Accident Analysis and Prevention 128 (2019) 164–174

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Modeling correlation and heterogeneity in crash rates by collision types using full bayesian random parameters multivariate Tobit model

Yanyong Guo[a,b], Zhibin Li[a,*], Pan Liu[a], Yao Wu[a]

[a] Jiangsu Key Laboratory of Urban ITS, Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, School of Transportation, Southeast University, Si Pai Lou #2, 210096, Nanjing, China

[b] Department of Civil Engineering, University of British Columbia, 6250 Applied Science Lane, V6T 1Z4, Vancouver, BC, Canada

ARTICLE INFO

Keywords:
Crash rates
Collision types
Multivariate Tobit model
Random parameters
Full bayesian estimation

ABSTRACT

Crashes present different collision types. There usually exist unobserved risk factors which could jointly affect crash rates of different types, resulting in correlation and heterogeneity issues across observations. The primary objective of the study is to propose a novel random parameters multivariate Tobit (RPMV-Tobit) model for evaluating risk factors on crash rates of different collision types. Crash data from 367 freeway diverge areas in a three-year period were obtained for modeling. Three major types of collisions including rear-end, sideswipe, and angle collisions were considered. The RPMV-Tobit model was structured to simultaneously accommodate correlations between crash rates across collision types and unobserved heterogeneity across observations. The RPMV-Tobit model was compared with a multivariate Tobit (MV-Tobit) model, a random effect multivariate Tobit (REMV-Tobit) model, and independent univariate Tobit (IU-Tobit) models under the Bayesian framework. The results showed that MV-Tobit model outperforms the IU-Tobit models on fitting crash rates, indicating that accounting for the correlation between crash types can improve model fit. The RPMV-Tobit model and REMV-Tobit model perform better than the MV-Tobit model, suggesting that accounting for the unobserved heterogeneous can further improve model fit. The improvement of model performance with the RPMV-Tobit model is higher than that with the REMV-Tobit model. The impacts of each risk factor on crash rates were estimated and some differences were found across different collision types. The lane-balanced design, number of lanes on mainline, speed limit, and speed difference present significant heterogeneous effects on crash rates. Findings suggest that the RPMV-Tobit model is a superior approach for comprehensive crash rates modeling and traffic safety evaluation purposes.

1. Introduction

Freeway diverge areas play an important role in diverging the exiting traffic from the through traffic on freeway mainline. However, they have been considered to be crash-prone locations due to the disturbances caused by the intense lane changes at diverge areas (Chen et al., 2009; Liu et al., 2009; Li et al., 2015). These disturbances may result in many traffic conflicts, leading to high potential of traffic crashes. An efficient way to improve the safety of freeway diverge areas is to understand the risk factors affecting crashes at freeway diverge areas. As such, analytical tools need to be developed that can aid transportation safety professionals in mitigation of crashes. Previous studies have developed several safety performance functions (SPFs) to evaluate the safety performance of freeway exit ramp areas (Bauer and Harwood, 1998; McCartt et al., 2004; Lord and Bonneson, 2005; Moon and Hummer, 2009; Chen et al., 2009; Lu et al., 2010; Wu et al., 2014). Though it is expected that the impact of risk factors on different collision types are different, previous studies did not distinguish the collision types during the modeling analyses of crash rates. This study aims to add to the current literature by proposing a methodological approach that takes into account crash rates by collision types at freeway diverge areas. Moreover, this paper seeks to shed light on possible risk factors related to crash rates of different collision types at freeway diverge areas.

Crash rates analysis is now advocated as an alternative of crash frequency prediction models because of several advantages. Crash rates provide a standardized safety measure of road entities, which is more understandable and acceptable to the public. Moreover, crash rates are commonly used in the accident reporting systems (NHTSA, 2012). As such, the development of crash rates model has many potential

* Corresponding author.
E-mail addresses: yanyong.guo@ubc.ca (Y. Guo), lizhibin@seu.edu.cn (Z. Li), liupan@seu.edu.cn (P. Liu), wuyaomaster@126.com (Y. Wu).

https://doi.org/10.1016/j.aap.2019.04.013
Received 14 December 2018; Received in revised form 28 February 2019; Accepted 17 April 2019
Available online 29 April 2019

applications and benefits. For example, the crash rate models were considered to be a complementary tool for road safety diagnosis in hotspots identification (Ma et al., 2015a; Xu et al., 2014; Ma et al., 2018). The traffic safety evaluations in before-after studies could be enhanced by combining crash count models and crash rates models.

Different from the features of discrete and non-negative for crash frequency, crash rates are continuous data that is usually left-censored at zero because no crashes may be reported over a specified time period. To deal with the censoring problem, the Tobit model was proposed to model crash rates by previous study (Anastasopoulos et al., 2008). The model has attracted considerable interest in recent traffic safety studies (Anastasopoulos et al., 2012a; Xu et al., 2013; Chen et al., 2014; Xu et al., 2014; Ma et al., 2015b; Yu et al., 2015; Caliendo et al., 2016; Bin Islam and Hernandez, 2016; Anastasopoulos, 2016; Zeng et al., 2017a, b; Anderson and Hernandez, 2017; Sarwar and Anastasopoulos, 2017; Zeng et al., 2018).

Crashes present different collision types due to the particular geometric and traffic features at various freeway diverge areas. Compared with analysis of total crash rates at freeway diverge area, modeling crash rates by collision types can provide better understanding of the impact of risk factors on the crash rates with a particular collision type. However, if crash rates are modeled independently irrespective of correlations among crash rates across collision types, significant estimation error could be introduced because unobserved effects at freeway diverge areas are likely to be shared among different types of collisions (Anastasopoulos et al., 2012a; Anastasopoulos, 2016; Zeng et al., 2017a; Sarwar and Anastasopoulos, 2017; Zeng et al., 2018). Meanwhile, the effect of risk factors on crash rates may vary across observations. If such unobserved heterogeneous effect is ignored, the model fit will be reduced and may result in biased parameter estimation and erroneous inferences (Anastasopoulos et al., 2012b; Chen et al., 2014; Yu et al., 2015; Caliendo et al., 2016; Bin Islam and Hernandez, 2016; Zeng et al., 2017b).

The objective of this study is to simultaneously analyze crash rates by collision types at freeway diverge areas via developing a random parameters multivariate Tobit (RPMV-Tobit) model, which accommodates both correlation between crash rates across collision types and unobserved heterogeneity across observations. Three-year period (2004–2006) crash data, including three types of collisions (i.e., rear-end, sideswipe, and angle), from 367 freeway diverge areas are used for the analysis. Four types of freeway diverge areas are identified according to the arrangement of lanes for traffic to exit. Three candidate Tobit models, i.e. multivariate Tobit (MV-Tobit) model, random effect Tobit (REMV-Tobit) model, and independent univariate Tobit (IU-Tobit) model were also estimated and compared with the RPMV-Tobit model under the Bayesian framework.

2. Literature review

2.1. Safety analysis of freeway diverge areas

Previously, numerous studies have been conducted to evaluate the safety performance of freeway diverge areas using the SPFs such as Poisson and negative binomial (NB) model, Poisson-lognormal model, zero-inflated count model, negative multinomial model, and generalized estimating equation model (Bauer and Harwood, 1998; Bared et al., 1999; McCartt et al., 2004; Golob et al., 2004; Lord and Bonneson, 2005; Garcia and Romero, 2006; Moon and Hummer, 2009; Chen et al., 2009; Liu et al., 2009; Lu et al., 2010; Chen et al., 2011a; Wu et al., 2014; Li et al., 2015; Papadimitriou and Theofilatos, 2017). The findings showed that risk factors related to traffic volume (e.g. mainline AADT, and ramp AADT), traffic operation (e.g. speed limit, speed difference, and proportion of heavy vehicles), ramp design elements (e.g. lane-balanced design, arrangement of lanes, ramp configuration, and left/right-side off-ramps), geometric design (e.g. right shoulder width, ramp length, and length of speed-change lane), road environment (e.g. rural or urban area) have significant impact on crash counts at freeway diverge areas.

Some studies analyzed the crash injury severity at freeway diverge areas using discrete choice models and machine learning approaches (Wang et al., 2009; Chen et al., 2011b; Yang et al., 2011; Wang et al., 2011; Li et al., 2012, 2015). For example, Wang et al. (2011) and Yang et al. (2011) applied ordered Probit (OP) regression to relate the severity of crashes to various explanatory variables at freeway exit segments. Li et al. (2012) compared the support vector machine (SVM) model and OP model for crash injury severity analysis at freeway diverge areas. The findings from previous studies showed that factors including traffic volume, deceleration lane length, ramp lane length, number of mainline lane, curve and grade at diverge areas, light and weather conditions, road surface condition, land type, and crash type significantly affect the severity of crashes at freeway diverge areas. The only study on crash analysis by collision type at freeway diverge areas was by Li et al. (2015). They evaluated the crash risks by collision type using a multivariate Poisson-lognormal (MVPLN) model. The results showed that the MVPLN model can capture the correlation of latent effects among the crash counts of different collision types.

