Analytic Methods in Accident Research 20 (2018) 81–91

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

Investigating varying effect of road-level factors on crash frequency across regions: A Bayesian hierarchical random parameter modeling approach

Chunyang Han[a], Helai Huang[a], Jaeyoung Lee[b], Jie Wang[a,*]

[a] School of Traffic and Transportation Engineering, Smart Transport Key Laboratory of Hunan Province, Central South University, Changsha 410075, China

[b] Environmental and Construction Engineering, University of Central Florida, Orlando, FL 32816-2450, United States

HIGHLIGHTS

- Developed a hierarchical random parameter model to analyze crash frequency.
- Showed the proposed model superiority over other models in goodness-of-fit.
- Illustrated varying safety effects of road-level factors across regions.
- Provided new insights in modeling crash frequency with hierarchical structure.

ARTICLE INFO

Keywords:
Crash frequency
Road-level factor
Regional variation
Hierarchical random parameter model

ABSTRACT

This study aims to quantitatively examine the variations in effect of road-level factors on crash frequency across different regions. Treating the hierarchical structure existing in the crash data that road entity nested within the geographic region, a hierarchical random parameter model, which allows the coefficients of road-level variables to vary with regions, is proposed. A Poisson lognormal model and a hierarchical random intercept model are also built for the purpose of comparison. A specific roadway facility type, urban two-lane two-way roadway segments in Florida, with crash and road level data including traffic volume, road length, surface condition, and access density for three-year period are used for a case study. The result shows that, in the hierarchical-random parameter model, the local regression coefficients and marginal effects of the road level factors vary over a wide range in the selected counties, which clearly illustrates the non-stationary in the relationships between road level factors and crash frequency across the counties. In regard to the model comparison, the hierarchical random parameter model outperforms the Poisson lognormal model and the hierarchical random intercept model in term of deviance information criterion (DIC). This further confirms the necessity of the use of hierarchical random parameter model in analyzing the crash frequency for road entities in different regions. This study provides a potential in guidance of model construction that considers regional variations (heterogeneities) in safety effects of road-level factors.

* Corresponding author.
E-mail addresses: Sandiant@csu.edu.cn (C. Han), huanghelai@csu.edu.cn (H. Huang), jaeyoung@knights.ucf.edu (J. Lee), jie_wang@csu.edu.cn (J. Wang).

https://doi.org/10.1016/j.amar.2018.10.002
Received 16 June 2018; Received in revised form 12 October 2018; Accepted 12 October 2018
Available online 19 October 2018
2213-6657/

1. Introduction

Crash frequency model associates expected crash frequency for roadway types with traffic volumes and roadway features, and it is used as the important quantitative tool to predict crash frequency and identify high-crash locations. Various kinds of crash frequency models have been proposed for different roadway types, such as rural two-lane, rural multilane, and urban and suburban arterial roadway segments and intersections. From a methodological perspective, Poisson and Negative Binomial regression models are the most commonly used for developing crash frequency models. However, the fundamental assumption of the aforementioned basic modeling techniques is that the relationship between crash frequency and road-level factors is fixed across space (spatial stability) and/or constant over time (temporal stability). This assumption is often violated since the crash observations are space and/or time dependent. A detailed discussion of temporal instability that exists in crash data can be found in recent paper of Mannering (2018). The present study focuses on the issue of spatial instability/variations. The aim of study is to investigate the underlying variations in effects of road-entity related factors (roadway features and traffic volumes) on crash frequency across geographical regions. A hierarchical model with random parameters is proposed as a methodological alternative to account for the regional varying relationships between the crash frequency and road level factors by allowing the coefficients to vary among the highly aggregated geographic units (e.g., county). The data collected for urban two-lane two-way roadway segments in Florida are used as a case study for the proposed model and examine the varying safety effects of road level factors across counties. Road level factors including traffic volume, segment length, access density, surface condition and median type are examined in this study.

Traffic crashes are the result of combination of multiple factors. According with Miranda-Moreno et al. (2011), Mitra and Washington (2012), Ukkusuri et al. (2012), Ryb et al. (2007), Lee et al. (2015a), Wang et al., (2017) and Adanu et al. (2017), factors that may be associated with crash frequency of the road entity (e.g. road segment or intersection) in the given period could be divided into three types: (1) road-entity-level factors related to road features and traffic characteristics, (2) region-level (also called macro-level) factors related to the regional environment of where the road entity locates (e.g., driving regulations, land use patterns, socioeconomic characteristics), and (3) individual-level factors related to road-user characteristics and behaviors (e.g., gender, age, alcohol consumption, risky actions). Ideally, it would be the best to integrate all relevant risk factors for obtaining accurate model predictions. However, existing crash databases, which were extracted from the authorities (e.g. police, department of transportation, local governments) only cover a small fraction of the large number of these relevant factors. Many factors (especially related to region-level and individual-level factors) are usually unavailable or even unobservable (Mannering and Bhat, 2014; Mannering et al., 2016). As a result, many important factors have been often not considered in modeling crash frequency. From a statistical view, if these omitted variables are significantly correlated with selected variables, this omission could introduce variations in the effects of selected variables on crash likelihood (Lord and Mannering, 2010; Mannering et al., 2016; Mannering, 2018). The omitted factors constitute the so-called unobserved heterogeneity. An overview of the potential for unobserved heterogeneity in the context of crash data and analysis has been highlighted by Mannering et al. (2016).

Risk factors associated with crash frequencies exhibit a hierarchical structure (Huang and Abdel-Aty, 2010). Road entities and their located geographic region could be viewed as a hierarchical system of the road entity nested within the geographic region. The road entities within the same region may be correlated with each other because they share similar unobserved (regional) characteristics (e.g. enforcement of driving regulations, land use patterns, socioeconomic characteristics, road density and traffic environments), and thus the traffic participants' characteristics and behaviors tend to be similar (Christoffel and Gallagher, 1999; Adanu et al., 2017). The presence of these unobserved regional factors (or heterogeneities) and their interactions with observed road-entity-level factors can eventually result in variations in the effect of these factors on crash likelihood across different space unit. Consequently, the effects of the road-entity-level factors on crash frequency in the same region are seen to be more similar than the effects in different geographical regions (Raudenbush and Bryk, 2002; Gelman and Hill, 2007; Lee et al., 2017). For example, at intersections, the safety effect of traffic signal could be largely influenced by the enforcement of local traffic regulations and the compliance with these regulations (which may not be easily observed). Many previous studies have found that spatial variations in the relationship between risk factors and crash likelihood and severity may exist across states/provinces (Erdogan, 2009; Aguero-Valverde, 2013; Truong et al., 2016; Heydari et al., 2018), counties (Aguero-Valverde and Jovanis, 2006; Song et al., 2006; Quddus, 2008; Huang et al., 2010; Li et al., 2013; Liu and Sharma, 2017; Liu and Sharma, 2018), traffic analysis zones (Xu and Huang, 2015; Ding et al., 2016; Xu et al., 2017; Ding et al., 2017) and even different road entities (Wang and Huang, 2016; Barua et al., 2016; Wang et al., 2017; Alarifi et al., 2017). Ignoring these potential spatial effects in estimated parameters, could adversely affect the inferences drawn from model estimations, as well as their ability for forecasting and evaluating the effects of safety countermeasures (Mannering and Bhat, 2014; Mannering et al., 2016).

