Accident Analysis and Prevention 50 (2013) 73–82

Contents lists available at SciVerse ScienceDirect

Accident Analysis and Prevention

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

Modeling crash frequency and severity using multinomial-generalized Poisson model with error components

Yu-Chiun Chiou*, Chiang Fu

Institute of Traffic and Transportation, National Chiao Tung University, 4F, 118, Sec. 1, Chung-Hsiao W. Rd., Taipei 100, Taiwan

ARTICLE INFO

Article history:
Received 16 October 2011
Received in revised form 25 March 2012
Accepted 26 March 2012

Keywords:
Crash frequency
Crash severity
Multinomial-generalized Poisson
Error components

ABSTRACT

Since the factors contributing to crash frequency and severity usually differ, an integrated model under the multinomial generalized Poisson (MGP) architecture is proposed to analyze simultaneously crash frequency and severity—making estimation results increasingly efficient and useful. Considering the substitution pattern among severity levels and the shared error structure, four models are proposed and compared—the MGP model with or without error components (EMGP and MGP models, respectively) and two nested generalized Poisson models (NGP model). A case study based on accident data for Taiwan's No. 1 Freeway is conducted. The results show that the EMGP model has the best goodness-of-fit and prediction accuracy indices. Additionally, estimation results show that factors contributing to crash frequency and severity differ markedly. Safety improvement strategies are proposed accordingly.

1. Introduction

To improve traffic safety, numerous statistical models have been developed that identify factors contributing to crash frequency and severity. Most identify risk factors for either crash frequency or severity independently. When modeling crash frequency (the number of accidents on roadway segments or at intersections over a specified period), a considerable number of studies have used various methodological approaches. Due to the discrete and non-negative integer character of accident counts, count-data models such as the Poisson model (e.g., Jones et al., 1991; Miaou, 1994; Shankar et al., 1997), negative binomial model (e.g., Hadi et al., 1995; Shankar et al., 1995; Poch and Mannering, 1996; Milton and Mannering, 1998; Lord, 2006; Malyshkina and Mannering, 2010), Poisson lognormal model (e.g., Miaou et al., 2005; Lord and Miranda-Moreno, 2008), Gamma model (e.g., Oh et al., 2006), generalized Poisson model (e.g., Dissanayake et al., 2009; Famoye et al., 2004) as well as zero-inflated modeling and other flexible modeling techniques (e.g., Abdel-Aty and Radwan, 2000; Wang and Abdel-Aty, 2008; Park and Lord, 2009; Anastasopoulos and Mannering, 2009; see Lord and Mannering, 2010 for elaborate and complete reviews) have been applied to model crash counts.

Crash frequencies are commonly collected by severity on relatively homogenous roadway segments, supporting the development of crash count models. Thus, crash data are typically classified according to severity (e.g., property damage only, injury, and fatality) or collision type (e.g., rear-end, head-on, sideswipe, and right angle). With this data segmentation, separate severity–frequency models are developed for each accident severity level. In this way, a series of negative binomial accident frequency models were developed for each crash severity level to predict the number of accidents at each severity level on roadway segments. Unfortunately, such an approach can generate a statistical problem in that interdependence due to latent factors likely exists across crash rates at different severity levels for a specific roadway segment (Ma et al., 2008). For example, an increase in number of accidents that are classified as having a certain severity level is also associated with changes in the number of accidents that are classified with other severity levels, setting up a correlation among various injury-outcome crash frequency models (Lord and Mannering, 2010).

Considerable research effort has focused on modeling accident severity from an individual perspective using such methodological approaches as logistic regression (e.g., Lui et al., 1988; Yau, 2004), bivariate models (e.g., Saccomanno et al., 1996; Yamamoto and Shankar, 2004), the multinomial and nested logit structures to evaluate accident-injury severities (e.g., Shankar et al., 1996; Chang and Mannering, 1999; Carson and Mannering, 2001; Lee and Mannering, 2002; Ulfarsson and Mannering, 2004; Khorashadi et al., 2005), and the discrete ordered probit model (e.g., O'Donnell and Connor, 1996; Duncan et al., 1998; Renski et al., 1999; Kockelman and Kweon, 2002; Khattak et al., 2002; Kweon and Kockelman, 2003; Abdel-Aty, 2003). For more details on accident severity models may refer to Savolainen et al. (2011).

Although these models have been applied by a number of researchers with a considerable success, Milton et al. (2008) indicated that these studies relied heavily on detailed data in individual accident reports and they have been proved to be difficult to use

* Corresponding author. Tel.: +886 2 23494940; fax: +886 2 23494953.
E-mail address: ycchiou@mail.nctu.edu.tw (Y.-C. Chiou).

0001-4575/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.aap.2012.03.030

in safety programming because a large number of event-specific explanatory variables need to be estimated to produce useable severity forecasts. Moreover, significant contributory factors in the model are usually not closely related to traffic management strategies, roadway geometrics, and weather-related factors; therefore, the corresponding countermeasures are difficult to propose accordingly. Furthermore, as different data scales are used by frequency models and the severity model, integration is extremely difficult.

Obviously, crash frequency and severity are two key indices that measure risk for a roadway segment. Either one only generates partial insights for crash risk. Increased scope and in-depth insights cannot be obtained without considering both indices together. Thus, two possible integrated modeling approaches were attempted. The first approach uses a conventional frequency model to predict total number of crashes and a severity model, such as the multinomial logit model, nested logit model, ordered probit model, or mixed logit model, to predict aggregate severity probability (e.g., Yamamoto et al., 2008; Kim et al., 2008; Milton et al., 2008). However, the assumption that crash frequency and severity are mutually independent still exists. The second approach applies multivariate regression models to predict crash frequencies for different severity levels. Multivariate regression models simultaneously develop crash frequency models by severity (Bijleveld, 2005; Ma and Kockelman, 2006; Song et al., 2006; Park and Lord, 2007; Ma et al., 2008; Aguero-Valverde and Jovanis, 2009; El-Basyouny and Sayed, 2009; Ye et al., 2009) to overcome the correlation problem among crash frequencies at different severity levels. However, this approach requires a complex estimation procedure combined with a subjectively preset correlation matrix of severity levels, making field validation very difficult.

