Incorporating temporal correlation into a multivariate random parameters Tobit model for modeling crash rate by injury severity

Qiang Zeng, Ph.D., Post-doctoral Research,
School of Civil Engineering and Transportation,
South China University of Technology, Guangzhou, Guangdong, 510641 P.R. China.
Tel: 86 137 8778 3119
E-mail: zengqiang@scut.edu.cn

Huiying Wen, Ph.D., Professor,
School of Civil Engineering and Transportation,
South China University of Technology, Guangzhou, Guangdong, 510641 P.R. China.
Tel: 86 136 0006 6536
E-mail: hywen@scut.edu.cn

Helai Huang*, Ph.D., Professor,
Urban Transport Research Center, School of Traffic and Transportation Engineering,
Central South University, Changsha, Hunan, 410075 P.R. China.
Tel: 86 731 82656631
E-mail: huanghelai@csu.edu.cn

Xin Pei, Ph.D., Lecturer,
Department of Automation,
Tsinghua University, Beijing, P.R. China.
Tel: 861062795043
E-mail: peixin@mail.tsinghua.edu.cn

S.C. Wong, Ph.D., Chair Professor,
Department of Civil Engineering,
The University of Hong Kong, Pokfulam Road, Hong Kong.
Tel:852 28591964
E-mail: hhecwsc@hku.hk

* Corresponding author

This research was jointly supported by the Natural Science Foundation of China (Nos. 51378222, 51578247, 71371192 and 71301083), and a grant from the Joint Research Scheme of the National Natural Science Foundation of China/Research Grants Council of Hong Kong (Nos. 71561167001 and N_HKU707/15).

Incorporating temporal correlation into a multivariate random parameters Tobit model for modeling crash rate by injury severity

ABSTRACT

This study develops three temporal multivariate random parameters Tobit models to analyze crash rate by injury severity; these models simultaneously accommodate temporal correlation and unobserved heterogeneity across observations and correlations across injury severity. The three models are estimated and compared in the Bayesian context with a crash dataset collected from Hong Kong's Traffic Information System, which contains crash, road geometry, traffic, and environmental information on 194 directional road segments over a five-year period (2002 to 2006). Significant temporal effects are found in all of the temporal models, and the inclusion of temporal correlation considerably improves the goodness-of-fit of the multivariate random parameters Tobit regression, according to the results of deviance information criteria (DIC) and Bayesian R^2, indicating the strength of considering cross-period temporal correlation. Moreover, after accounting for temporal effects, the magnitude of the correlation between the crash rates at various injury degrees decreases, probably because a portion of the correlation may be attributed to unobserved or unobservable factors with time-dependent or autoregressive safety effects. Among the three candidate temporal models, the one with independent temporal effects has lower DIC and R^2 values, which suggests better model fit performance than the two with constant or correlated temporal effects. This finding supports the model with independent temporal effects as a good alternative for traffic safety analysis.

Keywords: Crash rate by severity; temporal correlation; random parameters; multivariate Tobit model.

1. Introduction

To better understand how the safety performance of a roadway or an intersection is affected by factors related to traffic, geometric design, the environment, and even regulations and laws (Dong et al., 2017; Wu et al., 2017), the application of novel approaches to model crash frequency (i.e., the crash count at certain road sites over a specified period) has long been a research focus in the field of traffic safety analysis. As crash frequencies are given as non-negative integers, most of the advocated approaches are statistical count models. Poisson regression is the basic model that assumes crash occurrence to be a Poisson process (Jovanis and Chang, 1986). To deal with various issues related to crash-frequency data (e.g., over-dispersion, under-dispersion, excess zero observations, multilevel structure, spatiotemporal correlation, and unobserved heterogeneity), numerous Poisson model variations have been proposed. Lord and Mannering (2010) and Mannering and Bhat (2014) presented detailed descriptions and assessments of these models.

In the past decade, much research effort has been devoted to the development of innovative methods to analyze crash rates (such as the number of crashes per 100 million vehicle miles traveled), which can be regarded as good alternatives to the traditional crash-frequency prediction models (Anastasopoulos et al., 2008). Compared with crash frequency, crash rates may be more appealing because they (1) are a standardized measure of the relative safety performance of a roadway site, which is more directly useable for road safety evaluation by traffic agencies (Anastasopoulos et al., 2008); (2) clearly reflect the risk of accident involvement and hence are more understandable to the public (Ma et al., 2015b); (3) may be more effective criteria for ranking sites in terms of safety improvement (Xu et al., 2014); and (4) are commonly used in crash reporting systems. For example, the National Highway Traffic Safety Administration uses fatality and injury rates per 100 million vehicle miles traveled to describe traffic safety in the United States (NHTSA, 2012).