A vast array of studies has been devoted to utilize statistical and econometric models to account for unobserved heterogeneity (Lord and Mannering, 2010, Savolainen et al., 2011, Mannering and Bhat, 2014). The random parameter modeling approaches have gained considerable attentions in crash frequency analysis (Anastasopoulos and Mannering, 2009; Venkataraman et al., 2014; Anastasopoulos and Mannering, 2016; Barua et al., 2016; Sarwar et al., 2017; Alarifi et al., 2017; Shaon et al., 2018) and crash severity analysis (Kim et al., 2010; Venkataraman et al., 2013; Russo et al., 2014; Anastasopoulos, 2016; Bhat et al., 2017; Behnood and Mannering, 2017; Fountas and Anastasopoulos, 2017; Fountas et al., 2018; Heydari et al., 2018). They address data heterogeneity by allowing the model parameters to vary across observations. The parameter is treated as a random variable whose probability distribution is usually defined by analysts. However, the traditional random parameters modeling approach assumes the sources of heterogeneity are independent over the sample population (Fountas et al., 2018; Heydari et al., 2018). As mentioned previously, there is strong possibility for the road-entity-level factors to be correlated across space, since the possible interactions among factors from different regions (or other levels). Traditional random parameters modeling approach cannot provide parameter estimates that account for the possible spatial correlation effects among road entities in crash data with hierarchical structure (Fountas et al., 2018; Heydari et al., 2018).

One method of addressing the aforementioned hierarchical structure is to separately estimate different parameter sets for each region (such as using traditional random parameter models). This is a straightforward approach for considering cross-region differences in safety effects of risk factors. However, such independent analysis often yields unreliable estimates, especially when the sample size is limited. Hierarchical modeling approach is another choice which has the potential to deal with this issue of hierarchical structure by simultaneously incorporating different level models. Recently, the hierarchical modeling technique is recommended application in traffic safety analysis (Shankar et al., 1998; Jones and Jorgensen, 2003; Quddus, 2008; Huang et al., 2010; Emmanuelle et al., 2013; Coruh et al., 2015; Wang and Huang, 2016; Cai et al., 2018). In this study, a random parameter model is incorporated into the hierarchical modeling framework to account for regional-specific heterogeneity for investigating varying relationships between the crash frequency and road-entity-level factors across regions. The proposed model could distinctly address and properly estimate the hierarchical structure of data, but also mitigate the adverse impacts of these omitting variables (i.e., region-level and individual-level) by allowing the regression coefficients to vary with regions.

The rest of the paper is structured as follows. In the next section, a description of the statistical models applied in this study is presented. This is followed by a brief description of the data used for model development. Then, the results of parameter estimation obtained from the developed models are presented. Finally, conclusions are drawn and research implications are discussed.

2. Methodology

We begin this section with a quick review of the fixed coefficients model (Poisson lognormal model) that is commonly used for modeling crash frequency, then we move on to detail how this method could be readily generalized to the hierarchical random parameter model. Next, Bayesian inference used for model parameter estimation and the measurement for model performance comparison are also introduced.

2.1. Model specification

Traffic crashes have been conformed to an expanded Poisson distribution. Poisson model and its variants (such as negative binomial and Poisson lognormal model), are widely used and proven to be successful as they are capable of effectively modeling the rare, random, sporadic, and non-negative crash data. Compared with Poisson model, Poisson lognormal model can provide a more credible estimation by incorporating random effects of the linear predictor to address overdispersion for unobserved heterogeneity (Lord and Mannering, 2010). The formulation for Poisson lognormal model can be presented as follows:

$$Y_i \sim \text{Poisson } (\lambda_i) \tag{1}$$

$$\log \lambda_i = \beta_0 + \sum_{m=1}^{M} X_{mi}\beta_m + \varepsilon_i \tag{2}$$

where Y_i is the observed number of crash frequency for road entity i, and λ_i is the expectation of Y_i. X_{mi} is the mth factor for road entity i. β_0 and β_m are the intercept coefficient and the slope coefficient, respectively. ε_i denotes the random effect which is assumed to follow independently a normal distribution with zero mean and standard deviation $\sigma(\sigma > 0)$, that is,

$$\varepsilon_i \sim N(0, \ \sigma^2)$$

The aforementioned Poisson lognormal model assumes safety effects of factors (regression coefficients) to be fixed. In fact, this fixed setting of regression coefficient builds on an underlying assumption that the observations should be mutually independent. However, this fundamental requirement may always be violated. As mentioned before, road entities and the geographic region where they are located in could be viewed as a two-level hierarchical structure with road entity being the first level and geographic region being the second level. Due to the shared unobserved (regional) characteristics, crashes that occurred in a same region are possibly more similar comparing to those in other regions. In this case, a hierarchical random parameter model is further proposed to model crash frequency for road systems with hierarchical structure. In the hierarchical model with random parameters, regression coefficients (the intercept coefficient and the slope coefficient) are not fixed but vary with regions. That is, the $(\beta_{0,} \beta_1, \ldots, \beta_M)$ in Eq. (2) are set to be varying coefficient $(\beta_{0j,} \beta_{1j}, \ldots, \beta_{Mj})$ as:

$$Y_{ij} \sim \text{Poisson } (\lambda_{ij}) \tag{3}$$

$$\log\lambda_{ij} = \beta_{0j} + \sum_{m=1}^{M} X_{mij}\beta_{mj} + \varepsilon_i \tag{4}$$

$$\beta_{0j} = \gamma_{00} + \mu_{0j} \tag{5}$$

$$\beta_{mj} = \gamma_{m0} + \mu_{mj} \tag{6}$$

where Y_{ij} is the observed number of crashes for road entity i (e.g., segment) in region j (e.g., county), γ_{00} is the mean of the intercept. γ_{m0} is the average effect of mth factor. ε_i is the random effect representing within-region variance. μ_{0j} and μ_{mj} are random effects representing between-region variances, which are consistent for road entity in the same region but vary between different regions.μ_{0j} and μ_{mj} reflect unique effects associated with region *j*. ε_i, μ_{0j} and μ_{mj} are generally assumed to follow normally distributed with mean zero, that is,

$$\varepsilon_i \sim N(0, \ \sigma^2)$$

$$\mu_{0j} \sim N(0, \ \delta_0^2)$$

$$\mu_{mj} \sim N(0, \ \delta_m^2)$$

A simplified equation of hierarchical random parameter model is to let the intercept vary ($\beta_{0j} = \gamma_{00} + \mu_{0j}$) with fixed slope ($\beta_{mj} = \ \gamma_{m0}$), forming a hierarchical random intercept model. This means that the between-region variance only works on the intercept β_{0j}. In this study, a hierarchical model with random intercept is also provided as a reference for model performance comparisons.