Another drawback of the multivariate modeling approach is its inability to grasp associated changes related to severity and frequency variables only. If one fails to observe separately the effects of factors contributing to crash frequency and severity, the multivariate model may be partly limited for practical program evaluation. An appealing idea is to view risk factors according to their accident descriptive components (i.e., severity and frequency) individually under an integrated framework. However, expected difficulties arise when analyzing subjects and procedures. Consequently, using a conceptual model combining both crash frequency and severity is worthwhile.

Thus, this paper aims to develop a novel multinomial generalized Poisson (MGP) model to simultaneously model crash frequency (count data) and severity (ratio data). Furthermore, the proposed model considers the substitution pattern among severity levels and constructs a shared error structure as a correlation matrix through error components specified under an integrated model framework. A case study of Taiwan freeway crash data is utilized to assess the applicability of the proposed model. The remainder of this paper is organized as follows. Section 2 presents the proposed MGP model. Section 3 addresses data collection and descriptive statistics of the accident dataset for Taiwan's No. 1 Freeway. Section 4 presents model estimation results and comparisons. Section 5 discusses safety implications based on estimation results. Section 6 gives concluding remarks and suggestions to further research.

2. The proposed models

The MGP model is an extension of the multinomial-Poisson (MP) regression model (Terza and Wilson, 1990). In the context of crash frequency and severity modeling, we assume that accidents can be classified into a finite number of clusters according to severity levels and that the frequency of each severity level follows a conditional multinomial distribution, which is expressed as follows:

$$f\left(Y \middle| \sum_{j=1}^{J} y_j = N\right) = \frac{N! \prod_{j=1}^{J} \pi_j^{y_j}}{\prod_{j=1}^{J} y_j!} \tag{1}$$

where $f(\cdot)$ is the conditional probability of Y; $Y = [y_1, y_2, \ldots y_j, \ldots, y_J]$ and $\sum_{j=1}^{J} y_j = N$; $y_j = 0, 1, 2, \ldots, \infty$, for $j = 1, 2, \ldots, J$, is a random vector representing the observed crash counts of segment t within a given period (e.g., 1 year) at severity level j; J is the total number of severity levels determined in advance; $\pi_1, \pi_2, \ldots, \pi_J$ are multinomial probabilities of severity levels $1, 2, \ldots, J$, respectively; $\pi_j = y_j/N$ and $\pi_1 + \pi_2 + \ldots + \pi_J = 1$; and N is the total number of accidents across different severity levels of segment m within a given period. Thus, the conditional multinomial distribution can be used to determine crash frequencies at various severity levels, i.e., $y_1, y_2, \ldots, y_J$, given total number of accidents, N. Furthermore, the joint probability of these crash frequencies $h(y_1, y_2, \ldots, y_J)$ can be expressed as the product of conditional probability and marginal probability:

$$h(y_1, y_2, \ldots, y_J) = f\left(Y \middle| \sum_{j=1}^{J} y_j = N\right) \cdot g\left(\sum_{j=1}^{J} y_j = N\right) \tag{2}$$

where $g(\cdot) = g\left(\sum_{j=1}^{J} y_j = N\right)$ is the marginal probability of crash counts. Terza and Wilson (1990) assumed that the marginal (unconditional) probability has the following Poisson distribution:

$$g(\cdot) = \frac{\lambda^N \exp(-\lambda)}{N!} \tag{3}$$

where $g(\cdot)$ is the probability that N accidents occurred, and λ is the expected number of accidents. For estimation purposes, λ is usually specified as

$$\lambda = \exp(\beta' X) \tag{4}$$

where X and β' are vectors of explanatory variables and estimated parameters, respectively. The formulation of the multinomial Poisson (MP) model is then derived by substituting Eqs. (1) and (3) into Eq. (2).

The Poisson model assumes that variance equals mean. If observed data exhibit over-dispersion (under-dispersion), this assumption does not hold. This leads to estimation inefficiency because inference was invalidated by unreliable estimated standard errors. We can relax this assumption using the generalized Poisson (GP) model (Famoye et al., 2004; Dissanayake et al., 2009). The probability function of total accidents at any segment, N, can be written as Eq. (5):

$$g(\cdot) = \left(\frac{\lambda}{1+\eta\lambda}\right)^N \frac{(1+\eta N)^{N-1}}{N!} \exp\left(\frac{-\lambda(1+\eta N)}{1+\eta\lambda}\right) \tag{5}$$

where η is the dispersion parameter. If $\eta > 0$, the GP model indicates the over-dispersion feature in the empirical data. If $\eta = 0$, the probability function degenerated to the Poisson model. In contrast, if $\eta < 0$, the GP model denotes the under-dispersion feature in the empirical data. All other involved arguments associated with Eq. (5) are as defined previously. The mean and variance of N are represented by Eqs. (6) and (7), respectively:

$$E(N|X) = \lambda \tag{6}$$

$$V(N|X) = \lambda(1+\eta\lambda)^2 \tag{7}$$

According to Eq. (6), the probability function in Eq. (5) degenerates into the original Poisson model as $\eta = 0$. Hence, the GP model is a generalized Poisson model. Interested readers can refer to Famoye (1993) for detailed proofs. In accordance with the derivation by

Terza and Wilson (1990), formulation of the MGP model can be derived by substituting Eqs. (1) and (5) into Eq. (2) as follows:

$$h(y_1, y_2, \ldots y_J) = \frac{\prod_{j=1}^{J} \pi_j^{y_j}}{\prod_{j=1}^{J} y_j!} \left[1 + \eta \left(\sum_{j=1}^{J} y_j\right)\right]^{\sum_{j=1}^{J} y_j - 1} \times \left(\frac{\lambda}{1+\eta\lambda}\right)^{\sum_{j=1}^{J} y_j} \exp\left[\frac{-\lambda\left(1+\eta\sum_{j=1}^{J} y_j\right)}{1+\eta\lambda}\right] \quad (8)$$

where $\lambda = \sum_{j=1}^{J} \lambda_j$ is expected total accidents. $\lambda_j = \lambda\pi_j$ is expected crash count at the jth severity level. π_j is the probability of severity level j.

We assume probability can be determined by the multinomial logit (MNL) model:

$$\pi_i = \frac{\exp(s_i)}{\sum_{j=1}^{j} \exp(s_j)} \quad (9)$$

where $S_j = \gamma' Z + v_j$ is a linearly additive function for measuring risk of severity level j; Z is a vector of non-random explanatory variables, such as roadway geometrics, traffic factors, land use, and weather condition; γ is a vector of unknown parameters; v_j is a random error term, which we assume is a Gumbel distribution across all observations (McFadden, 1981).