2.2. Crash rates analysis

Compared to the traditional and widely used count-data models for crash frequency (Lord and Mannering, 2010; Mannering and Bhat, 2014), crash rates analysis is an emerging technique during the past decade. To date, crash rates data are mainly analyzed using the Tobit model which is firstly introduced by Anastasopoulos et al. (2008) to transportation safety field. Since then, various random parameters Tobit (RP-Tobit) models were proposed to account for the unobserved heterogeneity across observations (Anastasopoulos et al., 2012b; Chen et al., 2014; Yu et al., 2015; Ma et al., 2015b; Caliendo et al., 2016; Bin Islam and Hernandez, 2016; Zeng et al., 2017b; Anderson and Hernandez, 2017). The findings from these studies showed that the RP-Tobit models have better model performance than the fix parameters Tobit model.

Multivariate Tobit models are recently developed to jointly model crash rates by injury severity (Anastasopoulos et al., 2012a; Anastasopoulos, 2016; Sarwar and Anastasopoulos, 2017). All the model estimation results showed that significant correlation exists between crash rates at various severity levels. Moreover, the heterogeneous effects of some risk factors on crash injury-severity rates were found in the RPMV-Tobit model used by Anastasopoulos (2016) and Zeng et al. (2017a). Zeng et al. (2018) and Dong et al. (2018) also found significant temporal effects on crash injury-severity rates using multivariate temporal/dynamic Tobit model.

2.3. Unobserved heterogeneity in crash analysis

A growing body of advanced methodologies can be seen in accident analysis (Sarwar et al., 2017; Alarifi et al., 2018; Lee et al., 2018; Chen et al., 2018; Hossain et al., 2019; Song and Noyce, 2019), particularly those focusing on dealing with the unobserved heterogeneity in crash severity (Fountas and Anastasopoulos, 2017, 2018; Fountas et al., 2018a, b), crash frequency (Guo et al., 2019a; Xie et al., 2019), crash rates (Ulak et al., 2018; Zeng et al., 2018), and occurrence of accident (Fountas et al., 2018c; Guo et al., 2019b).

2018;Fountas and Anastasopoulos (2017,2018) and Fountas et al. (2018a, b) proposed various random parameters- and latent class- ordered probit models to account for the unobserved heterogeneous effects of explanatory variables and/or unobserved threshold heterogeneity in accident injury-severity analysis. Guo et al. (2019a) and Xie et al. (2019) developed random parameters multivariate Poisson-lognormal model and random parameters multivariate Poisson-Gamma model to accommodate the unobserved heterogeneity across observations in crash frequency analysis by collision types and by injury-severity respectively. Zeng et al. (2018) and Ulak et al. (2018) developed multivariate random parameters Tobit models in crash rates analysis and found significant heterogeneous effects of predictors as well as temporal and/or spatial variations across observations. Fountas et al. (2018c) and Guo et al. (2019b) utilized a grouped random parameters- and a random effect- binary logit regression in analyzing the occurrence of accident and secondary accident respectively.

The body of literature review shows that advanced methodologies are booming in accident analysis, focusing on specific issues such as unobserved heterogeneity. In contrast to many studies on crash counts, no study was conducted to analyze crash rates by collision types at freeway diverge areas. Previous studies have limited capacity to explain the risk factors related to crash rates of different collision types at freeway diverge areas. An evaluation on crash rates by collision types could enhance and provide a more comprehensive understanding in the mechanism of different crash risks at freeway diverge areas.

3. Data

Crash data for a three-year period were collected from 367 freeway diverge areas in the State of Florida, United States. Risk factors including the road geometries and traffic exposures were collected as explanatory variables (See Table 1). The diverge area defined in this paper covers a deceleration lane and an exit. More specifically, the diverge contains two influence areas, including an area located within 1500 ft upstream of the painted nose, and an area located within 1000 ft downstream of the pained nose (See Fig. 1). Thus, the freeway diverge area has a consistent length of 2500 ft (762 m).

The selected freeway diverge areas were classified into four types based on the arrangement of lanes for the freeway exits as shown in Fig. 1 (Chen et al., 2009; Li et al., 2015). Type 1 diverge area is installed with a single lane exit ramp with a tapered design. Type 2 diverge area is designed as a single lane exit ramp with the outer lane of freeway becoming a drop lane at the exit gore. Type 3 diverge area is installed with a two-lane exit with an optional lane. Type 4 diverge area is designed as a two-lane exit with the outer lane of the freeway dropped at the exit gore. Types 1 and type 3 diverge areas are considered as lane-balanced designs, while Type 2 and type 4 diverge areas are considered as lane-unbalanced designs (AASHTO, 2001).

Three major types of collisions are identified in our dataset [i.e., rear-end, sideswipe, and angle]. The proportions of other collision types are less than 5%. Thus, the research reported in this paper focused on three major collision types. In total, 3315 major crashes were collected at the selected 367 freeway diverge areas, including 2193 rear-end crashes (66.2%), 727 sideswipe crashes (21.9%), and 395 angle crashes (11.9%). Totally, the dataset consisted of 367 observations for each collision type.

Crash rates by collision type, which is used as the dependent variable in this study, is calculated as crashes per million vehicle kilometers traveled as follows

$$CR_i^k = \frac{10^6 \times N_i^k}{365 \times T \times L \times \sqrt{AADT_{mainline} \times AADT_{ramp}}} \tag{1}$$

where CR_i^k is the crash rates for collision type k at freeway diverge area i, N_i^k is the number of crashes for collision type k at freeway diverge area i, T is the number of years of study period (T = 3), L is the length of the freeway diverge area (L = 0.762Km), $AADT_{mainline}$ is the annual average daily traffic on mainline, and $AADT_{ramp}$ is the annual average daily traffic on exit ramp. According to the definition of crash rate (Anastasopoulos et al., 2012a; Zeng et al., 2017a), AADT should be treated as the exposure in calculating crash rate. In this study, the selected freeway diverge area covers both mainline and exit ramp. As such, both mainline AADT and ramp AADT are contributable to the crashes and should be considered. To avoid multicollinearity, they are combined by calculating the square root of the multiplication of mainline AADT and ramp AADT (Washington et al., 2010; Chen et al., 2009).

Among the 367 observations for each collision type, the rear-end crash rates of 121 (34.6%) observations, the sideswipe crash rates of 155 (42.2%) observations, and the angel crash rates of 191 (52%) observations are 0, indicating that the Tobit model is appropriate in fitting the data due to its ability to handle the censoring problem (left-censored at zero). Average crash rates by collision type for different freeway diverge areas were compared in Fig. 2. Type 4 diverge area has the highest crash rates among all collision types followed by type 2 diverge area. Type 3 and type 1 diverge areas are found to have the lowest crash rates. In general, crash rates by collision type for lane-balanced design diverge areas (type 1 & type 3) vary from 0 to 13.22 with a mean of 0.88, while crash rates for lane-unbalanced design diverge areas (type 2 & type 4) vary from 0 to 17.86 with a mean of 1.52. *T*-test shows that crash rates between lane-balanced and lane-unbalanced diverge areas are significantly different at the 95% confidence

Table 1
Variables definition and descriptive statistics summary.

Variables	Description	Mean	SD	Min	Max	Frequency (%)
Dependent variable						
Rear-end	Rear-end crashes per million vehicle kilometers traveled	2.06	2.86	0	17.86	367
Sideswipe	Sideswipe crashes per million vehicle kilometers traveled	0.71	1.01	0	5.58	367
Angle	Angle crashes per million vehicle kilometers traveled	0.42	0.67	0	5.24	367
Explanatory variables						
Ln(MAADT)	Logarithm of mainline AADT	6.93	0.67	4.88	9.47	367
Ln(RAADT)	Logarithm of ramp AADT	4.42	0.88	1.39	7.38	367
Mainlane	Number of lanes on mainline	3.41	1.02	2	6	367
Ramlane	Number of lanes on exit ramp	1.22	0.42	1	2	367
LengRam	Exit ramp length (mi)	0.36	0.20	0.10	1.88	367
LengDec	Deceleration lanes length (mi)	0.05	0.03	0.01	0.28	367
SpeedLimit	Post speed limit on mainline (mi/h)	64.85	6.27	50	70	367
DiffSpeed	Difference in speed limit between mainline and exit ramp (mi/h)	28.87	8.51	10	45	367
ShouWidth	Right shoulder width (ft)	10.06	0.99	1	15	367
LaneBala	Lane-balanced design					367
	1, Type 1 or 3					263(71.7%)
	0, Type 2 or 4					104(28.3%)
SurfType	Road surface type					367
	1, blacktop					333(90.7%)
	0, others					34(9.3%)
ShouType	Right shoulder type					367
	1, paved					192(52.3%)
	0, unpaved					175(47.7%)