In this study, the intra-class correlation coefficient (ICC) is employed to examine the correlation among road entities within the same region. The ICC is calculated as the portion of total variance explained by between-region variance (Raudenbush and Bryk, 2002), that is,

$$ICC = \frac{\delta_0^2}{\sigma^2 + \delta_0^2} \tag{7}$$

where δ_0^2 is the between-region variance of random effects μ_{0j} and σ^2 is the within-region variance of random effects ε_i. The value of ICC ranges from 0 to 1. A large ICC value close to 1 indicates the greater correlation among road entities within a region that a hierarchical model is preferable.

2.2. Bayesian inference

Compared to maximum likelihood estimation, Bayesian inference method is able to model parameter estimates with posterior distributions and predict new observations from a given sample of data. Several researchers have shown the advantages of Bayesian inference method in achieving valid results using smaller samples, and in accommodating complex model structures, such as the hierarchical structure (Ntzoufras, 2009). The theoretical framework for Bayesian inference can be expressed as:

$$\pi(\theta|y) = \frac{L(y|\theta)\pi(\theta)}{\int L(y|\theta)\pi(\theta)d\theta} \tag{8}$$

where y is the vector of observed data, θ the vector of parameters required for the likelihood function, $L(y|\theta)$ the likelihood function, $\pi(\theta)$ the prior distribution of θ, $\int L(y|\theta)\pi(\theta)d\theta$ the marginal distribution of observed data, and $\pi(\theta|y)$ the posterior distribution of θ given y.

In this study, non-informative priors are specified for parameters and the hyper-parameters, which are estimated based on previous studies (Xu, et al., 2017; Huang, et al., 2017). Specifically, a diffused normal distribution (0, 1000) is used as the priors of γ_{00} and γ_{m0}, and a diffused gamma distribution gamma (0.001, 0.001) is used as the priors of precisions of the normal distributions, $1/\sigma^2$, $1/\delta_0^2$, and $1/\delta_m^2$. The model simulation procedure is conducted with a Monte Carlo Markov Chain (MCMC) algorithm in the WinBUGS platform.

2.3. Model assessment

Under the Bayesian framework, Deviance Information Criterion (DIC) is commonly used to compare complex models since it offers a Bayesian measure of model fit and complexity (Spiegelhalter et al., 2002). Specifically, DIC is defined as:

$$DIC = \overline{D(\theta)} + P_D \tag{9}$$

where $\overline{D(\theta)}$ is the posterior mean deviance that can be taken as a Bayesian measure of goodness-of-fit, and P_D is a complexity measure for the effective number of parameters. Models with lower DIC values are preferred. Generally, differences in DIC of more than 10 definitely rule out the model with the higher DIC, differences between 5 and 10 are considered substantial, while a difference of less than 5 indicates that the models are not statistically different (Spiegelhalter et al., 2002).

3. Data preparation

The data collected for a specific roadway facility type, two-lane two-way urban road, in Florida are used to illustrate the proposed hierarchical random parameter model and examine varying effects of road level factors across regions. As shown in Fig. 1, the road entity level is the two-lane two-way urban segment (level 1) and the region level under consideration is the county (level 2).

Three years of crash data (2005–2007) were obtained from the Florida Department of transportation (FDOT) Crash Analysis Reporting System. Meanwhile, the shape files of site characteristics were downloaded from the website of the FDOT Transportation Data and Analytics Office. Then, geographical information system (GIS) technique was used to extracted two-lane two-way urban road segments according to the road characteristics including function class and number of lanes. Subsequently, GIS was used to map crashes and site characteristics to these segments.

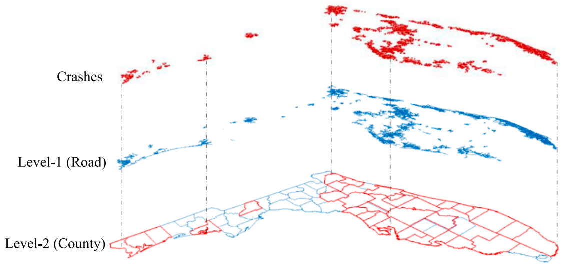

Fig. 1. Illustration of hierarchical relation among crashes, road entities, and counties.

Since the number of investigated segment type in some counties of Florida is small, with consideration of simple size in hierarchical modeling, we extracted two-lane two-way urban segments in 34 counties which have relatively higher number of this road type among all of 67 counties of Florida. As a result, a total of 3,857 road segments with 38,938 crashes occurred were selected for our analysis, and the number of segments ranges from 26 to 337 within each selected county. The descriptive statistics of crash data in a three-year period (2005–2007) and factors related with traffic volumes and roadway features for these selected segments are provided as in Table 1.

Collinearity diagnose for the factors of segment was conducted in SPSS software. As the result, the eigenvalue of the variable “median” is 0.015, and its condition index is more than 17, indicating that the presence of the median is significantly collinear with other factors. Therefore, the variable “median” was excluded from the models.

4. Results and discussions

Three models developed in this study are: 1) Bayesian hierarchical random parameter model that both the intercept coefficient and the slope coefficient vary with regions; 2) Bayesian hierarchical random intercept model that the intercept coefficient is varying but the slope coefficient is constant with regions; and 3) Bayesian Poisson lognormal model that both the intercept coefficient and the slope coefficient are set to be fixed over all the regions. For each model, the posterior estimates are obtained via two chains with 60,000 iterations, 20,000 of which are excluded as a burn-in sample using WinBUGS software. The ratios of the Monte Carlo errors relative to the standard deviations of the estimates are lower than 5%, and trace plots for all model parameters indicate convergence. The results of parameter estimation for these three models using a full set of data are summarized in Table 2.

4.1. Model comparison

As shown in Table 2, the DIC values are 19670.3, 19481.1 and 19454.9 for the Poisson lognormal model, the hierarchical random intercept model and the hierarchical random parameter model, respectively. Both hierarchical models have the significantly lower DIC values compared to the fixed-parameter Poisson lognormal model. It indicates that the hierarchical models outperform the Poisson lognormal model to a substantial extent. For the comparison of the two hierarchical models, the hierarchical random parameter model has a significantly lower value of the DIC, suggesting that the use of hierarchical random parameter model is more appropriate in terms of model goodness of fit.