Substantially different from crash frequencies, crash rates are continuous, non-negative numbers, and are usually left-censored at zero because no crashes may be observed at some sites during certain periods. Censoring refers to a limitation on data clustering that may result in a lower threshold (left-censored), an upper threshold (right-censored), or both. To handle the censoring issue, Anastasopoulos et al. (2008) first introduced the Tobit model to analyze crash rates, and studies have since focused on developing various forms of random parameters Tobit models to account for the heterogeneous effects of risk factors on crash rates (Anastasopoulos et al., 2012a; Caliendo el al., 2015; Ma et al., 2015a; Yu et al., 2015; Zeng et al., 2017b). Among these models, the random parameters Tobit model in refined temporal scale, advocated by Ma et al. (2015a), also demonstrates the significant serial correlation across observations in panel data.

To evaluate the safety performance of roadway sites more comprehensively, the relationship between the crash rate/frequency by injury severity and the observed risk factors (such as the traffic, geometric, and environmental characteristics of sites) has been further investigated (Zeng et al., 2016a). Models for the analysis of crash-injury-severity rates/frequencies may reveal deficiencies undetected by the overall-crash-rate/frequency models, by over-exposing crash severity (Ye et al., 2009, 2013). Hauer et al. (2004) suggested that using severity-weighted crash prediction models is the most cost-effective way to select sites for engineering improvement. Various multivariate count models have been developed to jointly analyze crash frequencies at different degrees of injury (Barua et al., 2014, 2016; Chiou and Fu, 2013, 2015; Chiou et al., 2014; Dong et al., 2016a; El-Basyouny and Sayed, 2009; Ma and Kockelman, 2006; Ma et al., 2008; Park and Lord, 2007; Song et al., 2006). However, only four studies (Anastasopoulos, 2016; Anastasopoulos et al., 2012b; Xu et al., 2014; Zeng et al., 2017a) have focused on joint modeling crash rate by injury severity. Their model estimation results all find that crash rates at various severity levels are significantly correlated. The significant correlation is attributable to common unobserved factors that affect crash rates across injury severity. The heterogeneous effects of certain risk factors are also found in the multivariate random parameters Tobit model proposed by Anastasopoulos (2016) and Zeng et al. (2017a), but none of these multivariate models account for the underlying temporal/serial correlation in the data of crash-injury-severity rates, which have been found in the prediction models for total crash rate (Ma et al., 2015a), total crash frequency (Castro et al., 2012; Dong et al., 2016b; Lord and Persaud, 2000; Noland et al., 2008; Quddus, 2008; Wang and Abdel-Aty, 2006), and crash frequency by severity (Chiou and Fu, 2015). Washington et al. (2011) pointed out that ignoring temporal correlation will lead to an underestimation the parameters' variances, and thus potentially result in the incorrect identification of the contributing factors, with significant consequences for safety.

The main objective of this study is thus to incorporate temporal correlation into a multivariate random parameters Tobit model to simultaneously analyze the crash rate by injury severity. The temporal correlation across observation periods will be accounted for, along with the correlation between crash rates at different severity levels and unobserved heterogeneity across observations. Three candidate temporal multivariate random parameters Tobit models are developed and compared in the Bayesian context, using the crash-injury-severity-rate data on 194 directional road segments over a five-year period (2002-2006) in Hong Kong.

The remainder of the paper is as follows. In the next section, the proposed models and criteria for model comparison are specified. The collected data for model demonstration are described in Section 3. The detailed estimation of the proposed models is introduced and the results of model comparison and parameter estimation are discussed in Section 4. Finally, concluding remarks and directions for future research are presented in Section 5.

2. Methods

In this section, the formulation of the base model, a multivariate random parameters Tobit regression, is first explicitly specified. Three forms of temporal models (uniform, correlated, and independent) are then proposed. Finally, two criteria, the deviance information criteria (DIC) and Bayesian R^2, are introduced for the purpose of model comparison in the context of Bayesian inference.

