Safety Science 47 (2009) 443–452

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/ssci

A simultaneous equations model of crash frequency by collision type for rural intersections

Xin Ye [a,*], Ram M. Pendyala [a,1], Simon P. Washington [a,2], Karthik Konduri [a,3], Jutaek Oh [b]

[a] Department of Civil and Environmental Engineering, Arizona State University, Room ECG252, Tempe, AZ 85287-5306, United States
[b] Department of Highway Research, The Korea Transport Institute, 2311 Daehwa-dong, Ilsan-gu, Goyang-si, Gyeonggi-do 411-701, Republic of Korea

ARTICLE INFO

Article history:
Received 4 March 2008
Received in revised form 20 June 2008
Accepted 24 June 2008

Keywords:
Intersection safety
Crash frequency
Collision type
Simultaneous equations model
Multivariate Poisson regression
Unobserved heterogeneity

ABSTRACT

Safety at roadway intersections is of significant interest to transportation professionals due to the large number of intersections in transportation networks, the complexity of traffic movements at these locations that leads to large numbers of conflicts, and the wide variety of geometric and operational features that define them. A variety of collision types including head-on, sideswipe, rear-end, and angle crashes occur at intersections. While intersection crash totals may not reveal a site deficiency, over exposure of a specific crash type may reveal otherwise undetected deficiencies. Thus, there is a need to be able to model the expected frequency of crashes by collision type at intersections to enable the detection of problems and the implementation of effective design strategies and countermeasures. Statistically, it is important to consider modeling collision type frequencies simultaneously to account for the possibility of common unobserved factors affecting crash frequencies across crash types. In this paper, a simultaneous equations model of crash frequencies by collision type is developed and presented using crash data for rural intersections in Georgia. The model estimation results support the notion of the presence of significant common unobserved factors across crash types, although the impact of these factors on parameter estimates is found to be rather modest.

1. Introduction

There are numerous vehicular conflict points that arise at intersections due to the variety of movements (both through and turning) that need to be accommodated at intersections. Traffic control devices may be used to minimize the number of conflicts and enhance the safety of operations at intersections. Despite such efforts, conflicts may still exist and/or driver errors may take place resulting in crashes at intersections. Transportation professionals are constantly searching for ways to design systems (roadway geometry and control device configuration) to enhance safety at intersections.

Due to the variety of vehicular movements at intersections, one encounters a variety of collision types at intersections. Collision types at intersections include rear-end, angle, sideswipe (same direction and opposite direction), head-on, and pedestrian-involved crashes. It is quite easy to see how this range of crash types can occur at intersections. When a vehicle stops at a traffic light, it is possible that a following vehicle rear-ends the vehicle in front. Turning movements, particularly those that are not protected, may result in angle crashes or head-on collisions. When drivers attempt swerving maneuvers to avoid a potential collision, it is possible to encounter sideswipes. If conflicts arise between pedestrian crossing and vehicular turning movements, then pedestrian-involved crashes may occur.

It is possible to model crash frequencies by collision type separately or independently as a function of intersection characteristics and environmental conditions. However, such independent equations of crash frequency (by collision type) ignore the potential correlations that may exist across collision types. Most notably, there may be common unobserved factors that simultaneously impact the frequency of crashes by collision type at intersections. For example, sight distance related factors or adjacent land use considerations may simultaneously influence the frequency of two or more crash or collision types. In this context, it is advisable to develop and deploy simultaneous equations models of crash frequencies by collision type to accurately estimate the impacts of various explanatory variables on crash frequencies. Ignoring such simultaneity, when in fact, it may be present, would result in inefficient estimates of (single-equation) model parameters. This paper aims to present a multivariate count data model that is capable of accounting for correlated unobserved factors across equations

* Corresponding author. Tel.: +1 480 965 2262; fax: +1 480 965 0557.
E-mail addresses: xin.ye@asu.edu (X. Ye), ram.pendyala@asu.edu (R.M. Pendyala), simon.washington@asu.edu (S.P. Washington), karthik.konduri@asu.edu (K. Konduri), jutaek@koti.re.kr (J. Oh).
[1] Tel.: +1 480 727 9164; fax: +1 480 965 0557.
[2] Tel.: +1 480 965 2220; fax: +1 480 965 0557.
[3] Tel.: +1 480 965 3589; fax: +1 480 965 0557.

0925-7535/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ssci.2008.06.007

representing frequencies of different collision types. 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 quantified error correlations in the joint model can aid in better understanding the role of unobserved factors that simultaneously affect the occurrence of crashes of various types.

The need for developing and estimating simultaneous equations model systems of crash frequencies by collision type has been recognized in the literature (Ma et al., 2008; Park and Lord, 2007); however, the formulation and estimation of such model systems is analytically complex and computationally challenging, particularly in the context of multivariate count data. This paper formulates and presents an n-dimensional multivariate count data model (multivariate 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 consistent and efficient parameter estimates, and test statistics for hypotheses testing. From an empirical standpoint, by presenting a simultaneous equations model system for intersection crash collision types, the paper provides key insights into the factors that impact crash frequencies of various collision types while accounting for the presence of error covariances (common unobserved factors).

Following a brief review of the literature in Section 2, the paper presents the modeling methodology adopted in this paper in Section 3. This is followed by a description of the dataset in Section 4. Model estimation results and key conclusions are presented in the Sections 5 and 6, respectively.

2. Literature on crash modeling

The literature provides plenty of evidence of the level of importance attached to intersection safety. Several researchers have analyzed intersection crash statistics and attempted to develop models of crash frequencies for intersections. Abdel-Aty and Keller (2005) model crash severity levels at signalized intersections. Abdel-Aty and Nawathe (2006) follow up on the previous study and present a novel approach for classifying and predicting signalized intersection crashes. Yan et al. (2005) present a study that focuses exclusively on modeling rear-end crashes at signalized intersections using a multiple logistic regression model while Wang and Abdel-Aty (2006) analyze temporal and spatial characteristics of such rear-end crashes. Pai and Saleh (2007a) use crash data from three-legged intersections in UK (involving motorcyclists). They present separate ordered probit models of crash severity for signalized and unsignalized intersections. They follow up on this study (Pai and Saleh, 2007b) and use the ordered probit methodology to develop multiple crash severity models for different crash types occurring at these intersections. Jonsson et al. (2007), Kim et al. (2006, 2007), Kim and Washington (2006), Lyon et al. (2003), Oh et al. (2003, 2004) constitute a series of papers that have examined intersection crash frequencies with a particular emphasis on safety considerations and factors influencing crash frequencies in rural environments. Other papers that have focused on safety in a rural environment include Ossenbruggen et al. (2001) and Zajac and Ivan (2003).