In the Poisson lognormal model, the variance of random effects (σ^2 = 0685) is found to be statistically and significantly different

Table 1
Summary of variable and descriptive statistics.

Variable	Description	Mean	S.D.	Min.	Max.
Response variable					
Crash frequency	Number of crashes per segment from 2005 to 2007	10.090	15.965	0	323
Road level factors					
Length	Segment length (mile)	1.497	1.005	0.600	7.876
ADT	Average daily traffic (10^3 pcu[a])	8.054	7.573	0.350	59.500
Access	Number of access roads/segment length	6.588	4.196	0	32.717
		Crash count (Proportion)		Segment count (Proportion)	
Median	Median barrier indictor:				
	Present	23,401 (60.1%)		2991 (77.5%)	
	Otherwise	15,537 (39.9%)		862 (22.5%)	
Surface Condition					
	Good	8173 (21.0%)		811 (21.0%)	
	Fair	28,216 (72.5%)		2755 (71.4%)	
	Poor	2549 (6.5%)		293 (7.6%)	

[a]pcu: passenger car units.

Table 2
Estimation results for three models.

Variables	Poisson lognormal model			Hierarchical random intercept model			Hierarchical random parameter model		
	Coefficient	S.D.	95% BCI	Coefficient	S.D.	95% BCI	Coefficient	S.D.	95% BCI
Fixed effect (mean of random parameter)									
Intercept	−0.692	0.321	(−0.733, −0.644)	−0.720	0.270	(−0.765, −0.676)	−0.711	0.284	(−0.757, −0.668)
ln(length)	0.943	0.041	(0.881, 1.000)	0.978	0.033	(0.921, 1.031)	0.953	0.037	(0.891, 1.015)
ln(AADT)	0.712	0.018	(0.677, 0.746)	0.623	0.017	(0.590, 0.655)	0.646	0.033	(0.582, 0.713)
Access	0.092	0.004	(0.084, 0.100)	0.082	0.004	(0.075, 0.089)	0.088	0.007	(0.074, 0.101)
Good surface condition	−0.192	0.039	(−0.270, −0.116)	−0.119	0.038	(−0.194, −0.044)	−0.152	0.057	(−0.270, −0.043)
Variance Component (variance of random parameter)									
Intercept (δ_0^2)				0.213	0.058	(0.126, 0.350)	0.108	0.111	(0.001, 0.379)
Length (δ_1^2)							0.005	0.002	(0.001, 0.010)
AADT (δ_2^2)							0.022	0.010	(0.008, 0.045)
Access (δ_3^2)							0.001	0.0007	(0.0004, 0.0017)
Good Surface Condition (δ_4^2)							0.041	0.020	(0.003, 0.117)
σ^2	0.685	0.030	(0.641, 0.731)	0.493	0.024	(0.460, 0.529)	0.465	0.030	(0.433, 0.499)
ICC				0.302					
DIC	19670.3			19481.1			19454.9		

from zero, as demonstrated by the lower standard deviation than mean. This result suggests that it is necessary to consider the unobserved heterogeneity and/or variability in modeling crash frequency. However, this Poisson lognormal model assumes that observations are mutually independent, and thus could not consider underlying correlations among crash occurrences within the same county. Furthermore, the hierarchical random intercept model could be structured to account for the heterogeneity/variability among analytic-specified groups (i.e., counties) of observations instead of independent observations. In the hierarchical random intercept model, the heterogeneity/variability is further distinguished into the within-county and the between-county variability. The results show that both the variance of within-county variability ($\sigma^2 = 0.493$) and between-county variability ($\delta_0^2 = 0.213$) are significantly different from zero. The interclass correlation (ICC) in the hierarchical random intercept model is 0.302 (0.302 = 0.213/(0.493 + 0.213)). This indicates that more than 30% of the total unexplained variation in crash frequency is attributable to the unexplained between-county variations.

Furthermore, by viewing the hierarchical random parameter model, we found that the variance of within-county random effect (0.465) is similar with that in the hierarchical random intercept model (0.493). However, the variance of between-county random effect reduced to 0.108 and becomes statistically insignificant (the posterior mean of variance is lower than its standard deviation). It implies that the between-region variation mainly works on the slope coefficients (the effects of road factors), rather than the intercept coefficient. This is to say that, the omitted region-level factors (e.g., traffic regulations, socioeconomic characteristics) may mainly and indirectly contribute to the crash occurrences by influencing road-user characteristics and behavior, leading to different safety effects of road level factors in different regions. This also further confirms the necessity of the use of hierarchical random parameter model to analyze the crash frequency for road entities with hierarchical structure.

4.2. *Varying effect of road level factors*

The hierarchical random parameter model can account for the varying relationships between the crash frequency and road level factors by allowing the regression coefficients to vary among higher-level units (i.e., counties). Based on the work of Gelman and Hill (2007), the rule to determine that the slope coefficient (the effect of factor) varies significantly with counties in this study is when the corresponding variance component (δ_m^2) is statistically significant and different from zero. As shown in Table 2, the estimated results of variance components indicate that all selected road level variables including ln(ADT), ln(length), access density and surface condition are associated with significant varying effects across counties.

A significant advantage of the hierarchical random parameter model is that, in addition to the global regression coefficients, local coefficients in each county can also be estimated. The local coefficient estimates for selected factors in this study are plotted by the solid lines in Fig. 2(a–d). Furthermore, since the model has nonlinear coefficients, direct parameters could not show a unit effect of explanatory variables on the number of crashes. The marginal effects are thus computed at the sample mean of the explanatory variables, and shown by the dotted lines in Fig. 2(a–d). Marginal effects for continuous variables measure the impact of a unit change in the exogenous variable on the frequency of crashes and the marginal effect for a dummy variable indicates a change in the crash counts with and without the presence of the variable. As shown in Fig. 2, these local coefficient estimates and corresponding marginal effects reveal obvious patterns of regional variations. The red horizontal straight lines in each figure are marked for estimated global means of regression coefficients. It is apparent that the local means deviated significantly from the global means estimated in the hierarchical random parameter model. The information from Table 2 and Fig. 2 illustrates that the effects of the road level factors on crash frequency appear to vary with counties rather than the potential assumption of the safety effects being fixed. Specifically, the mean of the regression coefficient and the marginal effect of ln(ADT) would be expected to vary in the range of (0.343, 0.823) and