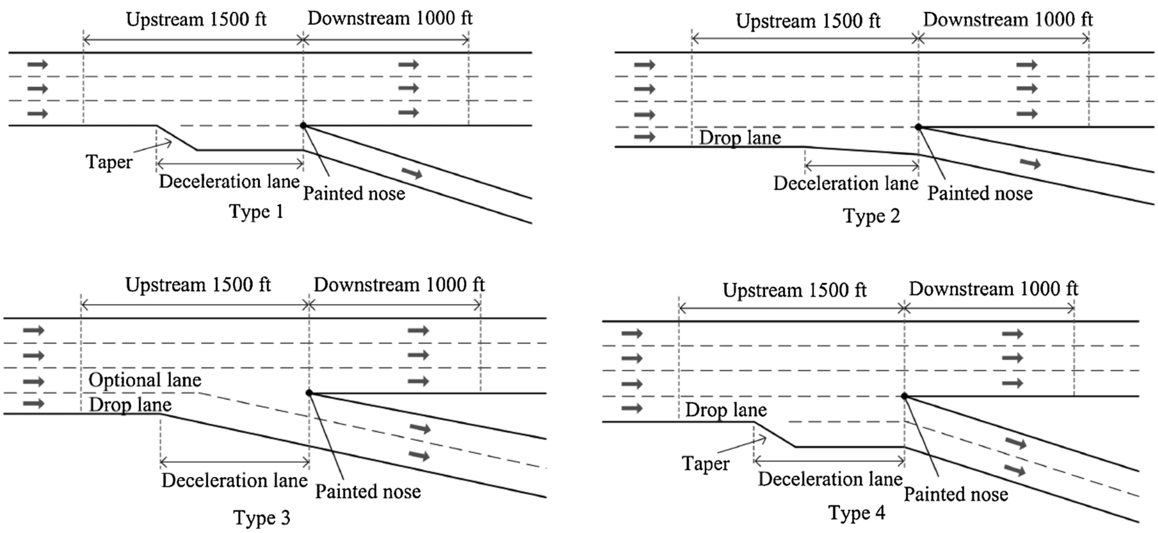

Fig. 1. Illustrations of four types of freeway diverge area.

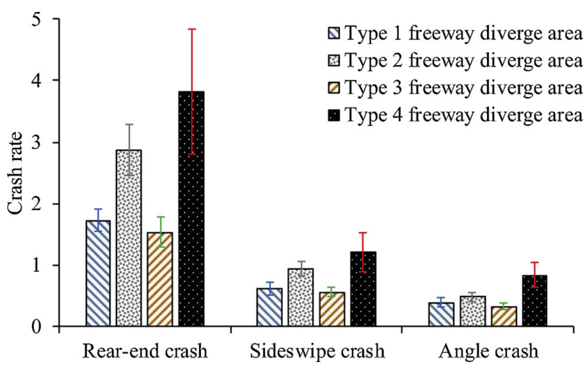

Fig. 2. Crash rates by collision type between freeway diverge areas.

level, indicating that the lane-balanced ramps are safer than those not balanced.

4. Methodology

4.1. Tobit model

The Tobit model was first proposed by Tobin (1958) for modeling the continuous dependent variable which is left-censored, right-censored, or both. Given that crash rates are usually left-censored at zero, Anastasopoulos et al. (2008) introduced the Tobit model into road safety evaluation. The Tobit model for fitting crash rates is given as

$$Y_i^* = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ,\cdots,+\beta_J X_{iJ} + \varepsilon_i \tag{2}$$

$$Y_i = \begin{cases} Y_i^*, & if\, Y_i^* > 0 \\ 0, & if\, Y_i^* \le 0 \end{cases}, i = 1, 2, \cdots N \tag{3}$$

where Y_i is the dependent variable (observed values of crash rates), X_{ij} is the jth ($j = 1, 2, \ldots, J$) explanatory variable for observation i, β_0 is the model intercept, $\beta_1, \beta_2, \ldots, \beta_J$ are the coefficients, Y_i^* is a latent variable observed only when positive, N is the number of observations, and ε_i denotes the unstructured error which is assumed to be normally and independently distributed with zero mean and variance σ_ε^2 as given in

$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2) \tag{4}$$

4.2. Multivariate Tobit model

The Tobit model can deal with crash rates for a specific collision type. However, crash rates may be correlated across various collision types because unobserved effects at the diverge areas are likely to be shared among different types of collision. To account for the possible correlation between crash rates across collision types, a multivariate Tobit (MV-Tobit) model was developed in our study. The MV-Tobit model for the joint modeling of crash rates by collision type is expressed as

$$Y_i^{k*} = \beta_0^k + \beta_1^k X_{i1} + \beta_2^k X_{i2} + ,\cdots,+\beta_J^k X_{iJ} + \varepsilon_i^k \tag{5}$$

$$Y_i^k = \begin{cases} Y_i^{k*}, \, if\, Y_i^{k*} > 0 \\ 0, \, if\, Y_i^{k*} \le 0 \end{cases}, i = 1, 2, \cdots N, \quad k = 1, 2, \cdots K \tag{6}$$

where Y_i^k is the dependent variable for the kth collision type for observation i, X_{ij} is the jth explanatory variable for observation i, β_0^k is the model intercept, $\beta_1^k, \beta_2^k, \ldots, \beta_J^k$ are the coefficients corresponding to kth collision type, Y_i^{k*} is a latent variable observed only when positive, N is the number of observations, K is the number of collision type, and ε_i^k denotes the multivariate normal errors distributed as $\boldsymbol{\varepsilon_i} \sim N_k(\mathbf{0}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\varepsilon_i} = \begin{pmatrix} \varepsilon_i^1 \\ \varepsilon_i^2 \\ \cdots \\ \varepsilon_i^K \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \cdots & \sigma_{1K}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \cdots & \sigma_{2K}^2 \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{K1}^2 & \sigma_{K2}^2 & \cdots & \sigma_{KK}^2 \end{pmatrix} \tag{7}$$

where the diagonal element σ_{kk}^2 ($k = 1, 2, \ldots, K$) represents the variance of error term ε_i^k, the off-diagonal element $\sigma_{k1,k2}^2$ ($k1 \neq k2$) denotes the covariance between ε_i^{k1} and ε_i^{k2}.

4.3. Random parameters multivariate Tobit model

The MV-Tobit model assumes that the parameters of explanatory variables are fixed across observations. However, heterogeneous effect of certain risk factors may exist across observations. To accommodate the unobserved heterogeneity in the MV-Tobit model, the random parameters multivariate Tobit (RPMV-Tobit) model is proposed in our study by setting the coefficients ($\beta_0^k, \beta_1^k, \beta_2^k, \ldots, \beta_J^k$) in Eq. (5) to be random parameters ($\beta_{i0}^k, \beta_{i1}^k, \beta_{i2}^k, \ldots, \beta_{iJ}^k$), which is given as

$$Y_i^{k*} = \beta_{i0}^k + \beta_{i1}^k X_{i1} + \beta_{i2}^k X_{i2} + ,\cdots,+\beta_{iJ}^k X_{iJ} + \varepsilon_i^k \tag{8}$$

The random parameters ($\beta_{i0}^k, \beta_{i1}^k, \beta_{i2}^k, \ldots, \beta_{iJ}^k$) are assumed to be multinormally distributed as $\boldsymbol{\beta}^{ij} \sim N_k(\boldsymbol{\beta}_j, \boldsymbol{\Phi}_j)$,where

$$\boldsymbol{\beta}^{ij} = \begin{pmatrix} \beta_{ij}^1 \\ \beta_{ij}^2 \\ \cdots \\ \beta_{ij}^K \end{pmatrix}, \boldsymbol{\beta}^{j} = \begin{pmatrix} \beta_{j}^1 \\ \beta_{j}^2 \\ \cdots \\ \beta_{j}^K \end{pmatrix}, \boldsymbol{\Phi}_j = \begin{pmatrix} \varphi_{11}^{j2} & \varphi_{12}^{j2} & \cdots & \varphi_{1K}^{j2} \\ \varphi_{21}^{j2} & \varphi_{22}^{j2} & \cdots & \varphi_{2K}^{j2} \\ \cdots & \cdots & \cdots & \cdots \\ \varphi_{K1}^{j2} & \varphi_{K2}^{j2} & \cdots & \varphi_{KK}^{j2} \end{pmatrix} \tag{9}$$

It should be noted that a random parameter β_{ij}^k for collision type k is used when its posterior estimate variance φ_{kk}^{j2} is significantly larger than 0; otherwise, the parameter β_j^k should be fixed across observations for collision type k (El-Basyouny and Sayed, 2009a; Anastasopoulos, 2016).