Much of the work on modeling crash frequencies involves the use of count data models. Literature on transportation safety is replete with papers of count data models for analyzing crash frequencies under a variety of conditions. Lee and Mannering (2002) model crash frequency and severity of run-off roadway crashes. Miaou et al. (1992) and Miaou (1994) examine the impacts of roadway geometrics on truck crash frequencies. Milton and Mannering (1998) present count data models of crash frequencies to examine the impacts of roadway geometrics and traffic characteristics on crash counts. Poch and Mannering (1996) present a negative binomial model of crash intersection frequencies. Shankar et al. (1997) is an example of a study that employs zero-inflated Poisson and Negative Binomial models for modeling crash frequencies. Shankar et al. (2003) present empirical models based on Negative Binomial distributions and mixing distributions to analyze pedestrian crashes using pedestrian crash data from Washington. Ma and Kockelman (2006a) present a model of crash frequency and severity using clustered data from Washington State. Lord et al. (2005) provide a discussion on the use of such models for modeling crash frequencies and suggest that caution needs to be exercised before using zero-inflated versions of count data models. In addition to count data modeling approaches, other approaches such as log-linear modeling methods have been employed to analyze crash frequencies (e.g., Golob and Recker, 1987). Nevertheless, count data modeling methods appear to be the standard for crash frequency analysis.

This paper may be considered an extension of the series of papers on the development of models of rural intersection crash frequencies by Kim et al. (2006, 2007), Kim and Washington (2006), Oh et al. (2003, 2004). All of those papers involve the use of count data models to examine crash frequencies at rural intersections. Kim et al. (2006) present six independent count data models – two Poisson and four Negative Binomial regression equations to predict crash frequencies by collision type.

The key issue associated with much of the literature cited here is that models are either developed for a single crash frequency type or separate model equations are estimated for a series of crash types. These single-equation approaches do not adequately account for the possibility that there may be common unobserved factors that affect crash frequencies of different crash types. In light of this limitation in existing approaches to crash frequency modeling, this paper extends previous work in this area by presenting a simultaneous equations system of crash frequencies by crash type using exactly the same data set and crash type definitions as used in Kim and Washington (2006). The intent is to explore whether there are common unobserved factors that simultaneously affect crash frequencies of different crash types.

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. After that initial attempt, Ma and Kockelman (2006b) and Ma et al. (2008) later estimated Poisson regression model formulations with multivariate normal heterogeneities. In these two papers, the authors adopted Bayesian methods 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, except that they focused on crash severity while this paper focuses on crash types).

Park and Lord (2007) present a Multivariate Poisson Regression – lognormal model for modeling crash frequencies by severity level at intersections. They employ a Markov Chain Monte Carlo (MCMC) simulation approach within a Bayesian framework to evaluate the multi-dimensional integral of the Poisson distribution and estimate parameters. 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. (2008) and Park and Lord (2007). Unlike their estimation approach, the simulation-based maximum likelihood estimation method employed in this paper uses Halton sequence draws to accurately compute the log-likelihood function (evaluate multi-dimensional 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). In addition, this paper focuses on crash collision type frequencies as opposed to severity frequencies at intersections.

3. Modeling methodology

This section presents the modeling methodology adopted in this paper. The Negative Binomial (NB) regression modeling methodology is commonly adopted for modeling crash frequencies (Cameron and Trivedi, 1986). This modeling approach employs a Poisson regression model that is modified to overcome over- or under-dispersion in the crash frequency variable (i.e., the variance and mean of a crash frequency variable are not equal to one another). This is done by accommodating additional heterogeneity in the random error component, which is assumed to be log-gamma distributed. The choice of the log-gamma distribution ensures that there is a closed form solution for the resulting log-likelihood function, thus providing computational tractability in the estimation of model parameters. However, when considering multiple crash frequency (count) variables simultaneously, it is preferable to assume that the random error component (heterogeneity) is normally distributed, thus leading to a multivariate normal (MVN) random error structure for the simultaneous equations system of interest. The MVN random error structure offers an analytically tractable and theoretically appealing framework for simultaneously modeling crash frequencies of various types.

3.1. Univariate poisson regression (UVP) model with normal heterogeneity

This section presents the formulation for a univariate Poisson regression model as a precursor to the presentation of the multivariate Poisson regression model. Let y_i represent crash frequency by collision type (with the subscript representing collision type suppressed without loss of generality) for intersection i. In the literature, crash frequency, y_i, is often represented using a class of count data models (e.g., Poch and Mannering, 1996; Oh et al., 2004; Lord et al., 2005; Jonsson et al., 2007; Ma et al., 2008). In the case of the Negative Binomial (NB) model, the expectation of y_i is assumed to be λ_i and,

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

where x_i is a row vector of explanatory variables indicating characteristics for intersection i and β is a column vector of coefficients associated with x_i. ε_i is a random error term representing heterogeneity that accounts for unobserved factors and other random disturbances.