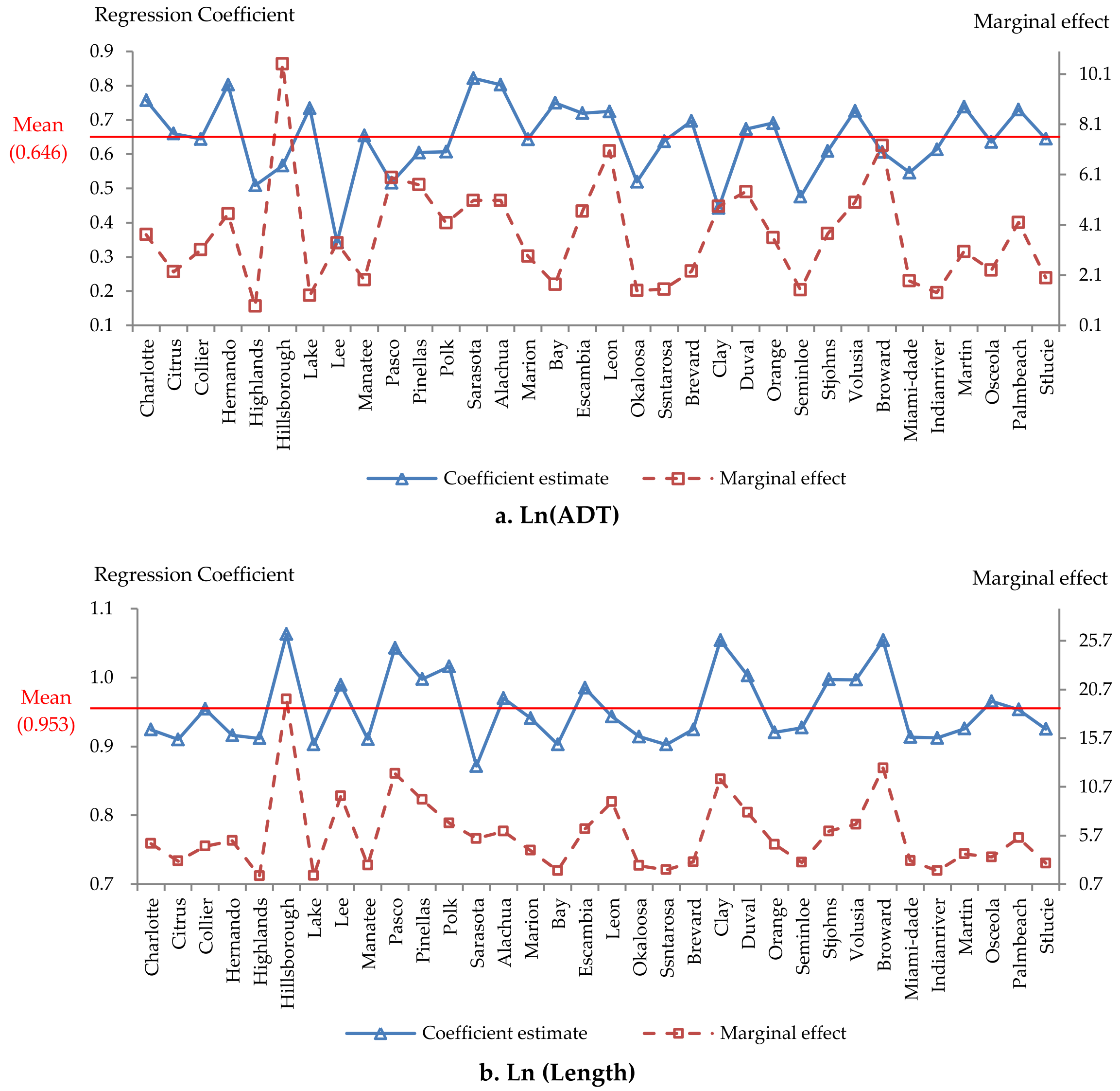

Fig. 2. Local regression coefficient estimates and marginal effects for selected road level factors in each county.

(0.867, 10.505) respectively, ln(Length) would be in the range of (0.872, 1.063) and (1.552, 19.705), access density of (0.050, 0.127) and (0.136, 1.606), and surface condition of (−0.302, 0.134) and (−1.950, 0.243).

As shown in Fig. 2, the signs of local regression coefficients for three risk-factor variables including ln(ADT), ln(Length), and access density are positive in all of selected 34 counties, indicating that these three variables have positive associations with crash frequencies in each county. But the magnitudes of positive effects are found to vary over a wide range in different counties.

The global estimated mean of the regression coefficient for ln(ADT) is 0.646 (Fig. 2a). The maximum value of the local coefficient estimation for this variable is 0.823 in Sarasota county, and the minimum is 0.343 in Lee county. Correspondingly, the global marginal effect of ln(ADT) is calculated as 4.576, the local marginal effect in Sarasota county is 5.073, and is 3.385 in Lee county. These mean that one unit increase in ln(ADT) will associate with a 1.525 (= 4.576/3) increase in average segment crash frequency per year, and with a 1.691 (= 5.073/3) increase in Sarasota county and a 1.128 (= 3.385/3) increase in Lee county. The positive effect of traffic volume is consistent with previous studies (Abdel-Aty and Wang, 2006; Gooch et al., 2016; Cai et al., 2018) that greater traffic volumes present more opportunities for exposure to traffic conflicts, and thus are associated with more crashes. Our study further reveals that the magnitude of safety effect of traffic volumes can vary significantly from one county to another.

The regional variation of coefficient estimation of ln(length) is relatively less than that of ln(ADT). The maximum value and minimum value of the local coefficient estimation of ln(length) are 1.063 in Hillsborough county and 0.872 in Sarasota county, respectively (Fig. 2b). Their corresponding marginal effects are 19.705 in Hillsborough county and 5.379 in Sarasota county, indicating that one unit increase in ln(length) would increase 6.568 (= 19.705/3) crashes in Hillsborough county per year and 1.793 (= 5.379/3) crashes in Sarasota county per year. The positive relationship between segment length and crash frequency is reasonable