Eq. (8) is a straightforward equation for integrating the two descriptive model components (i.e., the frequency and severity model), and the probability of accident frequency at road segment Σy_j is the weighted sum of crash counts of all severity levels on the same segment over a unit time span. However, an important property of the MNL model is its independence from irrelevant alternate severity outcomes. This independence may be a major concern if some crash-injury severity levels share unobserved effects. To overcome partly such a restriction, a nested logit (NL) model can group some possible levels that share unobserved effects into conditional nests (Koppelman and Wen, 1998). The NL model partitions a severity outcome set into several nests, each containing correlated levels. The NL model can be expressed as

$$\pi_i = \frac{\exp(s_i/\theta_k)\left(\sum_{j \in B_k} \exp(s_j/\theta_k)\right)^{\theta_k - 1}}{\sum_{l=1}^{L}\left(\sum_{j \in B_l} \exp(s_j/\theta_l)\right)^{\theta_l}} \quad (10)$$

where B_k represents nest k, which is a subset containing correlated outcomes with respect to crash severity; $(1-\theta_k)$ is a correlation measure of unobserved factors within nest k; and θ_k is in the range of 0–1. As the value of θ_k decreases, the strength of the correlation within the nest increases. Notably, θ_k is also called the inclusive value representing the degree of correlation among alternate severity levels within nest k. If $\theta_k = 1$, the NL model becomes an MNL model. If θ_k is equal to zero, perfect correlation is implied among the severity levels in the nest, indicating that the process by which crashes result in particular severity levels is deterministic.

Since related studies (Shankar et al., 1996; Lee and Mannering, 2002; Savolainen and Mannering, 2007; Savolainen et al., 2011) have revealed that two nearer accident severity levels such as "property damage only" and "possible injury", or "disabling injury" and "fatality", may tend to have strong correlations due to ordinal nature of crash severity data. Such problem violated MNL model's independence of irrelevant alternatives (IIA) property resulted in biased parameter estimates. In that case, an NL model is preferred. When the nested structure exists in π_j, the MGP model evolves into a flexible nested generalized Poisson (NGP) model, solving the problem of substitution patterns among severity levels.

Based on the work by Ye et al. (2009), specifying a partial or full error components structure may be an innovative choice comparing to the formulation of correlation matrix. The error components structure is considered in the expected frequency λ and severity function of three severity levels s_j (j = 1, 2, 3), which include fatality (s_1), injury (s_2), and property damage only (s_3). In this study, four random coefficients (σ_i) are specified to the frequency function λ and severity function s_j (j = 1, 2, 3) to model the following partial covariance structure:

$$\lambda = \exp(\beta' X + \varepsilon + \sigma_1 u) \quad (11)$$

$$s_1 = \gamma' Z + v_1 + \sigma_2 u \quad (12)$$

$$s_2 = \gamma' Z + v_2 + \sigma_3 u \quad (13)$$

$$s_3 = \gamma' Z + v_3 + \sigma_4 u \quad (14)$$

where u is an independent random variable, which is normally distributed; and σ_j are their corresponding coefficients to be estimated. To simplify the number of estimated parameters in the error components structure, $j-1$ standard deviation parameters are identified by subjectively setting one parameter equals to 1. We assume ε and v_j have Generalized Poisson and Gumbel distributions, respectively. Thus, the cumulative probability functions conditional on this random variable $h(y_1, y_2, \ldots, y_J | u)$ are expressed as

$$h(y_1, y_2, \ldots, y_J | u) = \frac{\prod_{j=1}^{J} \pi_j(Z, u)^{y_j}}{\prod_{j=1}^{J} y_j!} \left[1 + \eta \left(\sum_{j=1}^{J} y_j\right)\right]^{\sum_{j=1}^{J} y_j - 1} \times \left(\frac{\lambda(X, u)}{1+\eta\lambda(X, u)}\right)^{\sum_{j=1}^{J} y_j} \exp\left[\frac{-\lambda(X, u)\left(1+\eta\sum_{j=1}^{J} y_j\right)}{1+\eta\lambda(X, u)}\right] \quad (15)$$

To distinguish it from Eq. (8), Eq. (15) is called the multinomial generalized Poisson model with error components (EMGP). The unconditional cumulative probability function of the random multinomial can then be derived by integrating the conditional cumulative probability function over the distributional domain of the specified random variable:

$$H(y_1, y_2, \ldots, y_J) = \int h(y_1, y_2, \ldots, y_J) r(u) du \quad (16)$$

where $r(u)$ is the probability density function of u. As this integral does not have a neat closed-form expression, the unconditional probability function may be approximated by the following simulated probability function $H^s(y_1, y_2, \ldots, y_J)$:

$$H^s(y_1, y_2, \ldots, y_J) = \frac{1}{R}\sum_{r=1}^{R} h(y_1, y_2, \ldots, y_J | u_r) \quad (17)$$

The estimation procedure of the EMGP model typically follows the simulation-based maximum likelihood method, using Halton draws, which have a more efficient distribution of draws for numerical integration than purely random draws (Bhat, 2003; Train, 2003). In many empirical settings, the number of draws for simulation is determined according to the number of estimated variables, the complexity of model specification, and sample sizes. For accuracy purposes, the estimation results of proposed models are presented for 200 Halton draws. The estimated parameters do not vary markedly once the number of replications exceeds 150 in the empirical case (see Train (2003) for further technique details and simulation issues).

Table 1
Descriptive statistics of 124 segments.