As y_i constitutes count data, the probability of y_i conditional on ε_i is given as:

$$\Pr(y_i|\varepsilon_i) = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!}. \tag{2}$$

As mentioned previously, one may assume a normal distribution for the random error term, ε_i, as opposed to the log-gamma distribution that is typically assumed in the univariate NB model. Under the assumption of normality, one can integrate ε_i over its distributional domain and obtain the unconditional probability of y_i as:

$$\Pr(y_i) = \int_{-\infty}^{\infty} \frac{\exp[-\exp(x_i\beta + \varepsilon_i)][\exp(x_i\beta + \varepsilon_i)]^{y_i}}{y_i!} \frac{1}{\sqrt{2\pi}\sigma} \times \exp(-\frac{\varepsilon_i^2}{2\sigma^2})\,d\varepsilon_i. \tag{3}$$

The key difference between this model formulation and that of the traditional NB regression model is that Eq. (3) does not have a closed form expression. Therefore, one needs to approximate the unconditional probability as follows:

$$SP(y_i) = \frac{1}{R}\sum_{r=1}^{R} \frac{\exp[-\exp(x_i\beta + \sigma u_{ir})][\exp(x_i\beta + \sigma u_{ir})]^{y_i}}{y_i!}, \tag{4}$$

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 σ by multiplying them with a single factor σ. The Maximum Simulated Likelihood Estimation (MSLE) method can then be applied to efficiently estimate unknown parameters β and σ with the aid of quasi-random seeds (Bhat, 2003). This procedure is similar to that of the regular Maximum Likelihood Estimation (MLE) procedure except that the log-likelihood function and its first-order derivative are numerically approximated by making quasi-random draws from a normal distribution. Ordinary Least Square (OLS) estimates may be used as starting values for β. As for the initial value of σ, one may use the value of the standard deviation estimated from the conventional NB regression model.

3.2. 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 ($n \geqslant 2$) dependent (count) variables can be jointly modeled. The correlation among them 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.

In this paper, crash frequencies by type of collision are modeled jointly. Crashes are classified into six types: Angle, Head-on, Rear-end, Same Direction Sideswipe, Opposite Direction Sideswipe, and Pedestrian-involved crashes. Thus, there are six interrelated dependent count variables and the logarithm of expectations λ_1, λ_2, λ_3, λ_4, λ_5 and λ_6 for these six count variables may be formulated as:

$$\begin{cases} \ln(\lambda_{1i}) = x_{1i}\beta_1 + \varepsilon_1 = x_{1i}\beta_1 + \delta_1 u_{1i}, \\ \ln(\lambda_{2i}) = x_{2i}\beta_2 + \varepsilon_2 = x_{2i}\beta_2 + \delta_2 u_{1i} + \delta_3 u_{2i}, \\ \ln(\lambda_{3i}) = x_{3i}\beta_3 + \varepsilon_3 = x_{3i}\beta_3 + \delta_4 u_{1i} + \delta_5 u_{2i} + \delta_6 u_{3i}, \\ \ln(\lambda_{4i}) = x_{4i}\beta_4 + \varepsilon_4 = x_{4i}\beta_4 + \delta_7 u_{1i} + \delta_8 u_{2i} + \delta_9 u_{3i} + \delta_{10} u_{4i}, \\ \ln(\lambda_{5i}) = x_{5i}\beta_5 + \varepsilon_5 = x_{5i}\beta_5 + \delta_{11} u_{1i} + \delta_{12} u_{2i} + \delta_{13} u_{3i} + \delta_{14} u_{4i} + \delta_{15} u_{5i}, \\ \ln(\lambda_{6i}) = x_{6i}\beta_6 + \varepsilon_6 = x_{6i}\beta_6 + \delta_{16} u_{1i} + \delta_{17} u_{2i} + \delta_{18} u_{3i} + \delta_{19} u_{4i} + \delta_{20} u_{5i} \\ \qquad + \delta_{21} u_{6i}, \end{cases} \tag{5}$$

where u_{1i}, u_{2i}, u_{3i}, u_{4i}, u_{5i} and u_{6i} are six independent random variables, which are standard normally distributed and δ_i are coefficients to be estimated; x_i are vectors of explanatory variables and β_i are coefficients associated with variables in x_i.

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

$$\begin{cases} \Pr(y_{1i}|u_{1i}) = \frac{\exp(-\lambda_{1i})\lambda_{1i}^{y_1 i}}{y_1 i!}, \\ \Pr(y_{2i}|u_{1i}, u_{2i}) = \frac{\exp(-\lambda_{2i})\lambda_{2i}^{y_{2i}}}{y_{2i}!}, \\ \cdots \\ \Pr(y_{6i}|u_{1i}, u_{2i}, u_{3i}, u_{4i}, u_{5i}, u_{6i}) = \frac{\exp(-\lambda_{6i})\lambda_{6i}^{y_{6i}}}{y_{6i}!}. \end{cases} \tag{6}$$

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

$$\Pr(y_{1i}, y_{2i}, y_{3i}, y_{4i}, y_{5i}, y_{6i}) = \int \left[\prod_{j=1}^{6} Pr(y_{ji}|u_i)\right] g(u_i)\, du_i, \tag{7}$$

where $g(u_i)$ is the joint probability density function for independent random variables $u_{1i}, u_{2i}, \ldots, u_{6i}$ and the integral here represents a six-dimensional integral. As this complex multi-dimensional integral does not have a neat closed form expression, the unconditional probability function may be approximated by the simulated probability function given below:

$$\mathrm{SP}(y_{1i}, y_{2i}, y_{3i}, y_{4i}, y_{5i}, y_{6i}) = \frac{1}{R}\left\{\sum_{r=1}^{R}\left[\prod_{j=1}^{6} Pr(y_{ji}|u_{ir})\right]\right\}, \tag{8}$$

where u_{ir} are six columns of independent random seeds drawn from a standard normal distribution. These terms can be transformed into linear combinations of δ_i to realize a multivariate normal heterogeneity that may be prevalent in simultaneous equations models of crash frequencies by type. R is the number of Monte Carlo draws over which the probability function is approximated. According to the theory of simulation-based maximum likelihood estimators (Train, 2002), R needs to be greater than the square root of the sample size for obtaining consistent and efficient estimators. As the sample size is small in this study ($N = 165$), and the Halton sequence (Bhat, 2003) is used to generate quasi-random seeds, the value of R is conservatively set at 100. In this particular study, this is sufficient to accurately approximate the log-likelihood function and provide consistent and efficient parameter estimators. Variance, covariance, standard deviation, and correlation terms associated with the multivariate normal error structure may be calculated based on parameters δ_i in a straightforward manner. For example, consider ε_2 and ε_3. Then, the following illustrates how the various parameters of interest can be computed:

$$\mathrm{Var}(\varepsilon_3) = \delta_4^2 + \delta_5^2 + \delta_6^2, \tag{9}$$

$$\mathrm{Cov}(\varepsilon_2, \varepsilon_3) = \delta_2\delta_4 + \delta_3\delta_5, \tag{10}$$

$$\mathrm{Std}(\varepsilon_2) = \sqrt{\delta_2^2 + \delta_3^2}, \tag{11}$$

$$\mathrm{Corr}(\varepsilon_2, \varepsilon_3) = \frac{\delta_2\delta_4 + \delta_3\delta_5}{\sqrt{(\delta_2^2 + \delta_3^2)(\delta_4^2 + \delta_5^2 + \delta_6^2)}}, \tag{12}$$