2.1. Model specification

2.1.1. Multivariate random parameters Tobit model

As stated, crashes may not be reported at some sites during the analysis period, so crash rates are generally left-censored at zero. Empirically, on a specific site, the crash rate at a more severe injury degree is more likely to be zero. The Tobit regression is an appropriate method for analyzing censored data as it can avoid biased and inconsistent parameter estimates (Anastasopoulos et al., 2008). To accommodate the possible correlation between crash rates at various severity levels and heterogeneity in the effects of risk factors, a multivariate random parameters Tobit model has been proposed (Zeng et al., 2017a). Using a left-censored threshold of zero, the multivariate random parameters Tobit regression for the joint modeling of the crash rate by injury severity is expressed as follows:

$$Y_{it}^{k*} = \beta_{it}^{k0} + \sum_{m=1}^{M} \beta_{it}^{km} x_{it}^{m} + \varepsilon_{it}^{k}, \tag{1}$$

$$Y_{it}^{k} = \begin{cases} Y_{it}^{k*}, & if\ Y_{it}^{k*} > 0 \\ 0, & if\ Y_{it}^{k*} \le 0 \end{cases}, \quad i = 1, 2, \cdots, N,\ t = 1, 2, \cdots, T,\ k = 1, 2, \cdots, K, \tag{2}$$

where Y_{it}^{k*} and Y_{it}^{k} represent the unobservable and observed crash rates at injury severity level k and site i during period t, respectively; N, T, and K are the number of the observed sites, periods, and categorized injury severity levels, respectively; and $x_{it}^{1}, x_{it}^{2}, \cdots x_{it}^{M}$ are the observed values of M risk factors at site i during period t. ε_{it}^{k} denotes the random error term, which is assumed to follow a multi-normal distribution with zero mean, that is,

$$\boldsymbol{\varepsilon}_{it} \sim N_K(\mathbf{0}, \boldsymbol{\Sigma}),\ \boldsymbol{\varepsilon}_{it} = \begin{pmatrix} \varepsilon_{it}^{1} \\ \varepsilon_{it}^{2} \\ \cdots \\ \varepsilon_{it}^{K} \end{pmatrix},\ \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1K} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2K} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{K1} & \sigma_{K2} & \cdots & \sigma_{KK} \end{pmatrix}, \tag{3}$$

in which σ_{kk} ($k = 1, 2, \cdots K$) is the variance of error term ε_{it}^{k} while σ_{k_1,k_2} $(k_1 \ne k_2)$ is the covariance between $\varepsilon_{it}^{k_1}$ and $\varepsilon_{it}^{k_2}$.

To account for the possible heterogeneity in the effects of risk factors and their correlation, the random parameters $\beta_{it}^{k0}, \beta_{it}^{k1}, \cdots \beta_{it}^{kM}$ are also assumed to be multi-normally distributed as:

$$\boldsymbol{\beta}_{it}^{m} \sim N_K(\boldsymbol{\beta}_m, \boldsymbol{\Sigma}_m), \quad \boldsymbol{\beta}_{it}^{m} = \begin{pmatrix} \beta_{it}^{1m} \\ \beta_{it}^{2m} \\ \cdots \\ \beta_{it}^{Km} \end{pmatrix}, \quad \boldsymbol{\Sigma}_m = \begin{pmatrix} \sigma_{11}^{m} & \sigma_{12}^{m} & \cdots & \sigma_{1K}^{m} \\ \sigma_{21}^{m} & \sigma_{22}^{m} & \cdots & \sigma_{2K}^{m} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{K1}^{m} & \sigma_{K2}^{m} & \cdots & \sigma_{KK}^{m} \end{pmatrix}, \quad m = 0, 1 \cdots, M, \tag{4}$$

where $\boldsymbol{\beta}_m$ and $\boldsymbol{\Sigma}_m$ are the mean vector and the variance-covariance matrix of $\boldsymbol{\beta}_{it}^{m}$, respectively. If the covariance of two random parameters is not statistically significant (say, at the 95% credibility level), they are assumed to follow independent normal distributions (Zeng et al., 2017a). If a random parameter's variance is statistically insignificant (say, at the 95% credibility level), it is simplified to be fixed across observations (Anastasopoulos et al., 2012a).