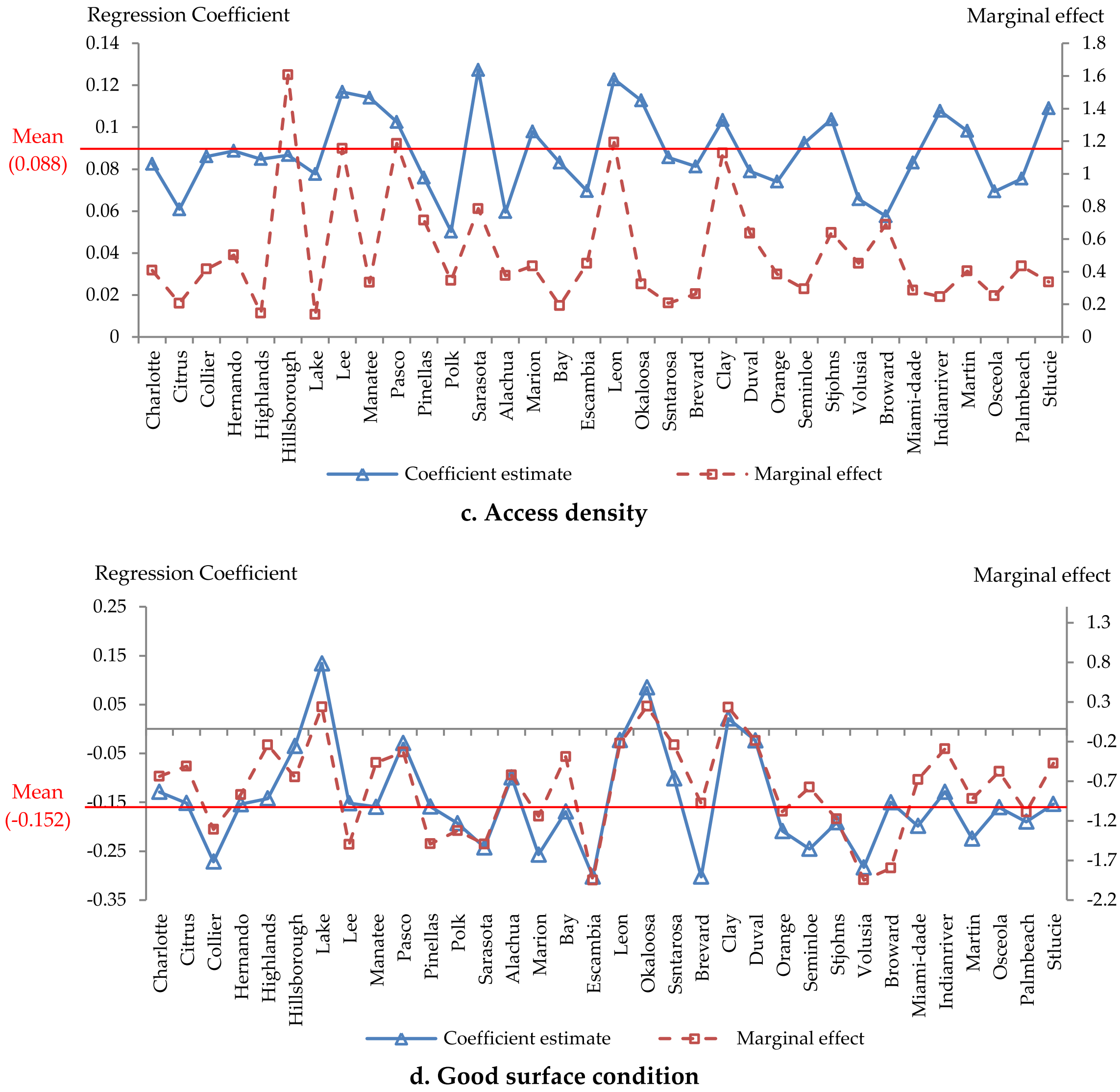

Fig. 2. (*continued*)

since a longer segment length is associated with the higher crash exposure, and thus increases the potential probability of crash occurrence.

With regard to the access density, this variable also exhibits an obvious pattern of regional inconsistency. The global estimated mean of the regression coefficient for the access density is 0.088 (Fig. 2c). County Sarasota has the maximum value (0.127) of the local coefficient estimation for this variable, and the minimum value (0.050) is in Polk county. Their marginal effects are 0.423 globally, 0.786 in Sarasota and 0.345 in Polk, respectively. It indicates that one unit increase in the access density associates with an approximate 0.141 (= 0.423/3) increase in average segment crash frequency per year, and with 0.262 (= 0.786/3) increase in Sarasota county and 0.115 (= 0.345/3) increase in Polk county per year. The positive effect of access density is generally expected and agrees with the preliminary studies (Zeng and Huang, 2014; Wang et al., 2014; Huang et al., 2014). This is because that more accesses may result in more conflict points and deteriorated safety.

What interesting to see are in Table 2 and Fig. 2d that the signs of local coefficients of the good surface condition are negative in most counties; but are positive in Lake, Okaloosa, and Clay. The global estimated mean of the regression coefficient for this factor is −0.152, and its corresponding marginal effect is −0.730. It indicates that urban two-lane two-way roadway segment with good surface condition would decrease crash frequency by −0.243 (= −0.730/3) in average per year. Overall, there is a negative association between the good surface condition and crash frequency (in 31 counties among 34 counties), which is consistent with prior studies (Anastasopoulos and Mannering, 2009; Anastasopoulos, et al., 2012; Lee, et al., 2015b). The counterintuitive safety effects of this variable in Lake, Okaloosa and Clay may due to the fact that drivers reside in these three counties are generally likely to drive with higher speed and less cautiously on the pavement with good surface condition, which is a manifestation of risk-compensation behavior (Chen et al., 2017). This inconsistent sign of coefficient estimations among different counties strongly indicates the safety effect of some road level variables could be largely influenced by regional-level and individual-level factors.

5. Conclusions and implications

This study seeks to quantitatively investigate the underlying regional variations in effects of road level factors on crash frequency. The relationship between crash frequency and road level factors should not be considered spatially constant, since many regional-level and individual-level unobserved or uncollected factors may be correlated with these road level factors. In modeling crash frequency, these correlations would introduce variations in the effects of selected variables on crash likelihood. To this end, a hierarchical random parameter model is proposed to investigate the regional varying relationships between the crash frequency and road level factors. For the purpose of comparison, a Poisson lognormal model and a hierarchical random intercept model are also developed.

Using the crash data during three-year period and their related road level factors for urban two-lane two-way highway segments in Florida as a case study, the results of model parameter estimations demonstrate that regression coefficients of all investigated factors (i.e., traffic volume, road length, access density and surface condition) and their corresponding marginal effects are significantly different over a wide range among different counties. In regard to the model comparison, the hierarchical random parameter model outperforms the hierarchical random intercept model and the Poisson lognormal model in terms of goodness-of-fit as measured by DIC. These results confirm the reasonability and necessity of the use of hierarchical random parameter model to analyze the crash frequency for road entities with hierarchical structure.

The results obtained from this study have important implications for crash analysts to develop accurate models and traffic safety engineers to establish effective countermeasures. First, this study provides a potential in guidance of model construction of crash frequency models that consider regional variations in safety effects of road level factors. By viewing road entities and their located geographic region as a two-level hierarchical system, the proposed hierarchical random parameter model can provide accurate crash frequency predictions that are adapt to multiple regions, and obtain customized-to-region information about expected crash frequency of road entities and their relationships with road level factors. Second, this study reveals that there is a large variation in effects of some road-level factors on crash frequency across counties. It implies that a direct use of crash frequency models from other regions for a specified region may produce incorrect inference, which agrees with previous points that findings from a traffic safety study are not necessarily transferable between distant geographic locations (Elvik, 2013; Liu et al., 2017; Farid et al., 2018). Thus, it is very essential for traffic agencies to use location-referenced crash data and safety contributing factors to develop local safety analysis models in order to accurately identify crash-prone location and better formulate safety plans to effectively enhance road traffic safety.