4.4. Random effect multivariate Tobit model

Another way to account for the unobserved heterogeneity in the MV-Tobit model is to allow only the intercept β_0^k in Eq. (5) to vary across observations, which leads to the random effect multivariate Tobit (REMV-Tobit) model. The REMV-Tobit model is given as

$$Y_i^{k*} = \beta_{i0}^k + \beta_1^k X_{i1} + \beta_2^k X_{i2} + ,\cdots,+\beta_J^k X_{iJ} + \varepsilon_i^k \tag{10}$$

The intercept is set to be random parameter that follows a normal distribution as following

$$\beta_{i0}^k \sim N(\beta_0^k, \sigma_0^{k2}) \tag{11}$$

4.5. Model transferability

The crash rates models may be different across various divergence types. Exploring whether separate models by divergence type are warranted will bring up possible policy implications. In this study, we adopted the transfer index, which is widely used in previous studies (Hadayeghi et al., 2006; Sikder et al., 2014; Farid et al., 2016, 2018), to assess the transferability of models across different types of freeway diverge areas. The measure is calculated as follows

$$TI = \frac{LL_j(\beta_i) - LL_j(\beta_{j,reference})}{LL_j(\beta_j) - LL_j(\beta_{j,reference})} \tag{12}$$

where $LL_j(\beta_i)$ is the log-likelihood of the SPF, which is developed based on the data from freeway diverge type i and being applied to data of freeway diverge type j; $LL_j(\beta_j)$ is the log-likelihood of freeway diverge type j's SPF; and $LL_j(\beta_{j,\ reference})$ is the log-likelihood of freeway diverge type j' constant only SPF. The TI has an upper bound, 1.0, and no lower bound (Hadayeghi et al., 2006). The higher the TI value the better is the performance relative to the constant only model. The closer TI is to 1, the SPF developed from data i, is more transferable to freeway diverge type j. A negative TI indicates that freeway diverge type j' constant only model performs better than the SPF of freeway diverge type i applied to freeway diverge type j.

4.6. Models comparison measure

Two measures, including the Deviance Information Criteria (DIC) and the Mean Absolute Deviance (MAD) were used for models comparison. The full Bayesian estimated models can be compared using DIC, which is a measure of model complexity as follows

$$DIC = \bar{D} + p_D;\ p_D = \bar{D} - \hat{D} \tag{13}$$

where D is the un-standardized deviance of the postulated model, $\bar{D}$ is the posterior mean of D, $\hat{D}$ is the point estimate obtained by substituting the posterior means of the model's parameters in D, and p_D is a measure of model complexity estimating the effective number of parameters. Generally, the model with a smaller DIC outperforms the model with a larger DIC. According to Spiegelhalter et al. (2002), the models which DIC values' difference lower than two are considered equally well, while models with DIC values' difference between 2–7 show a considerably less support to the higher DIC model.

The MAD provides a measure of prediction performance of the model, which can be calculated as

$$\mathrm{MAD} = \frac{1}{N}\sum_{i=1}^{N} |\hat{Y}_i^k - Y_i^k| \tag{14}$$

where Y_i^k is the observed crash rate for collision type k at site i, $\hat{Y}_i$ is the predicted crash rate for collision type k at site i. A smaller value of MAD suggests that on average the model predicts the observed data better.

5. Modeling results and discussion

5.1. Models estimation

Full Bayesian (FB) method has been advocated for model estimation as it can deal with sophisticated models, particularly for those do not have closed-form likelihood functions (Lord and Mannering, 2010). The specification of prior distribution of the model parameters are required before the FB estimates. To be specifically, the model parameters are coefficients (β_0, β_j) and variance σ_ε^2 in the IU-Tobit model; coefficients (β_0^k, β_j^k) and covariance matrix $\boldsymbol{\Sigma}$ in the MV-Tobit model; hyper-coefficients (β_0^k, β_j^k) and covariance matrix $\boldsymbol{\Phi}_j$ and $\boldsymbol{\Sigma}$ in the RPMV-Tobit model; hyper-intercept β_0^k and its variance σ_0^{k2}, coefficients β_j^k, and covariance matrix $\boldsymbol{\Sigma}$ in the REMV-Tobit model. Because of the absence of sufficient prior knowledge, non-informative priors were specified for the parameters. Generally, a diffused normal distribution $N(0, 10^4)$ is used as priors for coefficients (El-Basyouny and Sayed, 2009a). A diffused gamma distribution Gamma(0.001, 0.001) is used as the prior of precision for σ_ε^{-2} and σ_0^{k-2} (El-Basyouny and Sayed, 2009a; Guo et al., 2018a, b, c). A Wishart poir $W(P, r)$ is used as the priors of precisions for $\boldsymbol{\Phi}_j^{-1}$and $\boldsymbol{\Sigma}^{-1}$, where p is the $K \times K$ identity matrix, and $r = K$ is the degree of freedom (El-Basyouny and Sayed, 2009b).

The Markov chain Monte Carlo (MCMC) technique is applied using WinBUGS software to sample the posterior distribution (i.e., the posterior mean and standard deviation) of the model parameters. MCMC methods use sampling technique to generate chains of random points, the distribution of which converge to the target posterior distribution. The convergence was monitored by several ways (El-Basyouny and Sayed, 2009a, b; Guo et al., 2018a,b,c). First, two parallel chains with diverse starting values were tracked so that full coverage of the sample space is ensured. Brooks–Gelman–Rubin (BGR) statistic was also used, where convergence occurs if the value of the BGR statistic is less than 1.2. Moreover, convergence was checked by visually inspecting the MCMC trace plots of the model parameters. As a rule of thumb, convergence occurs when the ratios of the Monte Carlo errors relative to the respective standard deviations of the estimates are less than 0.05 (El-Basyouny and Sayed, 2009b). In this study, two Marko chains run for each model for 10,000 iterations. The first 5000 iterations are discarded as burn-in sample. After the first 5000 interactions, the posterior distributions arrive convergence. As such, the interactions after 5000 iterative were adopted for parameters estimation.

In order to reduce the model estimation bias caused by the multicollinearity between explanatory variables, the Pearson correlation coefficient and the Kendall's tau-b correlation coefficient between each pair of variables were estimated before the model estimation. If two variables were found significantly correlated in the correlation analysis, they were inputted into the model form one by one while monitoring the overall model fit and the significance of the variable. The procedure for selecting model variables is a forward stepwise procedure. The initial variables are added into the model one by one. Whether to keep or remove a variable from a model is decided based on two conditions. First, the variable's parameter needs to be significant. Second, the variable should exhibit low correlation with other variables that already exist in the model. Only variables that were significant with a 90% confidence level were kept in the model. The model form with the

Table 2
Model performance comparison results.

	IU-Tobit				MV-Tobit				RPMV-Tobit				REMV-Tobit			
	Mean	SD	5.0%	95.0%	Mean	SD	5.0%	95.0%	Mean	SD	5.0%	95.0%	Mean	SD	5.0%	95.0%
σ_{11}^2	6.215	0.466	5.490	7.020	6.187	0.456	5.480	6.969	6.282	0.472	5.556	7.104	6.190	0.461	5.472	6.978
σ_{22}^2	0.877	0.066	0.775	0.990	0.874	0.064	0.773	0.984	0.871	0.067	0.767	0.986	0.875	0.065	0.774	0.988
σ_{33}^2	0.376	0.028	0.332	0.424	0.375	0.028	0.331	0.423	0.373	0.028	0.330	0.421	0.372	0.028	0.328	0.421
σ_{12}^2 (σ_{21}^2)	–	–	–	–	1.028	0.134	0.817	1.257	1.033	0.137	0.818	1.266	1.018	0.134	0.804	1.247
σ_{13}^2 (σ_{31}^2)	–	–	–	–	0.523	0.085	0.389	0.669	0.527	0.086	0.391	0.671	0.516	0.085	0.381	0.660
σ_{23}^2 (σ_{32}^2)	–	–	–	–	0.222	0.032	0.171	0.277	0.224	0.033	0.173	0.280	0.220	0.033	0.170	0.276
$\rho_{12}(\rho_{21})$ ⋇	–	–	–	–	0.442	0.042	0.370	0.510	0.443	0.043	0.369	0.510	0.437	0.042	0.365	0.505
(ρ_{31} ρ_{31})	–	–	–	–	0.343	0.046	0.266	0.418	0.344	0.047	0.265	0.418	0.340	0.046	0.263	0.413
($\rho_{23}\rho_{32}$)	–	–	–	–	0.388	0.045	0.312	0.458	0.392	0.045	0.318	0.464	0.385	0.045	0.311	0.459
MAD	3.855△	4.023△△	4.325△△△		2.524△	2.551△△	2.782△△△		1.022△	1.351△△	1.476△△△		1.756△	1.823△△	2.126△△△	
DIC	1721.12*	1002.38**	691.18***		3278.69				3262.85				3267.31			

–indicates no parameters in this model.

⋇ $\rho_{xy} = \sigma_{xy}/\sqrt{\sigma_{xx}\sigma_{yy}}$.

△ represents MAD value for rear-end crash rates.

△△ represents MAD value for sideswipe crash rates.

△△△ represents MAD value for angle crash rates.

* represents DIC value for rear-end crash rates Tobit model.