2.1.2. Temporal multivariate random parameters Tobit models

The temporal correlation may derive from time-dependent factors, which are unobserved or unobservable, and factors with time-dependent or autoregressive safety effects that are not explicitly specified in the model (Wang and Abdel-Aty, 2006). Adding a residual term δ_{it}^{k} with a lag-1 dependence into the link function Eq. (1) is a common method of reflecting the possible temporal correlation (Huang et al., 2009). To explore the best form of temporal correlation under the multivariate modeling framework, three temporal multivariate random parameters Tobit models are formulated in this study, based on the lag-1 dependence.

(1) Temporal model I

We first consider the case that $\delta_{it}^{1} = \delta_{it}^{2} = \cdots = \delta_{it}^{k} = \cdots = \delta_{it}^{K} = \delta_{it}$, $i = 1, 2, \cdots, N$, $t = 1, 2, \cdots, T$. Here, it is assumed that, for site i and period t, the temporal effects of all K injury degrees are equal. Accordingly, the link function of Temporal model I is:

$$Y_{it}^{k*} = \beta_{it}^{k0} + \sum_{m=1}^{M} \beta_{it}^{km} x_{it}^{m} + \varepsilon_{it}^{k} + \delta_{it}, \tag{5}$$

where the temporal terms are assumed to follow the normal distributions, which are based on the stationarity assumption (Huang et al., 2009),

$$\delta_{i,1} \sim normal\left(0, \frac{\sigma_{\delta}^{2}}{1-\gamma^{2}}\right), \tag{6}$$

$$\delta_{i,t} \sim normal\left(\gamma \delta_{i,t-1}, \sigma_{\delta}^{2}\right), \quad for\ t > 1. \tag{7}$$

In the above two equations, γ is the autocorrelation coefficient and σ_{δ} is the standard deviation of the temporal terms.

(2) Temporal model II

Here, we consider the situation when $\delta_{it}^{k} = v_k \delta_{it}$, $i = 1, 2, \cdots, N$, $t = 1, 2, \cdots, T$. The temporal effects of all K injury degrees for any given site and period are thus correlated, and the link function is expressed as follows:

$$Y_{it}^{k*} = \beta_{it}^{k0} + \sum_{m=1}^{M} \beta_{it}^{km} x_{it}^{m} + \varepsilon_{it}^{k} + v_k \delta_{it}, \tag{8}$$

where v_k is the scale parameter for injury severity level k.

(3) Temporal model III

In this subsection, the temporal effects, $\delta_{it}^{1}, \delta_{it}^{2}, \cdots \delta_{it}^{k}, \cdots \delta_{it}^{K}$, are assumed to be mutually independent, and Temporal model III is expressed as:

$$Y_{it}^{k*} = \beta_{it}^{k0} + \sum_{m=1}^{M} \beta_{it}^{km} x_{it}^{m} + \varepsilon_{it}^{k} + \delta_{it}^{k}, \tag{9}$$

$$\delta_{i,1}^{k} \sim normal\left(0, \frac{\sigma_{\delta_k}^{2}}{1-\gamma_k^{2}}\right), \tag{10}$$

$$\delta_{i,t}^{k} \sim normal\left(\gamma_k \delta_{i,t-1}^{k}, \sigma_{\delta_k}^{2}\right), \quad for\ t > 1. \tag{11}$$

where γ_k and σ_{δ_k} are the autocorrelation coefficient standard deviation for injury severity level k, respectively.

2.2. Model comparison

The DIC and Bayesian R^2, which have been extensively used for model comparison in research on Bayesian modeling (Huang et al., 2016a; Zeng and Huang, 2014; Wang et al., 2017), evaluate the above candidate models.

As a Bayesian generalization of Akaike's information criteria, the DIC penalizes larger-parameter models. Specifically, it provides a Bayesian measure of model complexity and fitting (Spiegelhalter et al., 2005) and is defined as:

$$DIC = \overline{D(\theta)} + pD, \tag{12}$$

where $\overline{D(\theta)}$ is the posterior mean deviance that can be taken as a Bayesian measure of fitting, and pD is a complexity measure for the effective number of parameters. In general, models with lower DIC values are preferable, and over 10 differences can rule out models with a higher DIC (Spiegelhalter et al., 2005).