As in the case of the univariate model with normal heterogeneity (described in Section 3.1), the MSLE method can be applied to estimate unknown parameter β and δ_i using reasonable starting values. Ordinary Least Squares (OLS) estimates may be used as starting values for β. The overall standard deviation of error terms ε_1, ε_2, $\varepsilon_3, \ldots, \varepsilon_6$ can be approximated from corresponding values obtained in the Negative Binomial or Univariate Poisson regression models. Then, these estimated standard deviations are taken as starting values for the last δ_i terms in each line of Eq. group (5) (i.e. δ_1, δ_3, δ_6, δ_{10}, δ_{15}, and δ_{21}). The other δ_i terms start at zero. The log-likelihood function and its first-order derivative were coded in Gauss (Aptech, 2006) and the default BFGS (Broyden, Fletcher, Goldfarb, and Shanno) algorithm provided by the Maxlik module in Gauss was used for maximizing the log-likelihood function.

One shortcoming of the simulation method described above 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, δ_i). To overcome this shortcoming, a simulation-based hypothesis test (Ye and Pendyala, 2007) is applied to approximate statistical inferences for these calculated elements of interest (error variance-covariance terms). The idea is that all of the parameter estimators obtained through 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 this procedure, a few columns of random variables that provide a multivariate normal distribution for relevant estimators are drawn. Then, the relevant columns of random variables can be picked out 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 $\mathrm{Cov}(\varepsilon_2, \varepsilon_3)$ is of interest. Note that $\mathrm{Cov}(\varepsilon_2, \varepsilon_3) = \delta_2\delta_4 + \delta_3\delta_5$. Then, one needs to pick out four columns of random variables corresponding to δ_2, δ_4, δ_3, and δ_5, and calculate the expression $\delta_2\delta_4 + \delta_3\delta_5$ for each set of random variables to generate a column of "randomized" $\mathrm{Cov}(\varepsilon_2, \varepsilon_3)$. The percent of positive or negative counts among the random seeds for $\mathrm{Cov}(\varepsilon_2, \varepsilon_3)$ can approximate the significance level of $\mathrm{Cov}(\varepsilon_2, \varepsilon_3)$, i.e., the probability that $\mathrm{Cov}(\varepsilon_2, \varepsilon_3)$ is positive or negative. In addition, one may approximate the expectation and standard deviation of the elements of interest based on the mean value and standard deviation of these random seeds. However, as these elements are not normally distributed, the traditional t-statistic (obtained by dividing the mean by the standard deviation) cannot be used for statistical inference.

4. Data description

The data set used in this study is derived from two years (1996 and 1997) of rural two-lane intersection crash records for 38 counties in the State of Georgia. This data set has been used extensively in the past for intersection safety analysis and provides crash frequencies by collision type for 51 signalized and 114 unsignalized rural intersections in the state (Kim et al., 2006; Kim and Washington, 2006). The crash database incorporates frequencies based on a total of 837 crashes taking place at these intersections. Crashes occurring within the intersection or within 76 m (250 ft) of the center of the intersection along the major and minor road are included in the analysis. While this classification scheme may omit some intersection crashes and/or include some non-intersection crashes, it is commonly used in the United States as it is a non-arbitrary criterion that is easily repeatable and generalizable across jurisdictions. A series of explanatory variables describing roadway characteristics, intersection characteristics and geometry, and traffic volumes for both major and minor roads were extracted from roadway characteristics inventories and GIS roadway networks and merged with the crash frequency database.

A descriptive analysis of the database is presented in Table 1. Angle and rear-end collisions are found to show the highest average crash frequencies. These are followed by the pedestrian crash frequency. Head-on collisions and sideswipe collisions show the lowest average frequencies. The pedestrian crash frequency appears to be rather high (average of nearly one crash per intersection in the two year period). It is possible that pedestrian volumes are high as evidenced by the high minor road volumes. However, further investigation is warranted and data on pedestrian volumes is needed before any definitive statements can be made on the reasons for the high pedestrian crash frequency.

5. Model estimation results

Model estimation results are presented in Tables 2–4. Table 2 presents the estimates of the error correlations. The most important finding revealed by the table is that all inter-equation error correlations are positive and are statistically significant except for those representing pair wise correlations with pedestrian-involved crash frequencies. Very high error correlations (>0.8) are seen between angle and rear-end crashes, angle and sideswipe (same direction) crashes, and rear-end and sideswipe (same direc-

Table 1
Variable description ($N = 165$)

Variable name	Variable description	Min	Max	Mean	Std. dev.
ANGLE	Angle accident Frequency	0.00	33.00	2.20	3.58
HEADON	Head-on accident frequency	0.00	2.00	0.13	0.38
REAREND	Rear-end accident frequency	0.00	15.00	1.44	2.48
SIDESAME	Sideswipe (the same direction) accident frequency	0.00	6.00	0.25	0.72
SIDEOPPO	Sideswipe (the opposite direction) accident frequency	0.00	3.00	0.13	0.40
PEDESTRIAN	Pedestrian-involved accident frequency	0.00	4.00	0.92	0.98
LNADTMJ	Logarithm of annual average daily traffic on major road	6.04	9.63	8.04	0.94
LNADTMI	Logarithm of annual average daily traffic on minor road	4.38	9.25	6.57	0.99
ADTMJ/1000	Annual average daily traffic on major road /1000	0.42	15.20	4.53	3.72
ADTMI/1000	Annual average daily traffic on minor road /1000	0.08	10.40	1.27	1.79
SHLWDMAJ	Shoulder width on major road (feet)	0.00	10.00	1.43	1.42
LIGHTMAJ	Lighting exists on the major road	0.00	1.00	0.20	0.40
SPDLIMAJ/100	Speed limit on major road/100	0.25	0.55	0.46	0.08
LTLMIN	Number of left turn lanes on minor road	0.00	2.00	0.27	0.67
LTLMAJ	Number of left turn lanes on major road	0.00	2.00	0.49	0.86
RTLMAJ	Number of right turn lanes on major road	0.00	2.00	0.19	0.50
TERNMIN	Terrain on minor road (0 = flat, 1 = rolling, 2 = mountainous)	0.00	2.00	0.69	0.55
VIMAJ	Sum of absolute change of grade in percent per hundred feet for each curve on major road divided by the number of such curves	0.00	5.63	1.74	1.13