** represents DIC value for sideswipe crash rates Tobit model.

*** represents DIC value for angle crash rates Tobit model.

best statistical fit was considered as the final one.

5.2. Models comparison

Table 2 shows the results of DIC, MAD, and a number of hyper-parameters for models comparison. The MV-Tobit model provides a superior fit over the IU-Tobit models, indicating that accommodating correlation between crash rates across collision types can improve the model fit. The RPMV-Tobit model performs the best with the lowest MAD and DIC, followed by the REMV-Tobit model and the MV-Tobit model, suggesting that accommodating unobserved heterogeneity across observations could further improve statistical fits. Furthermore, the improvement of model performance with the RPMV-Tobit model is higher than that with the REMV-Tobit model.

Table 2 shows that the correlation coefficients (ρ_{12}, ρ_{13}, ρ_{23}) in the MV-Tobit model, RPMV-Tobit model, and REMV-Tobit are significant and strong in magnitude, demonstrating that correlations exist between crash rates across collision types. The significant correlations may be attributable to the common shared factors that affect crash rates across collision types simultaneously. Those factors could be weather conditions, driver familiarities to exit ramps, or other unobserved factors which are difficult to obtain and thus are not considered in the model. The correlation coefficients in the MV-Tobit model vary from 0.343 to 0.442, which is comparable with their respective counterparts that varies from 0.344 to 0.443 in the RPMV-Tobit model and that varies from 0.340 to 0.437 in the REMV-Tobit model. The random effects of rear-end crash rates (σ_{11}^2), sideswipe crash rates (σ_{22}^2), angle crash rates (σ_{33}^2), and their covariance (σ_{12}^2, σ_{13}^2, σ_{23}^2) in the RPMV-Tobit model are comparable with their respective counterparts in the REMV-Tobit model and the MV-Tobit model. These results are consistent with (Anastasopoulos, 2016) which shows that the MV-Tobit model has similar value in both correlation coefficients and error term covariance matrix with the RPMV-Tobit model. However, these results are different than (Zeng et al., 2017a) which shows that the correlation coefficient is increased dramatically from the MV-Tobit model to the RPMV-Tobit model, whereas random effects in the RPMV-Tobit model are lower than those in the MV-Tobit model.

5.3. Interpretation of parameter estimation

The parameter estimation results in the IU-Tobit models, MV-Tobit model, RPMV-Tobit model, and REMV-Tobit model are presented in Tables 3–6 respectively. Variables that are significant with a 90% confidence level are kept in the model. In general, the variables' associations are consistent in sign across all the models expect for the intercept. Nevertheless, variations in the parameters' estimates can be noticed across the models, which confirm the different mechanisms considered by the investigated models. Considering that the RPMV-Tobit model outperforms the other models, its risk factors associated with crash rates at freeway diverge areas are discussed.

According to the estimation results in Table 5, the risk factors as well as their impacts on different collision types are different. Seven risk factors [(Ln(MAADT), Ln(RAADT), LaneBala, Mainlane, LengDec, LengRam, and SpeedLimit] are found to have significant effect on rear-end crash rates. Six risk factors [(Ln(MAADT), Ln(RAADT), LaneBala, Mainlane, LengRam, and SpeedLimit] are significantly associated with sideswipe crash rates. Only four risk factors [(Ln(MAADT), Ln(RAADT), SpeedLimit, and DiffSpeed] are found to be contributed to angle crash rates. Table 7 shows the distributional effect of the random parameters of significant risk factors across observations. It is found that lane-balanced design, number of lanes on mainline, speed limit, and difference in speed limit present significant heterogeneous effect on crash rates across observations.

The coefficients of ramp AADT are negative across all collision types, indicating that crash rates decreased with the increase in ramp AADT. The result is consistent with previous studies (Anastasopoulos et al., 2012a; Zeng et al., 2017a, 2018) which showed negative correlations between crash rates and AADT. However, the mainline AADT is found to be positively related with crash rates. The result agrees with several studies (Xu, et al., 2014; Anastasopoulos, 2016; Sarwar and Anastasopoulos, 2017) and can be explained by that the trend of crash increment is larger than that of the mainline AADT at the freeway diverge areas. Chen et al. (2009) pointed out that the high mainline AADT can increase the weaving conflicts at diverge areas, resulting in a higher possibility of crash occurrence. These findings indicate the different mechanisms of impact of ramp traffic volume and mainline traffic volume on crash rates at freeway diverge areas.

Lane-balanced design (i.e. LaneBala) is negatively related to rear-end crash rates and sideswipe crash rates, implying that lane-balanced exits have lower crash rates than those not lane-balanced. The result is consistent with previous studies (Lu et al., 2010; Li et al., 2015) which showed safety benefits of lane-balanced design at freeway diverge areas. However, lane-balanced design is found to be not significantly associated with angle crash rates. It should be noted that lane-balanced design has significant heterogeneous effect on sideswipe crash rates. Table 7 shows that freeway diverge section with lane-balanced design can decrease sideswipe crash rates for 85.86% of the observations, whereas for the remaining 14.14% of the observations, the lane-balanced design can increase sideswipe crash rates.

The number of lanes on mainline (i.e. Mainlane) has a positive impact on rear-end crash rates and sideswipe crash rates, while it has no significant impact on angle crash rates. The result is somewhat inconsistence with Li et al. (2015) which showed that freeway diverge section with more lanes on mainline increased the number of crashes for all the three collision types. It is worth mentioning that this disparity may be caused by difference in modeling crash frequency and crash rates. The impact of this risk factor on rear-end crash rates and sideswipe crash rates varies across observations. Table 7 shows that the freeway diverge section with more lanes on mainline can increase rear-end crash rates for 99.87% observations and sideswipe crash rates for

Table 3
Parameter estimates of the IU-Tobit models.

Variable	Rear-end				Sideswipe				Angle			
	Mean	SD	5.0%	95.0%	Mean	SD	5.0%	95.0%	Mean	SD	5.0%	95.0%
Intercept (β_0)	4.956	2.651	0.498	9.291	0.861	0.996	−0.814	2.490	2.150	0.652	1.054	3.216
Ln(MAADT) (β_1)	0.817	0.264	0.377	1.255	0.240	0.099	0.075	0.405	0.158	0.065	0.049	0.265
Ln(RAADT) (β_2)	−0.660	0.173	−0.944	−0.376	−0.283	0.065	−0.389	−0.176	−0.255	0.042	−0.325	−0.185
LaneBala (β_3)	−0.942	0.321	−1.470	−0.412	−0.251	0.121	−0.449	−0.051	–	–	–	–
Mainlane (β_4)	0.413	0.169	0.133	0.690	0.190	0.063	0.085	0.293	–	–	–	–
LengDec (β_5)	10.240	4.130	3.462	16.940	–	–	–	–	–	–	–	–
LengRam (β_6)	−2.196	0.676	−3.296	−1.063	−0.752	0.254	−1.166	−0.327	–	–	–	–
SpeedLimit (β_7)	−0.079	0.034	−0.135	−0.023	–	–	–	–	−0.032	0.008	−0.046	−0.019
DiffSpeed (β_8)	–	–	–	–	–	–	–	–	0.011	0.006	0.002	0.021

– indicates that this variable was not significant at the 90% confidence level.

Table 4
Parameter estimates of the MV-Tobit model.

	Rear-end				Sideswipe				Angle			
Variable	Mean	SD	5.0%	95.0%	Mean	SD	5.0%	95.0%	Mean	SD	5.0%	95.0%
Intercept (β_0)	4.427	2.112	0.708	7.663	0.298	0.889	−1.207	1.779	1.815	0.540	0.935	2.707
Ln(MAADT) (β_1)	0.855	0.232	0.503	1.232	0.288	0.091	0.145	0.440	0.190	0.064	0.083	0.296
Ln(RAADT) (β_2)	−0.681	0.161	−0.954	−0.430	−0.286	0.061	−0.388	−0.184	−0.261	0.039	−0.326	−0.198
LaneBala (β_3)	−0.917	0.319	−1.448	−0.386	−0.247	0.122	−0.442	−0.042	–	–	–	–
Mainlane (β_4)	0.428	0.165	0.100	0.708	0.184	0.063	0.079	0.294	–	–	–	–
LengDec (β_5)	9.983	4.057	3.333	16.650	–	–	–	–	–	–	–	–
LengRam (β_6)	−2.198	0.672	−3.293	−1.068	−0.754	0.251	−1.162	−0.333	–	–	–	–
SpeedLimit (β_7)	−0.074	0.028	−0.116	−0.024	−0.037	0.011	−0.025	−0.019	−0.030	0.008	−0.044	−0.018
DiffSpeed (β_8)	–	–	–	–	–	–	–	–	0.011	0.006	0.001	0.020

– indicates that this variable was not significant at the 90% confidence level.

94.65% observations, whereas it can decrease rear-end crash rates for 0.13% observations and sideswipe crash rates for 5.35% observations.