Variable	Description	Mean	SE	Min	Max
Crash counts					
Y_3	Property damage only (PDO)	70.4	63.7	3.0	284.0
Y_2	Injury	4.1	3.5	0.0	17.0
Y_1	Fatality	0.5	0.8	0.0	4.0
N	Total	75.1	65.4	6.0	290.0
Crash counts per km (=crash counts/segment length)					
Z_3	PDO crashes per km	16.4	16.3	1.2	83.5
Z_2	Injury crashes per km	0.8	0.6	0.0	2.8
Z_1	Fatal crashes per km	0.1	0.2	0.0	1.5
N	Total crashes per km	17.3	16.6	2.2	85.3
Freeway geometrics					
GL	Segment length (km)	5.9	4.4	0.8	22.4
GN	Number of lanes	2.6	0.6	2.0	4.0
GC	Curvature (‰)	0.7	1.1	0.0	7.1
GU	Maximum upward slope (%)	1.3	2.1	0.0	13.7
GD	Maximum downward slope (%)	1.2	1.4	0.0	5.2
GO	Clothoid curve value (thousand degrees)	0.9	0.9	0.0	3.2
GS	Speed limit (*GS* = 1 for 110 km, *GS* = 0 else)	0.5	0.5	–	–
Rainfall					
RF	Annual rainfall (hundred millimeters)	21.1	7.5	11.1	38.9
Average daily traffic					
TTV	Total traffic (thousand passenger car units)	69.2	28.7	10.8	157.0
PSV	Percentage of small vehicles (%)	51.4	10.1	31.9	70.5
PLV	Percentage of large vehicles (%)	23.8	4.2	15.5	34.2
PKV	Percentage of trailer-tractors (%)	24.8	8.7	9.2	41.0
Freeway facilities (dummy variables, yes = 1; no = 0)					
PT	Presence of toll station	0.2	0.4	–	–
PR	Presence of rest area	0.1	0.3	–	–
PS	Presence of posted speed camera	0.3	0.5	–	–
Neighborhood (dummy variables, yes = 1; no = 0)					
AM	Adjacent to metropolitan	0.5	0.5	–	–
AP	Adjacent to airport, seaport or industry area	0.2	0.4	–	–

3. Data

The accident dataset for Taiwan's No. 1 Freeway in 2005 was collected. Data were from three sources: (1) the accident database; (2) geometric documents; and (3) the traffic database. The accident database, maintained by the National Highway Police Bureau (NHPB), contains accident information, such as crash severity, location and time of an accident, and number and types of vehicles involved. Geometric data were digitalized according to the official as-constructed freeway drawings, including number of lanes, slope, curvature degree, and clothoid curve value.

Taiwan's No. 1 Freeway runs north–south, is 373.3 km long, and has 63 interchanges. To facilitate model estimation, a study segment is formed by two adjacent interchanges. By considering north- and south-bound directions separately, 124 analytical samples are obtained. Since the lengths of segments remarkably differ, to better reflect the crash risk, the dependent variable is presented by the crash counts divided by the segment length (*GL*). The traffic database, maintained by the National Freeway Bureau (NFB), includes traffic volume, speed and occupancy of three vehicle types detected by loop detectors on a basic segment or on-ramp (small vehicles, large vehicles and trailer-tractors). Considering the values of passenger car equivalent (pce) of three types of vehicles, the total traffic at each road segment are measured in passenger car units. Table 1 gives descriptive statistics for these segments. The mean and standard deviation of accidents differ in either total accident cases or those cases at various severity levels, suggesting that the potential problem of over- or under-dispersion may cause inefficient model estimation and bias.

Table 2 presents the cross-tabulation of crash frequency and severity. In total, 67 (1%) fatal accidents occurred in 2005, and 8735 (94%) accidents were property-damage-only accidents. Moreover, all 124 segments had at least one PDO accident, while only 47 segments (38%) had at least one fatal accident.

4. Results

According to model formulation in Section 2, four possible models can be estimated: the MGP model with or without considering error components among severity levels, namely, the MGP and EMGP models. Two NGP models considering different nested structure among severity levels, namely, the NGP1 (nesting two severe severity levels: fatality and injury) and NGP2 (nesting two minor severity levels: PDO and injury) are tested. Unfortunately, according to the estimated inclusive values for two NGP models, we could not find any possible correlated nesting between two severe accident levels (i.e., injury and fatality with t-ratio of $\theta_k = 0.634$) nor two non-severe accident levels (i.e., injury and PDO with t-ratio of $\theta_k = 0.299$).

Tables 3 and 4 compare performance indices and prediction accuracy among models, respectively. In Table 3, the goodness-of-fit indices, including number of significant variables, means and standard deviations of predicted accident counts, η value, log-likelihood values, adjusted rho-square, and the Bayesian information criterion (*BIC*) are compared. The estimation results show that the model with the error component (i.e., the EMGP model) perform better than those models that do not consider error component (i.e., the MGP and NGP models), and two nested models do not perform better than the multinomial models (i.e., the MGP and EMGP models) in terms of BIC values. Additionally, according to the estimated dispersion parameter (η) of EMGP, which is decreasing from 0.082 to 0.062, the associated asymptotically t statistics are significantly different from zero as well, indicating that empirical data have a slightly over-dispersion problem. In addition,

Table 2
Cross-tabulation of crash frequency and severity.

	Severity level			Total crash
	PDO	Injury	Fatality	
Number of crashes				
Crash counts	8735 (94%)	509 (5%)	67 (1%)	9311 (100%)
Crash counts per km	2038 (95%)	96 (4%)	13 (1%)	2147 (100%)
Number of segments				
With at least one such crash	124 (100%)	110 (89%)	47 (38%)	
Without any such crash	0 (0%)	14 (11%)	77 (62%)	
Total	124 (100%)	124 (100%)	124 (100%)	

Table 3
Comparisons of goodness-of-fit among the models.

Models	Goodness of fit					
	K[a]	Crash (std.)[b]	η[c]	$LL(\beta)$	$Adj\text{-}\rho^2$ [d]	BIC[e]
Multinomial generalized Poisson (MGP)	26	18.33 (14.53)	0.082	−1108.130	0.147	2341.588
Nested generalized Poisson (NGP1)	27	18.33 (14.53)	0.082	−1108.095	0.147	2346.339
Nested generalized Poisson (NGP2)	27	18.33 (14.52)	0.082	−1106.496	0.148	2343.139
Multinomial generalized Poisson with error components (EMGP)	29	16.39 (12.76)	0.062	−1037.935	0.201	2215.659

[a] K: number of significant variables under $\alpha = 0.1$ level.
[b] Crash: mean predicted crash counts.
[c] η: dispersion parameter.
[d] $Adj\text{-}\rho^2$: *rho-square* adjusted in compassion with Null model (with three constants for crash severity and a single constant for generalized Poisson model).
[e] $BIC = -2 \times LL(\beta) + K \times \text{Ln}\, N$.

Table 4
Comparisons of prediction accuracy among the models.