Table 2
Matrix of error correlations (symmetric matrix)

Collision type	Angle	Head-on	Rear-end	Sideswipe (same direction)	Sideswipe (opposite direction)	Pedestrian-involved
Angle	1.0000					
Head-on	0.7529	1.0000				
	(0.2841)					
	[0.0247]					
Rear-end	0.9548	0.7859	1.0000			
	(0.0366)	(0.2613)				
	[0.0000]	[0.0206]				
Sideswipe (Same Direction)	0.8628	0.7524	0.8512	1.0000		
	(0.1029)	(0.2497	(0.1267)			
	[0.0000]	[0.0200]	[0.0003]			
Sideswipe (Opposite Direction)	0.5243	0.5981	0.5639	0.6524	1.0000	
	(0.2262)	(0.2719)	(0.2241)	(0.2085)		
	[0.0154]	[0.0331]	[0.0134]	[0.0063]		
Pedestrian-involved	0.1866	0.2501	0.2263	0.2345	0.3620	1.0000
	(0.2427)	(0.2954)	(0.2525)	(0.2774)	(0.3057)	
	[0.2166]	[0.1962]	[0.1842]	[0.1997]	[0.1298]	

The numbers in parentheses are standard deviations of estimated elements of the correlation matrix.
The numbers in square brackets are *p*-values denoting the significance of the estimated elements of the correlation matrix.

tion) crash frequencies. In other words, there are significant common unobserved factors affecting frequencies of these pairs of crash types. Statistically significant, but lower error correlations (between 0.5 and 0.8), are found between numerous other crash pairs. These include those between head-on and sideswipe crashes in the same direction, head-on and rear-end crashes, head-on and angle crashes, sideswipe crashes in the opposite direction and in the same direction, sideswipe crashes in the opposite direction and head-on crashes, sideswipe crashes in the opposite direction and rear-end crashes, and sideswipe crashes in the opposite direction and angle crashes. It is fairly easy to think of common unobserved factors that influence crash frequencies by collision type. There are a host of roadway surface conditions, environmental and weather conditions, driver population factors, adjacent land use characteristics, traffic composition variables (trucks, buses, etc.), vegetation and sight distance related factors, and traffic control device related configuration variables that likely influence crash frequencies of various types. Essentially, the modeling exercise of this paper has revealed that the presence of such common unobserved factors is significant and worthy of consideration in crash frequency analysis.

The only set of error correlations that turned out statistically insignificant is that involving pedestrian crash frequencies. All of the error correlations are less than 0.37 in value. When it comes to pedestrian-involved crash frequencies, it appears that the unobserved factors contributing to such crashes are more pedestrian-related. Perhaps they relate to the provision of pedestrian crosswalks, pedestrian signals, and other pedestrian safety-related features. These unobserved factors are more unique to pedestrian- involved crashes and as a result, the correlation between unobserved factors affecting vehicular crash frequencies and those affecting pedestrian-involved crash frequencies is likely to be lower. This is not to say that there are no common unobserved factors affecting these pairs of crash frequencies. The correlations are all positive in magnitude and show significance levels between 0.1 and 0.2. This suggests that error correlations involving pedestrian crashes are also worthy of consideration, but tend to be less significant in the context of a rural intersection-based crash frequency analysis.

Table 3 presents the model estimation results for different collision type frequencies. In general, most variables were retained in the final model specification based on a combination of their intuitive explanatory interpretation and statistical significance. The model associated with sideswipe – opposite direction frequency did not yield statistically significant results, possibly due to the very small numbers of such crashes at intersections and the large numbers of zeros in the data for this particular collision type.

Table 3
Estimation results for the multivariate poisson (MVP) regression model

Collision type	Angle		Head-on		Rear-end		Sideswipe (same direction)		Sideswipe (opposite direction)		Pedestrian- involved	
Variable	Coefficient	*t*-statistic	Coefficient	*t*-statistic	Coefficient	*t*-statistic	Coefficient	*t*-statistic	Coefficient	*t*-statistic	Coefficient	*T*-test
Constant	−4.9188	−4.9130	−16.6220	−4.1130	−12.3869	−8.7420	−1.5630	−1.4980	0.0951	0.0490	−2.2734	−3.9580
LNADTMJ	0.3632	2.9030	0.7351	1.9300	1.0132	6.2380	–	–	–	–	–	–
LNADTMI	0.3032	3.2180	0.7874	3.3140	0.5023	4.9690	–	–	−0.4621	−1.5140	0.3133	3.7140
SHLWDMAJ	0.0969	2.1330	–	–	–	–	0.1820	2.2640	–	–	0.1069	2.0050
LIGHTMAJ	−0.3948	−2.2130	–	–	–	–	–	–	–	–	-0.5757	-2.3720
SPDLIMAJ/100	–	–	5.6013	1.6000	–	–	−4.1353	−2.2160	–	–	–	–
LTLMAJ	–	–	–	–	–	–	0.7785	4.0650	–	–	–	–
LTLMIN	0.3495	3.2520	–	–	–	–	–	–	–	–	–	–
RTLMAJ	–	–	–	–	0.2194	1.7070	–	–	0.6300	1.7370	–	–
TERNMIN	–	–	–	–	0.2424	1.5720	–	–	–	–	–	–
VIMAJ	–	–	–	–	–	–	0.2559	1.7530	–	–	–	–
Coefficients for error structure												
δ_1, δ_2, δ_4, δ_7, δ_{11}, δ_{16}	0.8700	8.6310	0.6365	1.9730	0.8664	7.1620	1.0756	4.0190	0.7360	2.1520	0.0892	0.7850
δ_3, δ_5, δ_8, δ_{12}, δ_{17}	–	–	−0.3131	−0.7040	−0.1744	−1.4010	−0.2031	−0.5790	−0.5076	−0.8690	−0.1096	−0.5770
δ_6, δ_9, δ_{13}, δ_{18}	–	–	–	–	0.0379	0.2350	0.1325	0.4190	0.0253	0.0590	−0.0142	−0.0680
δ_{10}, δ_{14}, δ_{19}	–	–	–	–	–	–	0.1846	0.7320	0.5273	1.3890	0.1496	0.8400
δ_{15}, δ_{20}	–	–	–	–	–	–	–	–	0.3334	0.9030	0.1049	0.4990
δ_{21}	–	–	–	–	–	–	–	–	–	–	−0.0194	−0.0920
Goodness-of-fit measures												
LL(β)	−862.92											
LL(c)*	−1283.63											
Adj. ρ^2	0.2950											
Sample size	165											