The exit ramp length (i.e. LengRam) is found to be negatively associated with rear-end crash rates and sideswipe crash rates, suggesting that crash rates of these two collision types would decrease at longer ramp length. This is intuitive and reasonable since a longer ramp could store more vehicles and prevent ramp queue overspreading on the freeway mainline, resulting in lower crash rates at freeway diverge areas. The result is supported by previous studies (Chen et al., 2011a; Li et al., 2015). The positive sign for the deceleration lane length (i.e. LengDec) indicates that rear-end crash rates increase at exits with longer deceleration lanes. As pointed out by Garcia and Romero (2006), a long deceleration lane could encourage drivers to speed up before they exit the main freeway. As such, it has the potential to increase the rear-end crash risks.

The coefficients of post speed limit on mainline (i.e. SpeedLimit) are found to be negative across all collision types, implying that crash rates decrease with the increase of mainline speed limit. The result is in agreement with serval studies (Lave and Elias, 1994; Chen et al., 2009; Liu et al., 2009; Wang et al., 2011). The result is supported by the fact that freeway sections with a higher speed limit are usually designed according to higher standards with wider lanes, better signs or pavement marking, and better lighting conditions, which result in low crash rates. Table 7 shows that high speed limit on mainline can decrease rear-end crash rates, sideswipe crash rates, and angle crash rates for most of the observations, whereas it can increase crash rates by all collision types for a small proportion of observations.

Difference in speed limit (i.e. DiffSpeed) is found to be positively associated with angel crash rates, but not significantly associated with rear-end crash rates and sideswipe crash rates. This result is reasonable in that a larger difference in speed limits reflects disturbance between exiting traffic and through traffic, which contribute to the high angle crash rates. The effect of this risk factor is found to be heterogeneous across observations. Table 7 shows that large difference in speed limit can decrease angle crash rates for 93.52% observations, while it can increase angle crash rates for 6.48% observations. Furthermore, the difference in speed limit is not significant in univariate Tobit models. This indicates that the RPMV-Tobit model is capturing underlying effects (possibly those of the variables that are not significant in the univariate Tobit models) that may be masked in the error terms of the univariate models, hence explaining more of the variance in the data.

5.4. Models transferability

To test the model transferability across different types of freeway diverge areas, a RPMV-Tobit model for each type of freeway diverge area was estimated. The transfer index was calculated to assess the transferability of the SPF across freeway diverge types. The transferability assessment result is shown in Table 8.

As shown in Table 8, the application of an SPF of one freeway diverge type to itself yields a transfer index of 1 since such SPF is the representative of its local conditions. Additionally, most of the transfer indices on both sides of the diagonal are negative, indicating that the local constant only SPF outperforms the transferred SPF. The only

Table 5
Parameter estimates of the RPMV-Tobit model.

	Rear-end				Sideswipe				Angle			
Variable	Mean	SD	5.0%	95.0%	Mean	SD	5.0%	95.0%	Mean	SD	5.0%	95.0%
Intercept (β_0)	0.070	0.651	−0.899	1.188	−1.159	0.449	−2.048	−0.534	1.122	0.511	0.344	1.878
Ln(MAADT) (β_1)	1.021	0.242	0.630	1.445	0.304	0.091	0.159	0.455	0.171	0.093	0.011	0.308
Ln(RAADT) (β_2)	−0.519	0.172	−0.790	−0.224	−0.255	0.075	−0.373	−0.129	−0.242	0.065	−0.329	−0.154
LaneBala (β_3)	−0.976	0.118	−1.012	−0.812	−0.252	0.068	−0.322	−0.215	–	–	–	–
Mainlane (β_4)	0.519	0.183	0.239	0.856	0.217	0.098	0.054	0.373	–	–	–	–
LengDec (β_5)	9.575	0.721	8.214	10.433	–	–	–	–	–	–	–	–
LengRam (β_6)	−2.682	0.441	−3.338	−1.889	−0.923	0.247	−1.329	−0.506	–	–	–	–
SpeedLimit (β_7)	−0.038	0.030	−0.085	−0.025	−0.039	0.017	−0.068	−0.017	−0.024	0.016	−0.049	−0.016
DiffSpeed (β_8)	–	–	–	–	–	–	–	–	0.013	0.005	0.005	0.021
$\sigma_{\beta 0}^2$	0.661	3.906	2.64E-04	3.423	0.208	1.144	2.72E-04	1.085	0.091	0.468	2.57E-04	0.446
$\sigma_{\beta 1}^2$	0.010	0.045	2.50E-04	0.042	0.006	0.020	2.28E-04	0.022	0.007	0.021	2.49E-04	0.027
$\sigma_{\beta 2}^2$	0.006	0.038	2.24E-04	0.021	0.004	0.013	2.23E-04	0.013	0.005	0.028	2.15E-04	0.013
$\sigma_{\beta 3}^2$	0.059	1.289	2.65E-04	0.182	0.055	1.665	2.51E-04	0.107	–	–	–	–
$\sigma_{\beta 4}^2$	0.030	0.154	2.63E-04	0.134	0.018	0.061	2.94E-04	0.082	–	–	–	–
$\sigma_{\beta 5}^2$	0.170	1.916	2.49E-04	0.159	–	–	–	–	–	–	–	–
$\sigma_{\beta 6}^2$	0.020	0.152	2.59E-04	0.065	0.021	0.105	2.62E-04	0.087	–	–	–	–
$\sigma_{\beta 7}^2$	0.001	0.003	1.66E-04	0.004	0.001	0.002	1.48E-04	0.002	7.00E-04	1.15E-03	1.35E-04	0.002
$\sigma_{\beta 8}^2$	–	–	–	–	–	–	–	–	7.07E-05	1.34E-04	1.34E-04	0.002

– indicates that this variable was not significant at the 90% confidence level.

Table 6
Parameter estimates of the REMV-Tobit model.

	Rear-end				Sideswipe				Angle			
Variable	Mean	SD	5.0%	95.0%	Mean	SD	5.0%	95.0%	Mean	SD	5.0%	95.0%
Intercept (β_0)	4.375	2.732	0.001	8.858	1.260	0.807	0.130	2.851	2.486	0.488	1.728	3.346
$\sigma_{\beta 0}{}^2$	0.279	0.418	0.033	0.930	0.151	0.242	0.028	0.481	0.149	0.155	0.031	0.403
Ln(MAADT) (β_1)	0.898	0.254	0.463	1.323	0.227	0.077	0.089	0.347	0.145	0.054	0.048	0.228
Ln(RAADT) (β_2)	−0.650	0.188	−0.960	−0.340	−0.281	0.066	−0.393	−0.171	−0.250	0.043	−0.322	−0.183
LaneBala (β_3)	−0.931	0.504	−1.704	−0.133	−0.296	0.315	−0.713	−0.037	–	–	–	–
Mainlane (β_4)	0.420	0.171	0.137	0.693	0.190	0.065	0.081	0.296	–	–	–	–
LengDec (β_5)	10.140	4.023	3.553	16.870	–	–	–	–	–	–	–	–
LengRam (β_6)	−2.140	0.683	−3.274	−1.013	−0.733	0.254	−1.150	−0.315	–	–	–	–
SpeedLimit (β_7)	−0.082	0.035	−0.143	−0.027	−0.014	0.012	−0.035	−0.005	−0.037	0.008	−0.050	−0.024
DiffSpeed (β_8)	–	–	–	–	–	–	–	–	0.014	0.006	0.004	0.023

– indicates that this variable was not significant at the 90% confidence level.

Table 7
Distributional effect of the random parameters across the observations.

	Rear-end		Sideswipe		Angle	
	Above zero	Below zero	Above zero	Below zero	Above zero	Below zero
Ln(MAADT) (β_1)	100.00%	0.00%	100.00%	0.00%	98.11%	1.89%
Ln(RAADT) (β_2)	0.00%	100.00%	0.00%	100.00%	0.04%	99.96%
LaneBala (β_3)	0.00%	100.00%	14.14%	85.86%	–	–
Mainlane (β_4)	99.87%	0.13%	94.65%	5.35%	–	–
LengDec (β_5)	100.00%	0.00%	–	–	–	–
LengRam (β_6)	0.00%	100.00%	0.00%	100.00%	–	–
SpeedLimit (β_7)	13.19%	86.81%	8.82%	91.18%	17.78%	82.22%
DiffSpeed (β_8)	–	–	–	–	93.52%	6.48%

Table 8
Transfer index of SPFs for different types of freeway diverge areas.

	Application data			
SPF	Type1	Type 2	Type 3	Type 4
Type 1	1	−1.314	−0.262	−0.570
Type 2	−0.611	1	−2.875	−0.598
Type 3	0.096	−1.273	1	−0.792
Type 4	−0.838	−1.335	−2.846	1

exception is the transfer index which reflects the SPF of type 3 being applied to type 1, but with a value closed to 0 (0.096). The transfer indices indicate that the SPFs for different types of freeway diverge areas are not directly transferable. As such, separate SPF for each type of freeway diverge area is recommended to be developed in safety diagnosis and safety evaluation.