Severity	Accuracy	Model				Actual
		MGP	NGP1	NGP2	EMGP	
Fatality	Crash (%)	0.12 (0.82%)	0.12 (0.83%)	0.12 (0.82%)	0.10 (0.76%)	0.10 (0.82%)
	MAPE	0.319	0.321	0.333	0.307	–
	RMSE	0.200	0.200	0.205	0.198	–
Injury	Crash (%)	1.06 (7.48%)	1.06 (7.48%)	1.06 (7.49%)	0.90 (7.38%)	0.78 (7.48%)
	MAPE	0.816	0.816	0.814	0.695	–
	RMSE	0.746	0.746	0.748	0.654	–
PDO	Crash (%)	17.15 (91.70%)	17.15 (91.70%)	17.14 (91.69%)	15.39 (91.86%)	16.44 (91.70%)
	MAPE	0.698	0.698	0.698	0.617	–
	RMSE	13.451	13.450	13.447	13.111	–

Note: The percentages of the predicted crash counts at three severity levels are given in parentheses.

specifying the error component can mitigate variation in η. Hence, the estimated η of the EMGP model is lower than that of the MGP and two NGP models, but the inclusion of error components could not perfectly resolve the over-dispersion problem.

Table 4 compares the prediction accuracy of the four models by mean absolute percentage error (MAPE) and root-mean-square error (RMSE). The EMGP model performs best in comparison with other models, although all four models achieve relatively high prediction accuracy (Table 4).

For simplicity, only estimated parameters of two extreme models are reported and compared in Tables 5 and 6, respectively. This study sets $\alpha = 0.10$ as the variable selection criterion to avoid an excessive number of non-stable and insignificant variables adversely affecting efficiency in calculating numerical values and convergence results. Therefore, the potential variables, annual rainfall (*RF*) and percentage of trailer-tractors (*PKV*), are removed due to their insignificant effects. This study also tests all possible relationships among variables, including linear, squared, exponential, and natural log relationships.

5. Discussions

According to estimation results of the MGP and EMGP models (Tables 5 and 6), all significant variables are almost the same with a relatively similar magnitudes; however, variables of the EMGP model typically have more significant effects in terms of the t statistic, again demonstrating the superior performance of the EMGP model. Thus, only estimation results of the EMGP model are discussed below.

Only two variables of maximum downward slope (*GD*) and adjacent to metropolitan (*AM*) have significant effects on both crash frequency and severity, while other variables of crash frequency and severity model components are all different, suggesting that the factors contributing to crash frequency and severity differ.

First, in terms of geometric variables, maximum downward slope (*GD*) are significantly tested in both frequency and severity model components, while number of lanes (*GN*), exponential of maximum upward slope (*GU*), clothoid curve value (*GO*), and speed limit (*GS*) only significantly contribute to crash severity and curvature (*GC*) only affect crash frequency. The *GN* reduces PDO crashes but results into more severe crashes, indicating that more number of lanes may cause severe accidents. The exp (*GU*) has a negative coefficient on fatal crashes because drivers tend to drive at a lower speed on an upward-sloped segment and then largely mitigate the severity of crashes. Contrarily, both *GD* and *GO* have positive coefficients associated with the fatal and injury crashes, implying the crashes at high downward-sloped segments and curved transition curves are more severe. The *GS* has a positive effect on injury

Table 5
Model results of the multinomial generalized Poisson (MGP).

Variable		Severity level					
		Fatality		Injury		PDO	
		Para.	*t*-Stat	Para.	*t*-Stat	Para.	*t*-Stat
Logit crash severity model component							
Constant	–		1.407	2.147	5.330	8.113	
Freeway geometrics							
GN	Number of lanes	–		–		−0.259	−4.053
Exp(*GU*)	Exponential of maximum upward slope	−0.466	−2.097	–		–	
GD	Maximum downward slope	0.306	4.034	0.183	6.861	–	
GO	Clothoid parameter	0.250	2.341	0.177	4.158	–	
GS	Speed limit	–		0.280	3.017	–	
Traffic characteristics							
TTV	Total traffic	−0.709	−2.501	−0.424	−4.169	–	
PLV	Percentage of large vehicles	4.604	5.755	4.604	5.755	–	
Neighborhood							
AM	Adjacent to metropolitan	–		–		0.271	3.276
Freeway facilities							
PS	Presence of posted speed camera	–		−0.685	−2.880	−0.489	−2.175
PR	Presence of rest area	−1.361	−2.990	−0.321	−2.546	–	

Variable		Para.	*t*-Stat
Generalized Poisson crash frequency model component (for all severity levels)			
Constant		1.237	3.465
η		0.082	8.407
Freeway geometrics			
GC	Curvature	0.151	2.238
GD	Maximum downward slope	−0.555	−4.099
GD^2	Square of maximum downward slope	0.054	1.876
Traffic characteristics			
PSV	Percentage of small vehicles	3.050	4.291
Neighborhood			
AM	Adjacent to metropolitan	0.504	3.442
AP	Adjacent to airport, seaport or industry area	0.498	2.565
PT	Presence of toll station	−0.419	−2.568

Goodness of fit measures	
Log-likelihood (Null model)	−1298.488
Log-likelihood (Full model)	−1108.130
Adj-ρ^2	0.147
Samples	124

Note: Null model: with three constants for crash severity (market share) and a single constant for generalized Poisson model.

crashes because drivers tend to increase their speed at the segments with a higher speed limit, increasing accident severity once the accident occurred.

The *GC* affects crash frequency, suggesting that a high freeway curvature coefficient increases accident frequency. The *GD* has a polynomial effect (a negative linear effect and a positive squared effect) on accident frequency and a linear positive effect on accident severity (only for fatality and injury). By taking a derivative term of a variable, a 1° increase in *GD* has a marginal effect of increasing crash frequency by $-0.448 + 0.072 \times GD$, suggesting that a slight downward slope may reduce accident frequency. However, once a slope's grade exceeds 6.22%, crash frequency increases rapidly, suggesting that an abrupt downward slope significantly contributes to a reduced number of PDO crashes and, in turn, increases accident severity. As slope increases, driver awareness increases, reducing accident frequency for gentle slopes. However, when a slope exceeds a threshold (6.22% in this study), stopping becomes increasingly difficult, resulting in severer accident frequency.

