td>0.00</td>
<td>4.00</td>
<td>0.39</td>
<td>0.74</td>
</tr>
<tr>
<td>Number of ramp entries and exits in section</td>
<td>0.00</td>
<td>26.00</td>
<td>2.02</td>
<td>2.65</td>
</tr>
<tr>
<td>Number of median crossovers in section and the opposite direction</td>
<td>0.00</td>
<td>19.00</td>
<td>1.20</td>
<td>2.13</td>
</tr>
<tr>
<td>Average annual snowfall in inches</td>
<td>0.00</td>
<td>54.33</td>
<td>1.26</td>
<td>3.55</td>
</tr>
</tbody>
</table>
Table 2
Estimation Results of Recursive Poisson Regression Model with Normal or Log-gamma Error Structure

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Property-Damage-Only Crash Freq</th>
<th>Possible Injury Crash Freq</th>
<th>Injury/Fatal Crash Freq</th>
<th>Total Crash Freq</th>
<th>Total Crash Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crash Type</td>
<td>OLS</td>
<td>Univariate</td>
<td>Poisson</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-4.5353</td>
<td>-7.962</td>
<td>-5.0908</td>
<td>-10.301</td>
<td>-2.2712</td>
</tr>
<tr>
<td>Logarithm of AADT</td>
<td>0.5771</td>
<td>14.881</td>
<td>0.6026</td>
<td>13.081</td>
<td>0.6884</td>
</tr>
<tr>
<td>Length of section (miles)</td>
<td>0.1553</td>
<td>10.436</td>
<td>0.1075</td>
<td>6.121</td>
<td>0.1318</td>
</tr>
<tr>
<td>Maximum grade in section</td>
<td>0.0166</td>
<td>2.132</td>
<td>0.0374</td>
<td>3.976</td>
<td>0.0384</td>
</tr>
<tr>
<td>Maximum central angle in section / 100</td>
<td>0.2220</td>
<td>2.263</td>
<td>0.2400</td>
<td>2.086</td>
<td>0.7361</td>
</tr>
<tr>
<td>Single-unit trucks’ percent in volume</td>
<td>-0.0191</td>
<td>-0.780</td>
<td>-0.0901</td>
<td>-3.011</td>
<td>-0.0299</td>
</tr>
<tr>
<td>Double-unit trucks’ percent in volume</td>
<td>-0.0252</td>
<td>-3.260</td>
<td>-0.0013</td>
<td>-0.138</td>
<td>-0.0293</td>
</tr>
<tr>
<td>Friction factor / 10</td>
<td>-0.1620</td>
<td>-3.340</td>
<td>-0.1188</td>
<td>-2.288</td>
<td>-0.1725</td>
</tr>
<tr>
<td>Number of grade breaks in section</td>
<td>0.0142</td>
<td>1.495</td>
<td>0.0111</td>
<td>0.984</td>
<td>0.0373</td>
</tr>
<tr>
<td>Number of interchanges in section</td>
<td>0.3321</td>
<td>10.991</td>
<td>0.3486</td>
<td>9.333</td>
<td>0.2893</td>
</tr>
<tr>
<td>Number of over-crossings in section</td>
<td>0.0481</td>
<td>1.811</td>
<td>0.0887</td>
<td>2.343</td>
<td>0.0448</td>
</tr>
<tr>
<td>Number of ramp entries and exits in section</td>
<td>0.0128</td>
<td>1.296</td>
<td>0.0304</td>
<td>2.493</td>
<td>0.0496</td>
</tr>
<tr>
<td>Number of median crossovers in section and the opposite direction</td>
<td>-0.0045</td>
<td>-0.616</td>
<td>0.0038</td>
<td>0.430</td>
<td>-0.0473</td>
</tr>
<tr>
<td>Average annual snowfall in inches</td>
<td>0.0049</td>
<td>0.510</td>
<td>0.0179</td>
<td>1.491</td>
<td>-0.0019</td>
</tr>
</tbody>
</table>

Parameters in Error Structure

| $f_1, f_2, f_4, \sigma, \alpha$ | 1.0128 | 44.949 | 0.0000 | -- | 0.0000 | -- | 0.7170 | 39.076 | 0.6986 | 42.584 |
| $f_3$ | -- | -- | 1.0731 | 35.814 | 0.0000 | -- | -- | -- | -- | -- |
| $f_6$ | -- | -- | -- | -- | 1.0537 | 44.153 | -- | -- | -- | -- |