6. Conclusions

This study evaluated the impact of various risk factors on crash rates of three collision types (i.e. rear-end, sideswipe, and angle) at freeway diverge areas by developing a RPMV-Tobit model. The model can accommodate jointly the correlation between crash rates across collision types and unobserved heterogeneity across observations, which are caused by the existence of unobserved risk factors that could jointly affect crash rates of different collision types. Data from 367 freeway diverge areas in the State of Florida, United States, were used to estimate the proposed models. Three candidate Tobit models, i.e. MV-Tobit model, REMV-Tobit model, and IU-Tobit model were also developed and compared with the RPMV-Tobit model. The models comparison indicates that the MV-Tobit model outperforms IU-Tobit models on fitting crash rates, as reflected by lower DIC and MAD values. The result suggests that accounting for correlation between crash rates across collision types is able to improve model fit. Moreover, the RPMV-Tobit model and REMV-Tobit model provide a better performance than the MV-Tobit model, indicating that accounting for the unobserved heterogeneous effect of risk factors can further improve model fit.

The impact of various risk factors on the crash rates across collision types were evaluated. The results showed that the impact of risk factors on different collision types were different. The lane-balanced design, number of lanes on mainline, and ramp length have significant impact on rear-end crash rate and sideswipe crash rates, but they have no impact on angle crash rates. The deceleration lane length has positive impact on rear-end crash rates, however, it is not related to sideswipe crash rates and angle crash rates. The difference in speed limit has a positive impact on angle crash rates, whereas it has no impact on rear-end crash rates and sideswipe crash rates. These findings demonstrate the varied impact of freeway geometric designs on crash rates of different collision types. As such, collision types need to be distinguished when developing crash rates models for freeway diverge areas. The model estimates showed that the mainline AADT has a positive impact on crash rates while the ramp AADT has a negative on crash rates, indicating the different mechanisms of the impact of traffic volume on crash rates at freeway diverge areas. The result showed that lane-balanced design, number of lanes on mainline, speed limit, and difference in speed limit present significant heterogeneous effect on crash rates across observations. The heterogeneity may be caused by unobserved road characteristics, traffic characteristics, environmental factors, driver behaviors as well as other omitted collision types across observations. Additionally, the models transferability analysis showed that the SPFs for different types of freeway diverge areas are not directly transferable, suggesting separate SPF should be developed for each diverge type.

The analysis suggests the considerable potential of the proposed model in crash rates analysis at freeway diverge areas. Although some findings regarding collision types were obtained, there are some limitations to this study. First, research efforts still need to be conducted with crash data collected from other states or areas to validate the results of this paper before the results can be used to direct the exit ramp designs. Second, this study did not give enough insight into occurrence mechanism for different crashes. Rear-end crashes could be related to abrupt deceleration at diverge bottlenecks, and sideswipe/angle crashes could be caused by intense lane changes. As such, high-resolution traffic data from loop detectors on freeway mainlines and ramps could be collected to develop the real time crash risk prediction models in future study. Third, the insignificant variables that excluded in the final models can introduce omitted variable bias, which should be considered by optimizing the statistical modeling frameworks. There are several areas of further research that can be investigated to improve the current study. First, due to the limited access to the crash severity, this paper did not incorporate injury-severity in the developed models. An extension of the current study can examine the injury-severity rates across collision types. This will provide new insights regarding the injury-severity mechanism across different collision types. Second, this study did not consider the temporal and spatial correlation in crash rates, further study could be conducted to extend the current model to accommodate the temporal and spatial effect. Moreover, further research efforts could also be made to compare the risk factors with crash frequency and that with crash rates. Last but not least, future study can further decompose the Tobit coefficients to the effect of the parameters on the overall rates, and on the probability of an observation being in the zero-state. They will help with the model parameters' inferences.

Acknowledgement

This research was sponsored by the National Natural Science Foundation of China (71871057, 71701046, 6521000176,6521000160), and the Fundamental Research Funds for the Central Universities (2242018R20003; 2242017K40130; YBJJ1533).