In terms of traffic characteristics, the *TTV* has negative effects on two severe severity levels, implying the crash severity can be lowered at the segments with high traffic flow because of lower travel speed caused by traffic congestion. However, the *PLV* has relatively high effects on two severe crashes, suggesting the higher percentage of large vehicles, the more severe of the crashes. Additionally, the *PSV* has a positive effect on crash frequency. As the percentage of small vehicles increases, the percentage of large vehicles and tractor-trailers is then decreased and drivers' awareness may be reduced and travel speed is increased, resulting into a high crash potential condition. This result is similar to the findings of Hiselius (2004) in Sweden.

The variable of adjacent to metropolitan (*AM*) increases crash frequency and severity for PDO only, suggesting that a high number of accidents occur on segments close to urban areas and, fortunately, these accidents have low severity. This is because traffic volume on segments neighboring urban arterials is usually heavy and vehicles travel at a relatively slow speed, increasing the potential for PDO accidents. The variable of adjacent to airport, seaport or industry park (*AP*) also increases crash frequency. It is because there are more trucks traveling at the segments near airports, seaports and industry parks, making crash potential high.

In terms of freeway facilities, posted speed cameras (*PS*) decrease the potential of non-severe crashes (injury and PDO) given the number of accidents unchanged, suggesting that although a *PS*

Table 6
Model results of multinomial generalized Poisson with error components (EMGP).

		Severity level					
		Fatality		Injury		PDO	
		Para.	t-Stat	Para.	t-Stat	Para.	t-Stat
Logit crash severity model component							
Constant		–		2.000	65.822	5.587	85.503
Freeway geometrics							
GN	Number of lanes	–		–		−0.152	−6.421
exp(GU)	Exponential of maximum upward slope	−0.496	−15.989	–		–	
GD	Maximum downward slope	0.342	11.068	0.185	6.316	–	
GO	Clothoid parameter	0.231	7.459	0.158	5.248	–	
GS	Speed limit	–		0.369	11.954	–	
Traffic characteristics							
TTV	Total traffic	−0.724	−23.387	−0.738	−25.559	–	
PLV	Percentage of large vehicles	6.488	89.411	6.488	89.411	–	
Neighborhood							
AM	Adjacent to metropolitan	–		–		0.200	6.447
Freeway facilities							
PS	Presence of posted speed camera	–		−0.711	−22.936	−0.483	−15.604
PR	Presence of rest area	−1.382	−44.519	−0.579	−18.647	–	
Error component in crash severity model							
σ_S		1.000	–	0.262	8.459	0.824	26.527

		Para.	t-Stat
Generalized Poisson crash frequency model component (for all severity levels)			
Constant		0.982	31.684
η		0.062	6.390
Freeway geometrics			
GC	Curvature	0.152	4.925
GD	Maximum downward slope	−0.448	−14.797
GD^2	Square of maximum downward slope	0.036	3.352
Traffic characteristics			
PSV	Percentage of small vehicles	3.260	95.035
Neighborhood			
AM	Adjacent to metropolitan	0.428	13.763
AP	Adjacent to airport, seaport or industry area	0.534	17.217
Freeway facilities			
PT	Presence of toll station	−0.387	−12.458
Error component in crash frequency model			
σ_{GPM}		0.246	7.923

Goodness of fit measures	
Log-likelihood (Null model)	−1298.488
Log-likelihood (Full model)	−1108.130
$Adj\text{-}\rho^2$	0.201
Samples	124

Note: Null model: with three constants for crash severity (market share) and a single constant for generalized Poisson model.

does not reduce crash frequency, it may increase crash severity. The cause and effect relationship may be reversed. That is, posted speed cameras are usually installed at the segments with high potential for fatal crashes. If a segment has a rest area (*PR*), it has an effect in contrast to that of *PS*, because merging and diverging maneuvers on these segments slow traffic down and reduce potential for severe accidents. Meanwhile, if the segment has a toll station, the frequency of accidents is reduced. Drivers would be more careful while traversing toll stations at a lower speed due to more complicated driving maneuvers required than traveling at other segments, so the crash potential is mitigated.

The estimated parameters of the explanatory variables in proposed model results do not directly show the magnitude of the effects on the expected frequency for each level and all severities. Moreover, some explanatory variables (i.e., *AM* and *GD*) do not carry the same effects and implications on crash frequency model and severity model components, respectively. To better understand the impact of contributory factors, Table 7 further reports the elasticity effects of significant variables on individual severity levels (i.e., PDO, injury and fatality) and on aggregate level. Since calculation formulas and implications of dummy variables and continuous variables are different, they cannot be compared and should be described respectively.

Aggregate level elasticity values for continuous variables are computed based on the estimated EMGP model by Eq. (18):

$$\xi_{tjk} = \left(\frac{\partial E(y_{jt})}{\partial x_{jtk}}\right)\left(\frac{x_{jtk}}{E(y_{jt})}\right) \tag{18}$$

where $E(y_{jt}) = \pi_{jt}(x_{jtk})\lambda_{jt}(x_{jtk})$; $E(y_{jt})$ is expected frequency of severity level j at segment t; and x_{jtk} is the contributory variable k of accident frequency at severity level j on segment t. As x_{jtk} has changed, the accident frequency and severity are adversely affected, such that elasticity represents the effect of the corresponding

Table 7
Aggregate elasticity estimates of the EMGP model.

Variable		Severity level			Frequency
		Fatality	Injury	PDO	
Continuous variable					
Freeway geometrics					
GN	Number of lanes	0.383	0.377	−0.024	0.000
GC	Curvature	0.243	0.142	0.147	0.147
GU	Maximum upward slope	−0.098	0.001	0.001	0.000
GD	Maximum downward slope	0.226	−0.081	−0.244	−0.232
GO	Clothoid curve value	0.208	0.132	−0.009	0.000
Traffic characteristics					
TTV	Total traffic	−0.659	−0.671	0.043	0.000
PSV	Percentage of small vehicles	1.820	1.757	1.887	1.880
PLV	Percentage of large vehicles	1.422	1.440	−0.093	0.000
Dummy variable					
Freeway geometrics					
GS	Speed limit	−0.623	10.790	−0.626	0.000
Neighborhood					
AM	Adjacent to metropolitan	−0.313	3.877	−0.478	−0.238
AP	Adjacent to airport, seaport or industry area	39.476	46.454	33.394	34.147
Freeway facilities					
PS	Presence of posted speed camera	22.259	−9.784	0.430	0.000
PT	Presence of toll station	−22.654	−23.631	−23.710	−23.699
PR	Presence of rest area	−61.650	−35.320	2.456	0.000

factor on crash frequency at each severity level. Additionally, "elasticity effects" of dummy variables are computed by altering the value of the variable to "1" for the subsample of observed segments for which the variable takes a value of "0", and to "0" for the subsample for which the variable takes a value of "1". We then sum the shifts of expected frequencies in the two subsamples after reversing the sign of the shifts in the second subsample, and compute an effective percentage change in expected aggregate frequency. Thus, the dummy elasticity effect could be interpreted as the percentage change at the expected frequency of an injury severity level due to change in the dummy variable from 0 to 1 (for more details see Eluru and Bhat, 2007).