* LL(c) is measured by the log-likelihood value at convergence when only six constants are specified into the model.

Table 4
Estimation results for the univariate poisson (UVP) regression model

Collision type	Angle		Head-on		Rear-end		Sideswipe (same direction)		Sideswipe (opposite direction)		Pedestrian- involved	
Variable	Coefficient	*t*-statistic	Coefficient	*t*-statistic	Coefficient	*t*-statistic	Coefficient	*t*-statistic	Coefficient	*t*-statistic	Coefficient	*T*-test
Constant	−4.7831	−4.5600	−16.3635	−4.2160	−11.2033	−8.1400	−1.1273	−1.0470	0.8551	0.5050	−2.2152	−4.0080
LNADTMJ	0.4468	3.6770	0.7741	2.0870	0.9696	6.2430	–	–	–	–	–	–
LNADTMI	0.1949	1.8760	0.7020	3.0530	0.3906	3.7680	–	–	−0.5292	−1.9410	0.3086	3.7740
SHLWDMAJ	0.0751	1.2900	–	–	–	–	0.2058	2.0540	–	–	0.1040	1.9740
LIGHTMAJ	−0.5194	−2.2780	–	–	–	–			–	–	−0.5878	−2.4870
SPDLIMAJ/100	–	–	6.0488	1.7760	–	–	−4.0005	−1.9490	–	–	–	–
LTLMAJ	–	–	–	–	–	–	0.6982	3.6100	–	–	–	–
LTLMIN	0.4889	3.6990	–	–	–	–	–	–	–	–	–	–
RTLMAJ	–	–	–	–	0.4979	3.3490	–	–	0.7783	2.1240	–	–
TERNMIN	–	–	–	–	0.2513	1.3910	–	–	–	–	–	–
VIMAJ	–	–	–	–			0.1716	1.0340	–	–	–	–
Coefficients for error structure												
$\delta_1, \delta_2, \delta_4, \delta_7, \delta_{11,} \delta_{16}$	0.7671	7.8360	0.0000	–	0.0000	–	0.0000	–	0.0000	–	0.0000	–
$\delta_3, \delta_5, \delta_8, \delta_{12}, \delta_{17}$	–	–	0.2273	0.4430	0.0000	–	0.0000	–	0.0000	–	0.0000	–
$\delta_6, \delta_9, \delta_{13}, \delta_{18}$	–	–	–	–	0.6546	5.3940	0.0000	–	0.0000	–	0.0000	–
$\delta_{10}, \delta_{14}, \delta_{19}$	–	–	–	–	–	–	0.6877	2.2260	0.0000	–	0.0000	–
δ_{15}, δ_{20}	–	–	–	–	–	–	–	–	0.7354	1.4450	0.0000	–
δ_{21}	–	–	–	–	–	–	–	–	–	–	0.1077	0.4920
Goodness-of-fit measures												
LL(β)	−909.25											
LL(c)*	−1283.63											
# of parameters	27											
Adj. ρ^2	0.2706											
Sample size	165											

* LL(c) is measured by the log-likelihood value at convergence when only six constants are specified into the model.

The findings suggest that major road traffic volume contributes positively to the frequency of angle, head-on, and rear-end crashes. The same results are seen with respect to minor road traffic volume. In addition, an increase in traffic volume on the minor road was found to significantly increase pedestrian-involved crashes; it appears that a higher volume of traffic on the minor road may lead to greater conflicts with pedestrian movements or, it is also possible that pedestrians pay more attention to major road traffic and less attention to minor road traffic, thus leading to a higher involvement of pedestrians in crashes when minor road traffic is high.

It is to be noted that the variable representing signalization control at the intersection did not turn out to be statistically significant when traffic volume variables were included in the model. Although it would have been reasonable to retain the signalization variable in the models despite the statistical insignificance, this variable was not retained for a couple of reasons. First, it is likely that the effect of signalization on crash frequencies is captured by traffic volume variables, thus contributing to its statistical insignificance. Second, signalization is likely to be an endogenous variable because it is a function of traffic volumes at the intersection. As signalization is dependent on traffic volumes, it was considered prudent to omit that variable (but retain traffic volume variables, which are truly exogenous) to avoid any potential for endogeneity bias. It would have been ideal to estimate separate models for signalized and unsignalized intersections; however, the limited sample size in this study precluded the ability to parse the data set on this basis for model estimation purposes.

Larger shoulder widths are associated with higher frequencies for angle, sideswipe, and pedestrian-involved crashes. While this finding merits further investigation, it appears that drivers try to utilize the shoulders (when they are wide) to navigate through the intersection and around other vehicles or pedestrians that may be plying through the intersection. When drivers try to use the shoulder to pass waiting vehicles, there is a higher likelihood of being involved in a sideswipe, interfering with a pedestrian, or not noticing another turning vehicle resulting in an angle crash. This finding points to the need to re-evaluate the use of wide shoulders in intersections, and determine an optimal shoulder width to balance the tradeoff between collision types that are reduced by wider shoulders (head-on, run-off road) and those that are increased (angle, sideswipe, and pedestrian-involved).