The Bayesian R^2, a global model-fit measurement, is used to estimate the ratio of the explained sum of squares to the total sum of squares (Zeng and Huang, 2014). The Bayesian R^2 values of crash rates at each injury severity k and all observations, represented by R_k^2 and R_T^2 respectively, are calculated as (Zeng et al., 2017a):

$$R_k^2 = 1 - \frac{\sum_{t=1}^{T}\sum_{i=1}^{N}\left(Y_{it}^k - \lambda_{it}^k\right)^2}{\sum_{t=1}^{T}\sum_{i=1}^{N}\left(Y_{it}^k - \bar{Y}_k\right)^2}, \quad k = 1, 2, \cdots, K, \tag{13}$$

$$R_T^2 = 1 - \frac{\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(Y_{it}^k - \lambda_{it}^k\right)^2}{\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(Y_{it}^k - \bar{Y}\right)^2}, \tag{14}$$

$$\lambda_{it}^k = \begin{cases} \beta_{k0} + \sum_{m=1}^{M}\beta_{km}x_{it}^m, & if\ \beta_{k0} + \sum_{m=1}^{M}\beta_{km}x_{it}^m > 0 \\ 0, & if\ \beta_{k0} + \sum_{m=1}^{M}\beta_{km}x_{it}^m \le 0 \end{cases}, \tag{15}$$

$$\bar{Y}_k = \frac{1}{T \times N}\sum_{t=1}^{T}\sum_{i=1}^{N} Y_{it}^k, \quad k = 1, 2, \cdots, K, \tag{16}$$

$$\bar{Y} = \frac{1}{K \times T \times N}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{i=1}^{N} Y_{it}^k. \tag{17}$$

In the above equations, λ_{it}^k represents the expected crash rate at injury severity level k and site i during period t. $\bar{Y}_k$ and $\bar{Y}$ are the means of the crash rate at injury severity k and all observations, respectively.

3. Data preparation and preliminary analysis

To demonstrate the temporal correlation and to compare the proposed multivariate random parameters Tobit models, a crash dataset is collected from the Traffic Information System (TIS) maintained by the Transport Department of Hong Kong. This dataset contains 97 road segments (as shown in Fig. 1), whose traffic volumes are continuously measured by 97 core stations of the Hong Kong Annual Traffic Census (ATC) system. The directional average annual daily traffic (AADT) for each road segment adjacent to the 97 core stations ($N = 194$) from 2002 to 2006 ($T = 5$) is extracted from the ATC system for the analysis.

[Insert Fig. 1 here]

Crashes are mapped to these directional segments using geographical information system (GIS) techniques. The TIS categorizes crashes into three groups according to the severity of casualties' injuries; fatal, serious injury, and slight injury. As fatalities are rare, fatal is combined with serious injury to form the category of killed and seriously injured (KSI) crashes ($K = 2$). The road geometric, traffic, and environmental information is also obtained from the TIS system.

The yearly crash rate (number of crashes per million vehicle-kilometers traveled) by injury severity, CR_{it}^k, which is used as the dependent variable in this study, is calculated as:

$$CR_{it}^{k} = \frac{No\_crash_{it}^{k}}{AADT_{i}^{t} \times L_{i} \times 365/1000,000}, \quad (18)$$
$$i = 1, 2, \cdots 194,\ k = 1, 2,\ t = 2002, 2003, 2004, 2005, 2006$$

in which $No\_crash_{it}^{k}$ is the number of crashes at injury severity degree k that occurred on road segment i in year t; $AADT_{i}^{t}$ is the AADT on road segment i in year t; and L_{i} is the length of segment i. Among the total observations, 64 (6.6%) slight injury crash rates and 316 (32.6%) KSI crash rates are 0. Table 1 summarizes the definitions and descriptive statistics of the variables used in the model development. The results of the correlation tests and multi-collinearity diagnoses suggest that the correlation and collinearity among these factors are insignificant.

[Insert Table 1 here]

The lane-changing opportunity (LCO) variable refers to the length-weighted average number of eligible opportunities to change lanes in a sub-segment with identical lane markings (Zeng et al., 2016b). In road sections with double continuous lines, lane changing is not allowed (as shown in Fig. 2(a)); thus, $LCO = 0$. In sections with one continuous line and one broken line, lane changing is only allowed from the side of the broken line to the side of the continuous line (shown in Fig. 2(b)); thus, $LCO = 1$. In sections with a single broken line, lane changing is allowed between both adjacent lanes (as shown in Fig. 2(c)); thus, $LCO = 2$. Pei et al. (2012) provided a more detailed description of LCOs. On urban roadways, a certain proportion of traffic collisions occur when one of the vehicles is changing lanes. The defined LCO variable is expected to estimate the effect of lane-changing maneuvers (such as overtaking) on crash rates.