References

AASHTO, 2001. A Policy on Geometric Design of Highways and Streets. American Association of State Highway and Transportation Officials, Washington, D.C.
Alarifi, S.A., Abdel-Aty, M., Lee, J., 2018. A Bayesian multivariate hierarchical spatial joint model for predicting crash counts by crash type at intersections and segments along corridors. Accid. Anal. Prev. 119, 263–273.
Anastasopoulos, P.C., 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. Anal. Methods Accid. Res. 11, 17–32.
Anastasopoulos, P.C., Tarko, A.P., Mannering, F.L., 2008. Tobit analysis of vehicle accident rates on interstate highways. Accid. Anal. Prev. 40 (2), 768–775.
Anastasopoulos, P.C., Shankar, V.N., Haddock, J.E., Mannering, F.L., 2012a. A multivariate tobit analysis of highway accident-injury-severity rates. Accid. Anal. Prev. 45, 110–119.
Anastasopoulos, P.C., Mannering, F.L., Shankar, V.N., Haddock, J.E., 2012b. A study of factors affecting highway accident rates using the random-parameters tobit model. Accid. Anal. Prev. 45, 628–633.
Anderson, J., Hernandez, S., 2017. Heavy-vehicle crash rate analysis: comparison of heterogeneity methods using idaho crash data. Transp. Res. Rec. (2637), 56–66.
Bared, J., Giering, G.L., Warren, D.L., 1999. Safety evaluation of acceleration and deceleration lane lengths. Inst. Transp. Eng. ITE J. 69 (5), 50.
Bauer, K.M., Harwood, D.W., 1998. Statistical Models of Accidents on Interchange Ramps and Speed-change Lanes. Federal Highway Administration (FHWA)-RD-97-106, Washington, DC.
Bin Islam, M., Hernandez, S., 2016. Fatality rates for crashes involving heavy vehicles on highways: a random parameter tobit regression approach. J. Transp. Saf. Secur. 8 (3), 247–265.
Caliendo, C., De Guglielmo, M.L., Guida, M., 2016. Comparison and analysis of road tunnel traffic accident frequencies and rates using random-parameter models. J. Transp. Saf. Secur. 8 (2), 177–195.
Chen, H., Liu, P., Lu, J.J., Behzadi, B., 2009. Evaluating the safety impacts of the number and arrangement of lanes on freeway exit ramps. Accid. Anal. Prev. 41 (3), 543–551.
Chen, H., Zhou, H., Zhao, J., Hsu, P., 2011a. Safety performance evaluation of left-side off-ramps at freeway diverge areas. Accid. Anal. Prev. 43 (3), 605–612.
Chen, H., Zhang, Y., Wang, Z., Lu, J., 2011b. Identifying crash distributions and prone locations by lane groups at freeway diverging areas. Transp. Res. Rec. (2237), 88–97.
Chen, F., Ma, X., Chen, S., 2014. Refined-scale panel data crash rate analysis using random-effects tobit model. Accid. Anal. Prev. 73, 323–332.
Chen, F., Chen, S., Ma, X., 2018. Analysis of hourly crash likelihood using unbalanced panel data mixed logit model and real-time driving environmental big data. J. Safety Res. 65, 153–159.
Dong, C., Xie, K., Zeng, J., Li, X., 2018. Multivariate dynamic Tobit models with lagged observed dependent variables: an effectiveness analysis of highway safety laws. Accid. Anal. Prev. 113, 292–302.
El-Basyouny, K., Sayed, T., 2009a. Accident prediction models with random corridor parameters. Accid. Anal. Prev. 41 (5), 1118–1123.
El-Basyouny, K., Sayed, T., 2009b. Collision prediction models using multivariate Poisson-lognormal regression. Accid. Anal. Prev. 41 (4), 820–828.
Farid, A., Abdel-Aty, M., Lee, J., Eluru, N., Wang, J.H., 2016. Exploring the transferability of safety performance functions. Accid. Anal. Prev. 94, 143–152.
Farid, A., Abdel-Aty, M., Lee, J., 2018. Transferring and calibrating safety performance functions among multiple States. Accid. Anal. Prev. 117, 276–287.
Fountas, G., Anastasopoulos, P.C., 2017. A random thresholds random parameters hierarchical ordered probit analysis of highway accident injury-severities. Anal. Methods Accid. Res. 15, 1–16.
Fountas, G., Anastasopoulos, P.C., 2018. Analysis of accident injury-severity outcomes: the zero-inflated hierarchical ordered probit model with correlated disturbances. Anal. Methods Accid. Res. 20, 30–45.
Fountas, G., Anastasopoulos, P.C., Abdel-Aty, M., 2018a. Analysis of accident injury-severities using a correlated random parameters ordered probit approach with time variant covariates. Anal. Methods Accid. Res. 18, 57–68.
Fountas, G., Anastasopoulos, P.C., Mannering, F.L., 2018b. Analysis of vehicle accident-injury severities: a comparison of segment-versus accident-based latent class ordered probit models with class-probability functions. Anal. Methods Accid. Res. 18, 15–32.
Fountas, G., Sarwar, M.T., Anastasopoulos, P.C., Blatt, A., Majka, K., 2018c. Analysis of stationary and dynamic factors affecting highway accident occurrence: a dynamic correlated grouped random parameters binary logit approach. Accid. Anal. Prev. 113, 330–340.
Garcia, A., Romero, M.A., 2006. Experimental observation of vehicle evolution on deceleration lanes with different lengths. 85th Annual Meeting of the Transportation Research Board.
Golob, T.F., Recker, W.W., Alvarez, V.M., 2004. Safety aspects of freeway weaving sections. Transp. Res. Part A Policy Pract. 38 (1), 35–51.
Guo, Y., Li, Z., Wu, Y., Xu, C., 2018a. Exploring unobserved heterogeneity in bicyclists' red-light running behaviors at different crossing facilities. Accid. Anal. Prev. 115, 118–127.
Guo, Y., Osama, A., Sayed, T., 2018b. A cross-comparison of different techniques for modeling macro-level cyclist crashes. Accid. Anal. Prev. 113, 38–46.
Guo, Y., Li, Z., Wu, Y., Xu, C., 2018c. Evaluating factors affecting electric bike users' registration of license plate in China using Bayesian approach. Transp. Res. Part F Traffic Psychol. Behav. 59, 212–221.
Guo, Y., Li, Z., Liu, P., Wu, Y., 2019a. Exploring risk factors with crashes by collision type at freeway diverge areas: accounting for unobserved heterogeneity. IEEE Access 7, 11809–11819.
Guo, Y., Yang, B., Xu, C., 2019b. Analysis of Freeway Secondary Crashes With a Two-step Method by Loop Detector Data 7. IEEE Access, pp. 22884–22890.
Hadayeghi, A., Shalaby, A.S., Persaud, B.N., Cheung, C., 2006. Temporal transferability and updating of zonal level accident prediction models. Accid. Anal. Prev. 38 (3), 579–589.
Hossain, M., Abdel-Aty, M., Quddus, M.A., Muromachi, Y., Sadeek, S.N., 2019. Real-time crash prediction models: state-of-the-art, design pathways and ubiquitous requirements. Accid. Anal. Prev. 124, 66–84.
Lave, C., Elias, P., 1994. Did the 65 mph speed limit save lives? Accid. Anal. Prev. 26 (1), 49–62.
Lee, J., Yasmin, S., Eluru, N., Abdel-Aty, M., Cai, Q., 2018. Analysis of crash proportion by vehicle type at traffic analysis zone level: a mixed fractional split multinomial logit modeling approach with spatial effects. Accid. Anal. Prev. 111, 12–22.
Li, Z., Liu, P., Wang, W., Xu, C., 2012. Using support vector machine models for crash injury severity analysis. Accid. Anal. Prev. 45, 478–486.
Li, Z., Wang, W., Liu, P., Bai, L., Du, M., 2015. Analysis of crash risks by collision type at freeway diverge area using multivariate modeling technique. J. Transp. Eng. 141 (6), 04015002.
Liu, P., Chen, H., Lu, J.J., Cao, B., 2009. How lane arrangements on freeway mainlines and ramps affect safety of freeways with closely spaced entrance and exit ramps. J. Transp. Eng. 136 (7), 614–622.
Lord, D., Bonneson, J., 2005. Calibration of predictive models for estimating safety of ramp design configurations. Transp. Res. Rec. (1908), 88–95.
Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transp. Res. Part A Policy Pract. 44 (5), 291–305.
Lu, J., Geng, N., Chen, H., 2010. Impacts of freeway exit ramp configurations on traffic operations and safety. Adv. Transp. Stud.: Int. J. (22), 999–1009.
Ma, L., Yan, X., Weng, J., 2015a. Modeling traffic crash rates of road segments through a lognormal hurdle framework with flexible scale parameter. J. Adv. Transp. 49 (8), 928–940.
Ma, X., Chen, F., Chen, S., 2015b. Modeling crash rates for a mountainous highway by using refined-scale panel data. Transp. Res. Rec. (2515), 10–16.
Ma, Hao, W., Xiang, W., Yan, W., 2018. The impact of aggressive driving behavior on driver injury severity at highway-rail grade crossings accidents. J. Adv. Transp. 2018, 9841498.
Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: methodological frontier and future directions. Anal. Methods Accid. Res. 1, 1–22.
McCartt, A.T., Northrup, V.S., Retting, R.A., 2004. Types and characteristics of ramp-related motor vehicle crashes on urban interstate roadways in Northern Virginia. J. Safety Res. 35 (1), 107–114.
Moon, J.P., Hummer, J.E., 2009. Development of safety prediction models for influence areas of ramps in freeways. J. Transp. Saf. Secur. 1 (1), 1–17.
NHTSA, 2012. 2010 Motor Vehicle Crashes: Overview.

Papadimitriou, E., Theofilatos, A., 2017. Meta-analysis of crash-risk factors in freeway entrance and exit areas. J. Transp. Eng. Part A Syst. 143 (10), 04017050.

Sarwar, M.T., Anastasopoulos, P.C., 2017. The effect of long term non-invasive pavement deterioration on accident injury-severity rates: a seemingly unrelated and multivariate equations approach. Anal. Methods Accid. Res. 13, 1–15.

Sarwar, M.T., Fountas, G., Anastasopoulos, P.C., 2017. Simultaneous estimation of discrete outcome and continuous dependent variable equations: a bivariate random effects modeling approach with unrestricted instruments. Anal. Methods Accid. Res. 16, 23–34.

Sikder, S., Augustin, B., Pinjari, A.R., Eluru, N., 2014. Spatial transferability of tour-based time-of-Day choice models: empirical assessment. Transp. Res. Rec. 2429 (1), 99–109.

Song, Y., Noyce, D., 2019. Effects of transit signal priority on traffic safety: interrupted time series analysis of Portland, Oregon, implementations. Accid. Anal. Prev. 123, 291–302.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. Series B Stat. Methodol. 64 (4), 583–639.

Tobin, J., 1958. Estimation of relationships for limited dependent variables. Econ.: J. Econom. Soc. 26, 24–36.

Ulak, M.B., Ozguven, E.E., Vanli, O.A., Dulebenets, M.A., Spainhour, L., 2018. Multivariate random parameter Tobit modeling of crashes involving aging drivers, passengers, bicyclists, and pedestrians: spatiotemporal variations. Accid. Anal. Prev. 121, 1–13.

Wang, Z., Chen, H., Lu, J., 2009. Exploring impacts of factors contributing to injury severity at freeway diverge areas. Transp. Res. Rec. (2102), 43–52.

Wang, Z., Cao, B., Deng, W., Zhang, Z., Lu, J.J., Chen, H., 2011. Safety evaluation of truck-related crashes at freeway diverge areas. 90th TRB Annual Meeting Compendium of Papers DVD 23–27.

Washington, S.P., Karlaftis, M.G., Mannering, F., 2010. Statistical and Econometric Methods for Transportation Data Analysis. Chapman and Hall/CRC.

Wu, W.Q., Wang, W., Li, Z.B., Liu, P., Wang, Y., 2014. Application of generalized estimating equations for crash frequency modeling with temporal correlation. J. Zhejiang Univ. Sci. A 15 (7), 529–539.

Xie, K., Ozbay, K., Yang, H., 2019. A multivariate spatial approach to model crash counts by injury severity. Accid. Anal. Prev. 122, 189–198.

Xu, X., Wong, S.C., Choi, K., 2014. A two-stage bivariate Logistic-Tobit model for the safety analysis of signalized intersections. Anal. Methods Accid. Res. 3, 1–10.

Yang, Z.H.A.N.G., Zhibin, L.I., Pan, L.I.U., Liteng, Z.H.A., 2011. Exploring contributing factors to crash injury severity at freeway diverge areas using ordered probit model. Procedia Eng. 21, 178–185.

Yu, R., Xiong, Y., Abdel-Aty, M., 2015. A correlated random parameter approach to investigate the effects of weather conditions on crash risk for a mountainous freeway. Transp. Res. Part C Emerg. Technol. 50, 68–77.

Zeng, Q., Wen, H., Huang, H., Pei, X., Wong, S.C., 2017a. A multivariate random-parameters Tobit model for analyzing highway crash rates by injury severity. Accid. Anal. Prev. 99, 184–191.

Zeng, Q., Wen, H., Huang, H., Abdel-Aty, M., 2017b. A Bayesian spatial random parameters Tobit model for analyzing crash rates on roadway segments. Accid. Anal. Prev. 100, 37–43.

Zeng, Q., Wen, H., Huang, H., Pei, X., Wong, S.C., 2018. Incorporating temporal correlation into a multivariate random parameters Tobit model for modeling crash rate by injury severity. Transp. A Transp. Sci. 14 (3), 177–191.