The improved visibility provided by lighting appears to significantly decrease angle and pedestrian-involved crash frequencies. This finding is consistent with expectations; these are the two types of crashes that are likely to be most affected by the quality of light provided. It is interesting to note that the speed limit is significantly associated only with two crash type frequencies, and that too, in an opposite manner. While higher speed limits understandably lead to a higher frequency of head-on collisions (possibly more so at unsignalized intersections with no median barriers separating the two directions of travel), it is also found that higher speed limits lead to reduced sideswipe (same direction) crash frequencies. Higher speed limits may be associated with wider roadways, thus contributing to a reduction in sideswipe collisions.

The number of turning lanes significantly impacts crash frequencies by collision type. The number of left turn lanes on the major road positively impacts sideswipe (same direction) crash frequencies. It is not immediately clear why this might be the case, although it is conceivable that vehicles sideswipe one another when transitioning into left turn lanes at an intersection. The number of left turn lanes on the minor street leads to more angle crashes. It is likely that minor streets do not have protected left turn signal indications, even when the intersection is signalized. As the presence of left turn lanes is likely an indicator of a higher level of left turning movements (from the minor street), this can contribute to a higher number of angle crashes. Finally, the number of right turn lanes on the major roadway contributes positively to rear-end collision frequencies and sideswipe (opposite direction) frequencies. The former finding is very consistent with expectations. As vehicles slow down to turn right in the right turn lane, vehicles coming from behind may rear-end the vehicle in front leading to more rear-end crashes. With respect to sideswipe (opposite direction) crashes, it is not clear why the number of right turn lanes on the major road would contribute positively to such crashes. This finding merits further investigation.

The last two variables in the model system denote the terrain and the geometry of the roadway (grade and curve). As expected, rolling and mountainous terrain constitutes more adverse conditions from a safety standpoint. Such terrain contributes positively to rear-end collisions and it is rather surprising that this variable does not significantly impact other collision type frequencies. The grade and curve variable (VIMAJ) positively impacts sideswipe (same direction) crash frequencies. It appears that such roadway geometry leads to driver maneuvers resulting in sideswipe (same direction) crashes. For example, some drivers may attempt to navigate a curve with a larger turning radius, thus potentially interfering with drivers in an adjacent lane.

Overall the findings reported here are quite consistent with expectations and shed considerable light on the factors affecting crash frequencies by collision type. The table also provides estimates of the δ_i parameters corresponding to each equation (see Eq. (5) for the definition of δ_i). There are several δ_i parameters that are statistically insignificant; however, they are retained in the model results to show their values and to use all δ_i values in the computation of the elements of the error variance-covariance matrix. The goodness-of-fit statistics are quite consistent with what one might expect from an intersection crash frequency model of this nature.

For comparison purposes, a simultaneous equations model of crash frequencies that ignores cross-equation error correlation was also estimated. Table 4 shows the model estimation results for such a system of univariate Poisson (UVP) regression models. In general, it is found that the estimates of model coefficients differ in magnitude between the two model systems, although the differences are rather modest when viewed from a qualitative perspective. This finding is not surprising because both the UVP and MVP (multivariate Poisson) models offer consistent and unbiased estimates. The MVP model offers more statistically efficient (i.e., lower standard error) estimates of coefficients. With respect to goodness-of-fit measures, the adjusted likelihood ratio index for the MVP model is 0.2950, which is about 10% higher than that of the UVP model (0.2706). Thus, the MVP model offers gains in efficiency and an overall improvement in the fit of the model to the data; however, in this particular context, it is found that the model offers parameter estimates rather similar to those obtained from a UVP model.

6. Conclusions

A critical need exists to be able to estimate crash frequencies by collision types at intersections, primarily to support safety management programs and to enable improved methods for detecting safety deficiencies at intersections. While past research has modeled crash frequencies by collision type for intersections as a function of various explanatory factors, such efforts have largely been confined to the development and estimation of single-equation models where crash frequencies by collision type are modeled independently of one another. Recent advances in crash frequency modeling have pointed out the limitations of such approaches and the need to simultaneously model multiple crash frequencies in a unifying framework that explicitly accounts for overdispersion in the data and the presence of common unobserved factors (error covariances) that affect multiple crash frequencies.

In this particular paper, a simultaneous equations model of crash frequencies by collision type is developed and estimated for a sample of 165 rural intersections in 38 counties of Georgia. A multivariate Poisson (MVP) regression model with multivariate normal heterogeneity is formulated and presented in this paper. The model system is a flexible specification that accommodates overdispersion in the data and the presence of error covariances. Simulation approaches are used to compute the log-likelihood function (that involves multi-dimensional integrals corresponding to the number of dependent variables or crash frequency types being considered) and classic optimization algorithms are used to maximize the simulated log-likelihood function and estimate parameters. The crash types considered in this paper are rear-end, head-on, sideswipe (same direction), sideswipe (opposite direction), angle, and pedestrian-involved crashes. This constitutes a simultaneous equations system with six dimensions.

Model estimation results show that a host of roadway geometry, intersection control, and traffic volume characteristics significantly impact crash frequencies of various types. More importantly, the model estimation results show significant presence of error correlation across pairs of crash frequency equations in the model system. The only equation that did not show statistically significant error correlations is the one describing pedestrian-involved crashes. It is not clear if this particular result is generalizable or a manifestation of the nature of the data set used in this study. The study supports recent research in this area that calls for the simultaneous equations modeling of crash frequencies using multivariate count data models that account for overdispersion and flexible error structures. Having said that, however, a comparison of the MVP model estimation results with results obtained from a system of univariate Poisson (UVP) regression models that ignore cross-equation error correlation showed that parameter estimates do not differ substantially between the two model formulations. Thus, while there is a gain in efficiency and goodness-of-fit from a statistical standpoint, it appears that the gain in predictive capability may be more modest.