[Insert Fig. 2 here]

4. Model estimation and result analysis

4.1. Model estimation

With advances in computing methods, Bayesian inference has been gaining popularity as it can deal with very complex models, particularly those that do not have easily calculable likelihood functions (Lord and Mannering, 2010). Freeware WinBUGS, which is a popular platform for Bayesian inference, can be used to construct a flexible programming environment. Therefore, all the candidate models are programmed, estimated, and evaluated in WinBUGS.

As in previous studies (Barua et al., 2016; El-Basyouny and Sayed, 2009; Zeng et al., 2017a), non-informative priors are specified for the parameters and the hyper-parameters. A diffused normal distribution $N(0, 10^4)$ is used as the priors of the elements of $\boldsymbol{\beta}_m$ $(m = 0,1,\cdots 12)$, γ, v_k and γ_k $(k = 1,2)$. A diffused gamma distribution $gamma(0.001, 0.001)$ is used as the priors of the precisions of temporal effects, $1/\sigma_\delta^2$ and $1/\sigma_{\delta_k}^2$ $(k = 1,2)$. A Wishart prior $W(\mathbf{P}, r)$ is used for $\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\Sigma}_m^{-1}$, where $\mathbf{P} = \begin{bmatrix} 1, & 0 \\ 0, & 1 \end{bmatrix}$ represents the scale matrix and $r = 2$ is the degrees of freedom. For each model, a chain of 500,000 iterations of the Markov Chain Monte Carlo (MCMC) simulation are made, with the first 4,000 iterations acting as burn-ins. The Gelman-Rubin statistics available in WinBUGS are used to evaluate the MCMC convergence.

4.2. Model comparison

The results of the DIC, Bayesian R^2, and hyper-parameters for model comparison are shown in Table 2. Compared to the multivariate random parameters Tobit base model, the three Temporal models obviously have much lower DIC values and much higher values of R_1^2, R_2^2, and R_T^2, suggesting that the Temporal models fit the crash-rate data much better than the base model (Huang et al., 2009). This can be further confirmed by the autocorrelation coefficients and standard deviations of the temporal terms, which are in the main statistically significant at the 95% credibility level. The structured temporal effects may partially capture the unstructured random effects, so the random effects of slight injury crash rates (σ_{11}) and KSI crash rates (σ_{22}) in the Temporal models are all lower than their respective counterparts in the multivariate random parameters Tobit model (Washington et al., 2011). Interestingly, the correlation coefficient ρ $(= \sigma_{12} / \sqrt{\sigma_{11}\sigma_{22}})$ varies from significantly positive in the base model to significantly negative in Temporal models I and II and to insignificantly negative in Temporal model III. Its magnitude drops dramatically from 0.84 to 0.37, 0.43, and 0.03 after the incorporation of the temporal terms, possibly because the correlation between crash rates at the two severity degrees is mainly due to unobserved or unobservable factors with time-dependent or autoregressive safety effects, which are accounted for by the Temporal models, particularly Temporal model III.

[Insert Table 2 here]

The Bayesian R^2 measure results for Temporal models I and II are comparable, but the latter's DIC value is smaller, probably due to the substantial difference between the scale parameters, $v_1 (= 2.50)$ and $v_2 (= 5.58)$. Temporal model III has the lowest DIC and the highest values of R_1^2, R_2^2, and R_T^2, which indicate that it outperforms the other candidate models and that the temporal effects of the crash rates at the two injury-severity levels are independent. According to the estimates in Temporal model III, we can see that the difference between the temporal effects mainly lies in the variation of the temporal terms' standard deviations, $\sigma_{\delta_1}(=0.10)$ and $\sigma_{\delta_2}(=0.40)$.

4.3 Interpretation of parameter estimation

The multivariate random parameters Tobit model with independent temporal effect (Temporal model III) significantly outperforms the other three models, so only its risk factors' parameter estimates (shown in Table 3) are discussed in this section. According to the estimation results in Table 3, the random parameters' standard deviations of two factors (bus stop and lane width)[1] are significant at the 95% credibility level for both slight injury and KSI crash rates, which demonstrates that these two factors have significantly heterogeneous effects on the crash rates at the two injury-severity levels.