Several limitations should be noted for this study. First, region-level factors are not considered in this study. In the follow-up study, important region-level demographic and socioeconomic variables (e.g., population by age group, industry, income, employment) could be incorporated into hierarchical random parameter models to further explore these variability patterns of road-level factors. For example, it is worthwhile to see if the effects of road-level variables are still significant different after controlling for region-level variables. Second, there may be micro-level spatial correlations among adjacent road entities, which are caused by similar traffic volumes, traffic controls and road features of adjacent road entities. The modeling approach that incorporates the spatial correlation into the hierarchical random parameter model will be further explored to track the source of varying effects of road-level factors. Third, only a specific roadway facility type, two-lane two-way urban road, has been explored. It is worthwhile to further investigate the varying relationships between intersection crash frequency and intersection-related road factors across different regions.

Acknowledgements

This work was jointly supported by: 1) the Joint Research Scheme of National Natural Science Foundation of China/Research Grants Council of Hong Kong (Project No. 71561167001 & N_HKU707/15), 2) the Natural Science Foundation of China (No. 713711921). We would like to thank Dr. Mohamed Abdel-Aty at the University of Central Florida and the Florida Department of Transportation for providing the data.

References

Abdel-Aty, M., Wang, X., 2006. Crash estimation at signalized intersections along corridors: analyzing spatial effect and identifying significant factors. Transportation Research Record 1953 (1), 98–111.

Adanu, E., Smith, R., Powell, L., Jones, S., 2017. Multilevel analysis of the role of human factors in regional disparities in crash outcomes. Accident Analysis and Prevention 109, 10–17.

Aguero-Valverde, J., 2013. Multivariate spatial models of excess crash frequency at area level: case of Costa Rica. Accident Analysis and Prevention 59, 365–373.

Aguero-Valverde, J., Jovanis, P., 2006. Spatial analysis of fatal and injury crashes in Pennsylvania. Accident Analysis and Prevention 38 (3), 618–625.

Alarifi, S., Abdel-Aty, M., Lee, J., Park, J., 2017. Crash modeling for intersections and segments along corridors: a Bayesian multilevel joint model with random parameters. Analytic Methods in Accident Research 16, 48–59.

Anastasopoulos, P., 2016. Random parameters multivariate Tobit and zero-inflated count data models: addressing unobserved and zero-state heterogeneity in accident injury-severity rate and frequency analysis. Analytic Methods in Accident Research 11, 17–32.

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., 2016. The effect of speed limits on drivers' choice of speed: a random parameters seemingly unrelated equations approach. Analytic Methods in Accident Research 10, 1–11.

Anastasopoulos, P., Shankar, V., Haddock, J., Mannering, F., 2012. A multivariate Tobit analysis of highway accident-injury-severity rates. Accident Analysis and Prevention 45 (1), 110–119.

Barua, S., El-Basyouny, K., Islam, M., 2016. Multivariate random parameters collision count data models with spatial heterogeneity. Analytic Methods in Accident Research 9, 1–15.

Behnood, A., Mannering, F., 2017. The effect of passengers on driver-injury severities in single-vehicle crashes: a random parameters heterogeneity-in-means approach. Analytic Methods in Accident Research 14, 41–53.

Bhat, C., Astroza, S., Lavieri, P., 2017. A new spatial and flexible multivariate random-coefficients model for the analysis of pedestrian injury counts by severity level. Analytic Methods in Accident Research 16, 1–22.

Cai, Q., Abdel-Aty, M., Lee, J., Wang, L., Wang, X., 2018. Developing a grouped random parameters multivariate spatial model to explore zonal effects for segment and intersection crash modeling. Analytic Methods in Accident Research 19, 1–15.

Chen, S., Saeed, T., Labi, S., 2017. Impact of road-surface condition on rural highway safety: a multivariate random parameters negative binomial approach. Analytic Methods in Accident Research 16, 75–89.

Christoffel, T., Gallagher, S., 1999. Injury Prevention and Public Health: Practical Knowledge, Skills, and Strategies. Aspen Publishers Inc., Gaithersburg, MD.

Coruh, E., Bilgic, A., Tortum, A., 2015. Accident analysis with the random parameters negative binomial panel count data model. Analytic Methods in Accident Research 7, 37–49.

Ding, C., Wang, Y., Yang, J., Liu, C., Lin, Y., 2016. Spatial heterogeneous impact of built environment on household auto ownership levels: evidence from analysis at traffic analysis zone scales. Transportation Letters 8 (1), 26–34.

Ding, C., Mishra, S., Lu, G., Yang, J., Liu, C., 2017. Influences of built environment characteristics and individual factors on commuting distance: a multilevel mixture hazard modeling approach. Transportation Research Part D 51, 314–325.

Elvik, R., 2013. International transferability of accident modification functions for horizontal curves. Accident Analysis and Prevention 59, 487–496.

Emmanuelle, D., Eleonora, P., Heike, M., George, Y., 2013. Multilevel analysis in road safety research. Accident Analysis and Prevention 60 (1), 402–411.

Erdogan, S., 2009. Explorative spatial analysis of traffic accident statistics and road mortality among the provinces of Turkey. Journal of Safety Research 40 (5), 341–351.

Farid, A., Abdel-Aty, M., Lee, J., 2018. Transferring and calibrating safety performance functions among multiple states. Accident Analysis and Prevention 117, 276–287.

Fountas, G., Anastasopoulos, P., 2017. A random thresholds random parameters hierarchical ordered Probity analysis of highway accident injury-severities. Analytic Methods in Accident Research 15, 1–16.

Fountas, G., Anastasopoulos, P., Abdel-Aty, M., 2018. Analysis of accident injury-severities using a correlated random parameters ordered Probit approach with time variant covariates. Analytic Methods in Accident Research 18, 57–68.

Gelman, A., Hill, J., 2007. Data Analysis Using Regression and Multilevel/ Hierarchical Models. Cambridge University Press.

Gooch, J., Gayah, V., Donnell, E., 2016. Quantifying the safety effects of horizontal curves on two-way, two-lane rural roads. Accident Analysis and Prevention 92, 71–81.

Heydari, S., Fu, L., Thakali, L., Joseph, L., 2018. Benchmarking regions using a heteroskedastic grouped random parameters model with heterogeneity in mean and variance: applications to grade crossing safety analysis. Analytic Methods in Accident Research 19, 33–48.

Huang, B., Lu, L., Zhang, Y., Lu, J., 2014. A new access density definition and its correlation with crash rates by microscopic traffic simulation method. Accident Analysis and Prevention 64, 111–122.