Although the differences between the two model systems are modest, this does not mean that one does not need to consider using a multivariate Poisson (MVP) regression model that accounts for cross-equation error correlation. First, the MVP model offers a strong theoretical basis for safety modeling in that it is capable of accounting for the presence of potential common unobserved factors that affect multiple crash types. Second, the MVP model estimation results in this study showed significant error correlations; this means that the MVP model estimates are statistically more efficient and as such, they are likely to provide more accurate coefficient estimates than a less statistically efficient UVP model. Finally, to minimize the potential error that might result from using a UVP model, it would be prudent to perform a comparison of models such as that presented here prior to proceeding with the use of a UVP model in a particular application context.

Future research efforts in this area should further explore the implications of using MVP-type simultaneous equations model formulations from a crash prediction perspective. Although the results in this study suggest that there may not be much difference from a prediction standpoint (because the coefficient estimates are quite similar), it is not clear whether this finding is generalizable and applicable to all crash contexts. Also, it would be useful to estimate model systems separately for signalized and unsignalized intersections.

References

Abdel-Aty, M., Keller, J., 2005. Exploring the overall and specific crash severity levels at signalized intersections. Accident Analysis and Prevention 37, 417–425.

Abdel-Aty, M., Nawathe, P., 2006. A novel approach for signalized intersection crash classification and prediction. Advances in Transportation Studies, an International Journal Section B3.

Aptech, 2006. GAUSS 8.0. Aptech Systems, Maple Valley, Washington.

Bhat, C.R., 2003. Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. Transportation Research Part B 37 (9), 837–855.

Cameron, C., Trivedi, P., 1986. Econometric models based on count data: comparisons and applications of some estimators and tests. Journal of Applied Econometrics 1 (1), 29–53.

Golob, T.F., Recker, W.W., 1987. An analysis of truck-involved freeway accidents using log-linear modeling. Journal of Safety Research 18, 121–136.

Jonsson, T., Ivan, J.N., Zhang, C., 2007. Crash prediction models for intersections on rural multilane highways: differences by collision type. In: Presented at the 86th annual meeting of the transportation research board, National Research Council, Washington, DC.

Kim, D., Washington, S., Oh, J., 2006. Modeling crash types: new insights into the effects of covariates on crashes at rural intersections. Journal of Transportation Engineering 132 (4), 282–292.

Kim, D., Lee, Y., Washington, S., Choi, K., 2007. Modeling crash outcome probabilities at rural intersections: application of hierarchical binomial logistic models. Accident Analysis and Prevention 39, 125–134.

Kim, D.-G., 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 38 (6), 1094–1100.

Lee, J., Mannering, F.L., 2002. Impact of roadside feature on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accident Analysis and Prevention 34, 149–161.

Lord, D., Washington, S.P., 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, 35–46.

Lyon, C., Oh, J., Persaud, B., Washington, S., Bared, J., 2003. Empirical investigation of the IHSDM accident prediction algorithm: rural intersections. Transportation Research Record, Journal of the Transportation Research Board 1840, 78–86.

Ma, J., Kockelman, K.M., 2006a. Crash frequency and severity modeling using clustered data from washington state. In: Proceedings of the IEEE Intelligent Transportation Systems Conference 2006, Toronto, Canada, September 17–20.

Ma, J., Kockelman, K.M., 2006b. Bayesian multivariate poisson regression for models injury count, by severity. Transportation Research Record, Journal of the Transportation Research Board 1950, 24–34.

Ma, J., Kockelman, K.M., Damien, P., 2008. A multivariate poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. Accident Analysis and Prevention 40, 964–975.

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, 471–482.

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

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

Oh, J., Lyon, C., Washington, S., Persaud, B., Bared, J., 2003. Validation of the FHWA crash models for rural intersections: lessons learned. Transportation Research Record, Journal of the Transportation Research Board 1840, 41–49.

Oh, J., Washington, S., Choi, K., 2004. Development of accident prediction models for rural highway intersections. Transportation Research Record, Journal of the Transportation Research Board 1897, 18–27.

Ossenbruggen, P.J., Pendharkar, J., Ivan, J.N., 2001. Roadway safety in rural and small urbanized areas. Accident Analysis and Prevention 33, 485–498.

Pai, C., Saleh, W., 2007a. An analysis of motorcyclist injury severity under various traffic control measures at three-legged junctions in the UK. Safety Science 45, 832–847.

Pai, C., Saleh, W., 2007b. Modeling motorcyclist injury severity by various crash types at T-junctions in the UK. Safety Science. doi:10.1016/j.ssci.2007.07.005.

Park, E.S., Lord, D., 2007. Multivariate poisson-lognormal models for jointly modeling crash frequency by severity. Transportation Research Record, Journal of the Transportation Research Board 2019, 1–6.

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

Shankar, V.N., Milton, J.C., Mannering, F.L., 1997. Modeling statewide accident frequencies as zero-inflated probability processes: an empirical inquiry. Accident Analysis and Prevention 29, 829–837.

Shankar, V.N., Ulfarsson, G.F., Pendyala, R.M., Nebergall, M.B., 2003. Modeling crashes involving pedestrians and motorized traffic. Safety Science 41, 627–640.

Train, K., 2002. Discrete Choice Methods with Simulation. Cambridge University Press.

Wang, X., Abdel-Aty, M., 2006. Temporal and spatial analyses of rear-end crashes at signalized intersections. Accident Analysis & Prevention 38 (6), 1137–1150.

Yan, X., Radwan, E., Abdel-Aty, M., 2005. Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model. Accident Analysis and Prevention 37, 983–995.

Ye, X., Pendyala, R.M., 2007. Analysis of the relationship between timing and duration of maintenance activities using alternative joint discrete–continuous

modeling systems. In: CD-ROM of the 86th Annual Meeting of the Transportation Research Board. National Research Council, Washington, DC.
Zajac, S.S., Ivan, J.N., 2003. Factors influencing injury severity of motor vehicle-crossing pedestrian crashes in rural Connecticut. Accident Analysis and Prevention 35, 369–379.
Zhao, Y., Kockelman, K.M., 2001. Household vehicle ownership by vehicle type: application of a multivariate negative binomial model. In: CD-ROM of the 81st Annual Meeting of the Transportation Research Board. National Research Council, Washington DC.