[Insert Table 3 here]

Specifically, the effects of BS (bus stop) on the slight injury and KSI crash rates are found to follow two independently normal distributions, with means of 0.296 and 0.759 and standard deviations of 0.472 and 0.685, respectively. Given these distributional parameters with their 95% credible intervals away from zero, the presence of bus stops on 73.4% and 86.6% of road segments increases the slight injury and KSI crash rates, respectively, which is attributable to the increased interaction between buses and other vehicles when they enter or leave bus bays (Pei et al., 2012; Zeng et al., 2016a); however, for the other 26.6% and 13.4% of road segments, the presence of a bus stop decreases the slight injury and KSI crash rates respectively, possibly due to the calming effect caused by the decreased speed when buses enter or leave their bays.

Width (lane width) results in random parameters that are normally distributed, with means of 0.224 and 0.392 and standard deviations of 0.092 and 0.231 for slight injury and KSI crash rates, respectively, which indicates that widening the lanes in the majority (99.2% and 95.4%, respectively) of roadway segments would increase the slight injury and KSI crash rates, whereas widening the lanes in the minority (0.8% and 4.6%, respectively) of road segments would have opposite effects. Gross and Jovanis (2007) also found a U-shaped relationship between lane width and crash risk.

With regard to the safety effects of other factors, which are constant across observations, the signs of each factor's coefficients for both slight injury and KSI

[1] The covariance of the two pair of random parameters is not significant at the 95% credibility level. Therefore, they are independently and normally distributed in Temporal model III.

crash rates are identical, which means that they have consistent effects on the crash rates at the two injury severity degrees. Nevertheless, some risk factors only have significant effects on slight injury or KSI crash rate.

The significant negative coefficients of AADT indicate that both the slight injury and KSI crash rates decrease with increasing daily traffic volume, which can be attributed to the reduced travel speeds caused by the greater traffic volume, thus decreasing the likelihood of a crash (Anastasopoulos et al., 2012a; Huang et al., 2016b; Zeng et al., 2017a). It is notable that the crash rates at both injury severity levels are lower on roadway segments with higher speed limits, which contradicts engineering intuition and the findings of many studies (Aguero-Valverde and Jovanis, 2008); however, some studies have suggested that roadway segments with high posted speed limits are usually high-grade highways with good planning, construction, management, and maintenance, thus promoting traffic safety (Milton and Mannering, 1998; Zeng et al., 2016a, 2016b, 2017a).

The estimates of the effects of Merge and Curvature on the KSI crash rate show that increasing the number of merging ramps or roadway curvature can significantly decrease the KSI crash rate. Similarly, Inter produces a significantly negative effect on the slight injury crash rate. The risk compensation theory may explain these findings; drivers may adapt to an adverse driving environment (more merging ramps or intersections, or a more curved roadway) by altering their driving behavior (such as being more careful or slowing down) (Mannering and Bhat, 2014; Zeng et al., 2017a). Some drivers may overcompensate for the adverse conditions, leading to a lower crash risk or a higher slight injury rate. Conversely, from the estimation results of the effect of Diverge, we find that an increase in the number of diverging ramps results in a higher slight injury crash rate. This conforms to engineering intuition and the findings of previous studies (Zeng et al., 2016a, 2017a), as the sites of approaches of diverging ramps are hazardous.

LCO (lane changing opportunity) is found to have a significantly positive effect on the KSI crash rate, which is consistent with many studies (Pei et al., 2012; Zeng et al., 2016a, b), as the increased vehicle interaction caused by lane changing maneuvers may raise the incidence of traffic conflicts.

5. Conclusions and future research

This study proposes three temporal multivariate random parameters Tobit models for the joint analysis of crash rate by injury severity by considering temporal correlation and unobserved heterogeneity across observations and correlation between crash rates at different severity levels. A crash dataset including crash, road geometric, traffic, and environmental information on 194 directional road segments for a five-year period (2002-2006) in Hong Kong is used to compare the candidate models. They are calibrated and evaluated in the Bayesian context via programming in the freeware WinBUGS.