Specifically, for continuous variables of *PSV* and *PLV* have an estimated elasticity >1 for severe accident types, suggesting that they are key factors to more severe accidents. According to the estimation results of the EMGP model (Table 6), an increase in *PSV* significantly increases the number of accidents but not crash severity. By elasticity estimates, *PSV* was actually identified as a key factor contributing to crash frequency with the similar marginal effects on each severity level. It is worth of noting that crashes at the segments with high heavy traffic tend to be less severe.

Comparing to the high elasticity effects of traffic characteristics, geometric variables have relatively lower effects on crash severity and frequency. It is because the geometric design standard for freeways is usually higher than surface roadways, making highly curved and sloped freeway segments barely existed. However, according to the computed elasticity effects, some geometric variables still affect crash severity and frequency. Generally, too curved and too many lanes freeway should be avoided in freeway planning and design. It is interesting to note that the maximum downward slope have a positive elasticity effect on fatal crashes but a negative elasticity effect on crash frequency, because drivers would be more carefully while traveling at the downward sloped segments, but once a crash occurs, the severity would be largely increased due to the difficulty in braking.

The elasticity effects of dummy variables are relatively larger than those of continuous variables because of their different formulas. Therefore, it is meaningless to compare the effects of continuous and dummy variables. However, among all dummy variables, *AP* has the largest positive elasticity effects on crashes at all severity levels. To install warning signs and to properly confine overtaking behaviors at the segments near airports, seaports and industry parks could effectively reduce crashes at all severity levels. Conversely, *PT* and *PR* have negative effects on severe crashes. The presence of toll station and rest area can slow vehicle speed and reduce the number of severe accidents. Notably, the presence of rest area can largely reduce severe accidents but slightly increases PDO accidents.

6. Conclusions

This study contributes to literature in several ways. First, this study integrates an accident frequency model with a severity model under the MGP architecture, and uses the integrated model to analyze accident data—count data (crash frequency) and ratio data (severity)—such that the MGP model is more efficient in evaluating and presenting accident data. Notably, according to estimation results, the factors contributing to accident frequency and severity differ markedly. Generally, traffic related factors have larger effects on crash severity and frequency than geometric factors.

Additionally, four models are developed and compared. This study adopted the shared error term to construct common errors and covariance structure so as to improve model explanatory capability and reliability. The estimation results show that the EMGP model performs best, as this model specifies the error component in the crash frequency and severity model by allowing different errors in crash frequency and severity. Thus, the estimation results show that the proposed covariance structure can further enhance the model performance.

Based on the proposed framework, future studies can introduce more flexible models in the context of frequency modeling, such as Poisson log-normal, random-parameters and other mixed distribution count models. For modeling severity outcomes, ordered probit, mixed logit (also called the random parameters logit model) and more compatible generalized extreme value models (GEV family model) like generalized nested logit (GNL) are recommended. Additionally, there is no segment with zero crash count due to the spatial segmentation used in this study, which might lead to biased estimation parameters. More refined spatial segmentation or other censored models (e.g., Tobit regression in Anastasopoulos et al., 2012) on accident rates can be considered. Furthermore, this study uses the shared error component to handle the common error term and covariance structure. The covariance structure can be derived to enhance model performance further. Additionally, it also deserves to compare prediction performances among different modeling frameworks in the context of crash severity and frequency, such as multivariate Poisson log-normal (MPLN) models, which aims to simultaneously modeling crash frequencies at all severity levels. Last, additional explanatory variables can be utilized to investigate their effects on accident frequency and severity to generate more effective safety improvement strategies.

Acknowledgements

The authors are indebted to three anonymous reviewers for their insightful comments and constructive suggestions, which help clarify several points made in the original manuscript. This study was financially sponsored by the ROC National Science Council (NSC 97-2628-E-009-035-MY3).

References

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

Abdel-Aty, M., 2003. Analysis of driver injury severity levels at multiple locations using ordered probit models. Journal of Safety Research 34 (5), 597–603.

Aguero-Valverde, J., Jovanis, P.P., 2009. Bayesian multivariate Poisson log-normal models for crash severity modeling and site ranking. In: Presented at the 88th Annual Meeting of the Transportation Research Board.

Anastasopoulos, P., Mannering, F., 2009. A note on modeling vehicle-accident frequencies with random-parameters count models. Accident Analysis and Prevention 41 (1), 153–159.

Anastasopoulos, P., Mannering, F., Shanker, V., Haddock, J., 2012. A study of factors affecting highway accident rates using the random-parameters Tobit model. Accident Analysis and Prevention 45 (1), 628–633.

Bhat, C., 2003. Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. Transportation Research Part B 37 (1), 837–855.

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

Carson, J., Mannering, F., 2001. The effect of ice warning signs on accident frequencies and severities. Accident Analysis and Prevention 33 (1), 99–109.

Chang, L.Y., Mannering, F., 1999. Analysis of injury severity and vehicle occupancy in truck and non-truck-involved accident. Accident Analysis and Prevention 31 (4), 579–592.

Dissanayake, D., Aryaijab, J., Wedagamac, P., 2009. Modelling the effects of land use and temporal factors on child pedestrian casualties. Accident Analysis and Prevention 41 (4), 1016–1024.

Duncan, C., Khattak, A., Council, F., 1998. Applying the ordered probit model to injury severity in truck-passenger car rear-end collisions. Transportation Research Record 1635, 63–71.

El-Basyouny, K., Sayed, T., 2009. Collision prediction models using multivariate Poisson-lognormal regression. Accident Analysis and Prevention 41 (4), 820–828.

Eluru, N., Bhat, C., 2007. A joint econometric analysis of seat belt use and crash-related injury severity. Accident Analysis Prevention 39, 1037–1049.