Huang, H., Abdel-Aty, M., 2010. Multilevel data and Bayesian analysis in traffic safety. Accident Analysis and Prevention 42 (6), 1556–1565.

Huang, H., Abdel-Aty, M., Darwiche, A., 2010. County-level crash risk analysis in Florida: Bayesian spatial modeling. Transportation Research Record 2148 (2148), 27–37.

Huang, H., Zhou, H., Wang, J., Chang, F., Ma, M., 2017. A multivariate spatial model of crash frequency by transportation modes for urban intersections. Analytic Methods in Accident Research 14, 10–21.

Jones, A., Jorgensen, S., 2003. The use of multilevel models for the prediction of road accident outcomes. Accident Analysis and Prevention 35 (1), 59–69.

Kim, J., Ulfarsson, G., Shankar, V., Mannering, F., 2010. A note on modeling pedestrian-injury severity in motor-vehicle crashes with the mixed logit model. Accident Analysis and Prevention 42 (6), 1751–1758.

Lee, J., Abdel-Aty, M., Cai, Q., 2017. Intersection crash prediction modeling with macro-level data from various geographic units. Accident Analysis and Prevention 102, 213–226.

Lee, J., Abdel-Aty, M., Choi, K., Huang, H., 2015a. Multi-level hot zone identification for pedestrian safety. Accident Analysis and Prevention 76, 64–73.

Lee, J., Nam, B., Abdel-Aty, M., 2015b. Effects of pavement surface conditions on traffic crash severity. Journal of Transportation Engineering 141 (10), 04015020.

Li, Z., Wang, W., Liu, P., Bigham, J., Ragland, D., 2013. Using geographically weighted Poisson regression for county-level crash modeling in California. Safety Science 58 (10), 89–97.

Liu, C., Sharma, A., 2017. Exploring spatio-temporal effects in traffic crash trend analysis. Analytic Methods in Accident Research 16, 104–116.

Liu, C., Sharma, A., 2018. Using the multivariate spatio-temporal Bayesian model to analyze traffic crashes by severity. Analytic Methods in Accident Research 17, 14–31.

Liu, J., Khattak, A., Wali, B., 2017. Do safety performance functions used for predicting crash frequency vary across space? Applying geographically weighted regressions to account for spatial heterogeneity. Accident Analysis and Prevention 109, 132–142.

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

Mannering, F., 2018. Temporal instability and the analysis of highway accident data. Analytic Methods in Accident Research 17, 1–13.

Mannering, F., Bhat, C., 2014. Analytic methods in accident research: methodological frontier and future directions. Analytic Methods in Accident Research 1, 1–22.

Mannering, F., Shankar, V., Bhat, C., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Analytic Methods in Accident Research 11, 1–16.

Miranda-Moreno, L., Morency, P., El-Geneidy, A., 2011. The link between built environment, pedestrian activity and pedestrian-vehicle collision occurrence at signalized intersections. Accident Analysis and Prevention 43 (5), 1624–1634.

Mitra, S., Washington, S., 2012. On the significance of omitted variables in intersection crash modeling. Accident Analysis and Prevention 49, 439–448.

Ntzoufras, I., 2009. Bayesian Modeling Using WinBUGS. Wiley.

Quddus, M., 2008. Modeling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data. Accident Analysis and Prevention 40 (4), 1486–1497.

Raudenbush, S., Bryk, A., 2002. Hierarchical Linear Models: Applications and Data Analysis Methods, second ed. Sage, Thousand Oaks, CA.

Russo, B., Savolainen, P., Schneider, W., Anastasopoulos, P., 2014. Comparison of factors affecting injury severity in angle collisions by fault status using a random parameters bivariate ordered Probit model. Analytic Methods in Accident Research 2, 21–29.

Ryb, G., Dischinger, P., Kufera, J., Soderstrom, C., 2007. Social, behavioral and driving characteristics of injured pedestrians: a comparison with other unintentional trauma patients. Accident Analysis and Prevention 39 (2), 313–318.

Sarwar, M., Anastasopoulos, P., Golshani, N., Hulme, K., 2017. Grouped random parameters bivariate Probit analysis of perceived and observed aggressive driving behavior: a driving simulation study. Analytic Methods in Accident Research 13, 52–64.

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

Shankar, V., Albin, R., Milton, J., Mannering, F., 1998. Evaluation of median crossover likelihoods with clustered accident counts: an empirical inquiry using the random effect negative binomial model. Transportation Research Record 1635, 44–48.

Shaon, M., Xiao, Q., Shirazi, M., Lord, D., Geedipally, S., 2018. Developing a random parameters negative Binomial-Lindley model to analyze highly over-dispersed

crash count data. Analytic Methods in Accident Research 18, 33–44.
Song, 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.
Spiegelhalter, D., Best, N., Carlin, B., Linde, V., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64 (4), 583–639.
Truong, L., Kieu, L., Vu, T., 2016. Spatiotemporal and random parameter panel data models of traffic crash fatalities in Vietnam. Accident Analysis and Prevention 94, 153–161.
Ukkusuri, S., Miranda-Moreno, L., Ramadurai, G., Isa-Tavarez, J., 2012. The role of built environment on pedestrian crash frequency. Safety Science 50, 1141–1151.
Venkataraman, N., Ulfarsson, G., Shankar, V., 2013. Random parameter models of interstate crash frequencies by severity, number of vehicles involved, collision and location type. Accident Analysis and Prevention 59 (4), 309–318.
Venkataraman, N., Shankar, V., Ulfarsson, G., Deptuch, D., 2014. A heterogeneity-in-means count model for evaluating the effects of interchange type on heterogeneous influences of interstate geometrics on crash frequencies. Analytic Methods in Accident Research 2, 12–20.
Wang, J., Huang, H., 2016. Road network safety evaluation using Bayesian hierarchical joint model. Accident Analysis and Prevention 90, 152–158.
Wang, J., Huang, H., Zeng, Q., 2017. The effect of zonal factors in estimating crash risks by transportation modes: motor vehicle, bicycle and pedestrian. Accident Analysis and Prevention 98, 223–231.
Wang, X., Song, Y., Yu, R., Schultz, G., 2014. Safety modeling of suburban arterials in Shanghai China. Accident Analysis and Prevention 70, 215–224.
Xu, P., Huang, H., 2015. Modeling crash spatial heterogeneity: Random parameter versus geographically weighting. Accident Analysis and Prevention 75, 16–25.
Xu, P., Huang, H., Dong, N., Wong, S., 2017. Revisiting crash spatial heterogeneity: a Bayesian spatially varying coefficients approach. Accident Analysis and Prevention 98, 330–337.
Zeng, Q., Huang, H., 2014. Bayesian spatial joint modeling of traffic crashes on an urban road network. Accident Analysis and Prevention 67, 105–112.