Goodness of Fit Measure

| LL($\beta$) | -10563.9 | -4682.99 | -4700.20 |
| LL($c$) | -23816.5 | -15847.5 | -15847.5 |
| Adj. $\rho^2(c)$ | 0.5547 | 0.7036 | 0.7026 |
| $N$ | 1375 | 1375 | 1375 |
Table 3
Estimation Results of Joint Poisson Regression Model with Trivariate Normal Error Structure

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Joint Poisson Model with Trivariate Normal Error Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crash Type</td>
<td>Property-Damage-Only Crash Freq</td>
</tr>
<tr>
<td>Variable</td>
<td>Coef</td>
</tr>
<tr>
<td>Constant</td>
<td>-4.8060</td>
</tr>
<tr>
<td>Logarithm of AADT</td>
<td>0.5808</td>
</tr>
<tr>
<td>Length of section (miles)</td>
<td>0.1899</td>
</tr>
<tr>
<td>Maximum grade in section</td>
<td>0.0111</td>
</tr>
<tr>
<td>Maximum central angle in section / 100</td>
<td>0.4308</td>
</tr>
<tr>
<td>Single-unit trucks’ percent in volume</td>
<td>-0.0402</td>
</tr>
<tr>
<td>Double-unit trucks’ percent in volume</td>
<td>-0.0281</td>
</tr>
<tr>
<td>Friction factor / 10</td>
<td>-0.0974</td>
</tr>
<tr>
<td>Number of grade breaks in section</td>
<td>0.0153</td>
</tr>
<tr>
<td>Number of interchanges in section</td>
<td>0.2142</td>
</tr>
<tr>
<td>Number of over-crossings in section</td>
<td>0.0645</td>
</tr>
<tr>
<td>Number of ramp entries and exits in section</td>
<td>-0.0025</td>
</tr>
<tr>
<td>Number of median crossovers in section and the opposite direction</td>
<td>0.0043</td>
</tr>
<tr>
<td>Average annual snowfall in inches</td>
<td>-0.0072</td>
</tr>
</tbody>
</table>

Parameters in Error Structure

\[ f_1, f_2, f_4, \sigma, \alpha \]

\[ 1.0205 \quad 39.143 \quad 0.2809 \quad 9.905 \quad 0.3451 \quad 11.782 \]

\[ f_3, f_5 \]

\[ -- \quad -- \quad 9.522 \quad 36.779 \quad 0.3463 \quad 10.747 \]

\[ f_6 \]

\[ -- \quad -- \quad -- \quad 0.9751 \quad 34.411 \]

Goodness of Fit Measure

\[ \text{LL}(\beta) \quad -10440.6 \]

\[ \text{LL}(c) \quad -23816.5 \]

\[ \text{Adj. } \rho^2(c) \quad 0.5597 \]

\[ N \quad 1375 \]
Table 4
Estimated Expectation and Standard Deviation of Elements in Error Covariance Matrix

<table>
<thead>
<tr>
<th>Elements</th>
<th>Triangular Method</th>
<th>Cholesky Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std Dev</td>
</tr>
<tr>
<td>$Var_1$</td>
<td>1.0421</td>
<td>0.0534</td>
</tr>
<tr>
<td>$Var_2$</td>
<td>0.9870</td>
<td>0.0507</td>
</tr>
<tr>
<td>$Var_3$</td>
<td>1.1925</td>
<td>0.0652</td>
</tr>
<tr>
<td>$Cov_{12}$</td>
<td>0.2868</td>
<td>0.0304</td>
</tr>
<tr>
<td>$Cov_{13}$</td>
<td>0.3523</td>
<td>0.0321</td>
</tr>
<tr>
<td>$Cov_{23}$</td>
<td>0.4268</td>
<td>0.0336</td>
</tr>
<tr>
<td>$Std_1$</td>
<td>1.0205</td>
<td>0.0262</td>
</tr>
<tr>
<td>$Std_2$</td>
<td>0.9932</td>
<td>0.0255</td>
</tr>
<tr>
<td>$Std_3$</td>
<td>1.0916</td>
<td>0.0299</td>
</tr>
<tr>
<td>$Corr_{12}$</td>
<td>0.2829</td>
<td>0.0276</td>
</tr>
<tr>
<td>$Corr_{13}$</td>
<td>0.3164</td>
<td>0.0281</td>
</tr>
<tr>
<td>$Corr_{23}$</td>
<td>0.3934</td>
<td>0.0226</td>
</tr>
</tbody>
</table>
Figure 1. Variation of Expectation and Standard Deviation of Log-gamma Heterogeneity by Alpha Value

Figure 2. Comparison of Probability Density Function of Estimated Log-gamma and Normal Heterogeneity