Famoye, F., 1993. Restricted generalized Poisson regression model. Communications in Statistics, Theory and Methods 22, 1335–1354.

Famoye, F., Wulu, J.T., Singh, K.P., 2004. On the generalized Poisson regression model with an application to accident data. Journal of Data Science 2, 287–295.

Hadi, M.A., Aruldhas, J., Chow, L.F., Wattleworth, J.A., 1995. Estimating safety effects of cross-section design for various highway types using negative binomial regression. Transportation Research Record 1500, 169–177.

Hiselius, L.W., 2004. Estimating the relationship between accident frequency and homogeneous and inhomogeneous traffic flows. Accident Analysis Prevention 36, 985–992.

Jones, B., Janssen, L., Mannering, F., 1991. Analysis of the frequency and duration of freeway accidents in Seattle. Accident Analysis and Prevention 23 (2), 239–255.

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

Khorashadi, A., Niemeier, D., Shankar, V., Mannering, F., 2005. Differences in rural and urban driver-injury severities in accidents involving large-trucks: an exploratory analysis. Accident Analysis and Prevention 37 (5), 910–921.

Kim, J.K., Ulfarsson, G., Shankar, V., Kim, S., 2008. Age and pedestrian injury severity in motor-vehicle crashes: a heteroskedastic logit analysis. Accident Analysis and Prevention 40 (5), 1695–1702.

Kockelman, K.M., Kweon, Y.J., 2002. Driver injury severity: an application if ordered probit models. Accident Analysis and Prevention 34 (3), 313–321.

Koppelman, F.S., Wen, C., 1998. Alternative nested logit models: structure, properties and estimation. Transportation Research Part B 32 (5), 289–298.

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

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

Lord, D., 2006. Modeling motor vehicle crashes using Poisson-gamma models: examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accident Analysis and Prevention 38 (4), 751–766.

Lord, D., Miranda-Moreno, L.F., 2008. Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: a Bayesian perspective. Safety Science 46 (5), 751–770.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transportation Research Part A: Policy and Practice 44 (5), 291–305.

Lui, K., McGee, D., Rhodes, P., Pollock, D., 1988. An application of a conditional logistic regression to study the effects of safety belts, principal impact points, and car weights on drivers' fatalities. Journal of Safety Research 19 (4), 197–203.

Ma, J., Kockelman, K.M., 2006. Bayesian multivariate Poisson regression for models of injury count by severity. Transportation Research Record 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 (3), 964–975.

Malyshkina, N., Mannering, F., 2010. Empirical assessment of the impact of highway design exceptions on the frequency and severity of vehicle accidents. Accident Analysis and Prevention 42 (1), 131–139.

McFadden, D., 1981. Econometric models of probabilistic choice. In: Manski, C.F., McFadden, D. (Eds.), Structure Analysis of Discrete Data with Econometric Applications. MIT Press, Cambridge, MA.

Miaou, S.P., Bligh, R.P., Lord, D., 2005. Developing median barrier installation guidelines: a benefit/cost analysis using Texas data. Transportation Research Record 1904, 3–19.

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

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

Milton, J., Shankar, V., Mannering, F., 2008. Highway accident severities and the mixed logit model: an exploratory empirical analysis. Accident Analysis and Prevention 40 (1), 260–266.

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

Oh, J., Washington, S.P., Nam, D., 2006. Accident prediction model for railway-highway interfaces. Accident Analysis and Prevention 38 (6), 346–356.

Park, B.J., Lord, D., 2009. Application of finite mixture models for vehicle crash data analysis. Accident Analysis and Prevention 41 (4), 683–691.

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

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

Renski, H., Khattak, A., Council, F., 1999. Effect of speed limit increases on crash injury severity: analysis of single-vehicle crashes on North Carolina interstate highways. Transportation Research Record 1665, 100–108.

Saccomanno, F., Nassar, S., Shortreed, J., 1996. Reliability of statistical road accident injury severity models. Transportation Research Record 1542, 14–23.

Shankar, V., Mannering, F., Barfield, W., 1995. Effect of roadway geometrics and environmental factors on rural accident frequencies. Accident Analysis and Prevention 27 (3), 371–389.

Shankar, V., Mannering, F., Barfield, W., 1996. Statistical analysis of accident severity on rural freeways. Accident Analysis and Prevention 28 (3), 391–741.

Shankar, V., Milton, J., Mannering, F., 1997. Modeling accident frequencies as zero-altered probability processes: an empirical inquiry. Accident Analysis and Prevention 29 (6), 829–837.

Savolainen, P., Mannering, F., 2007. Probabilistic models of motorcyclists' injury severities in single- and multi-vehicle crashes. Accident Analysis and Prevention 39 (6), 955–963.

Savolainen, P., Mannering, F., Lord, D., Quddus, M., 2011. The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. Accident Analysis and Prevention 43 (5), 1666–1676.

Song, J.J., Ghosh, M., Miaou, S., Mallick, B., 2006. Bayesian multivariate spatial models for roadway traffic crash mapping. Journal of Multivariate Analysis 97 (1), 246–273.

Terza, J.V., Wilson, P.W., 1990. Analyzing frequencies of several types of events: a mixed multinomial-Poisson approach. The Review of Economics and Statistics 72 (1), 108–115.

Train, K., 2003. Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge, UK.

Ulfarsson, G.F., Mannering, F., 2004. Differences in male and female injury severities in sport-utility vehicle, pickup and passenger car accidents. Accident Analysis and Prevention 36 (1), 135–147.
Wang, X., Abdel-Aty, M., 2008. Modeling left-turn crash occurrence at signalized intersections by conflicting patterns. Accident Analysis and Prevention 40 (1), 76–88.
Yamamoto, T., Shankar, V., 2004. Bivariate ordered-response probit model of driver's and passenger's injury severities in collisions with fixed object. Accident Analysis and Prevention 36 (5), 869–876.
Yau, K., 2004. Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong. Accident Analysis and Prevention 36 (3), 333–340.
Yamamoto, T., Hashiji, J., Shankar, V., 2008. Underreporting in traffic accident data, bias in parameters and the structure of injury severity models. Accident Analysis and Prevention 40 (4), 1320–1329.
Ye, X., Pendyala, R.M., Washington, S.P., Konduri, K., Oh, J., 2009. A simultaneous equations model of crash frequency by collision type for rural intersections. Safety Science 47 (3), 443–452.