The temporal effects, represented by (uniform, correlated, or independent) residual terms with a lag-1 dependence, are found statistically significant in the Temporal models. The results of DIC and Bayesian R^2 also show that the temporal

Tobit models have substantially better fit than the multivariate random parameters Tobit model, indicating that the consideration of temporal correlation across observations is reasonable. After accounting for the temporal correlation, the correlation between slight injury and KSI crash rates becomes smaller or even insignificant at the 95% credibility level. The multivariate random parameters Tobit regression with independent temporal effects (Temporal model III) is found to outperform the multivariate random parameters Tobit regressions with constant and correlated temporal effects (Temporal models I and II) in fitting the crash-rate data, which suggests that the temporal effects of slight injury and KSI crash rates are mutually independent. The results show that bus stop and lane width have heterogeneous effects on both slight injury and KSI crash rates.

In summary, empirical analysis demonstrates the superiority of the multivariate random parameters Tobit model with independent temporal effects and the significance of independent temporal correlations in crash-rate data, which indicates that the model should be proposed to analyze crash rate by severity over successive periods. However, due to the limitations of the collected dataset, only two levels of injury severity are considered in the empirical analysis. The models presented in the paper are essentially bivariate models. Nevertheless, the method can be applied for any number (≥2) of dependent variables. Field data with more (≥3) severity categories could be used to further compare the models' performance. Moreover, as with many other crash frequency/rate prediction models, the proposed model may suffer from underreporting, which would result in biased parameter estimates. Accounting for the effects of underreporting simultaneously would be merited. In addition, significant spatial correlation may exist across the crash rates of adjacent road sites (Zeng et al., 2017b), so it would also be beneficial for future research to incorporate spatial correlation into the multivariate random parameters Tobit model.

Acknowledgements

This research was jointly supported by the Natural Science Foundation of China (Nos. 51378222, 51578247, 71371192 and 71301083), and a grant from the Joint Research Scheme of the National Natural Science Foundation of China/Research Grants Council of Hong Kong (Nos. 71561167001 and N_HKU707/15).

References

Aguero-Valverde, J., and P. Jovanis. 2008. Analysis of road crash frequency with spatial models. *Transportation Research Record* 2061: 55-63.

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. *Analytic methods in accident research* 11: 17-32.

Anastasopoulos, P. C., F. L. Mannering, V. N. Shankar, and J. E. Haddock. 2012a. A study of factors affecting highway accident rates using the random-parameters tobit model. *Accident Analysis and Prevention* 45: 628-633.

Anastasopoulos, P. C., V. N. Shankar, J. E. Haddock, and F. L. Mannering. 2012b. A multivariate tobit analysis of highway accident-injury-severity rates. *Accident Analysis and Prevention* 45: 110-119.

Anastasopoulos, P. C., A. P. Tarko, and F. L. Mannering. 2008. Tobit analysis of vehicle accident rates on interstate highways. *Accident Analysis and Prevention* 40 (2): 768-775.

Barua, S., K. El-Basyouny, and M. T. Islam. 2014. A full Bayesian multivariate count data model of collision severity with spatial correlation. *Analytic Methods in Accident Research* 3: 28-43.

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

Caliendo, C., M. L. De Gugliemo, and M. Guida. 2015. Comparison and analysis of road tunnel traffic accident frequencies and rates using random-parameter models. *Journal of Transportation Safety and Security* 8 (2): 177-195.

Castro, M., R. Paleti, and C. R. Bhat. 2012. A latent variable representation of count data models to accommodate spatial and temporal dependence: Application to predicting crash frequency at intersections. *Transportation Research Part B: Methodological* 46: 253-272.

Chiou, Y. C., and C. Fu. 2013. Modeling crash frequency and severity using multinomial-generalized Poisson model with error components. *Accident Analysis and Prevention* 50: 73-82.

Chiou, Y. C., C. Fu, and H. Chih-Wei. 2014. Incorporating spatial dependence in simultaneously modeling crash frequency and severity. *Analytic Methods in Accident Research* 2: 1-11.

Chiou, Y. C., and C. Fu. 2015. Modeling crash frequency and severity with spatiotemporal dependence. *Analytic Methods in Accident Research* 5: 43-58.

Dong, C., D. B. Clarke, S. S. Nambisan, and B. Huang. 2016a. Analyzing injury crashes using random-parameter bivariate regression models. *Transportmetrica A: Transport Science* 12 (9), 794-810.

Dong, C., S. S. Nambisan, K. Xie, D. B. Clarke, and X. Yan. 2017. Analyzing the effectiveness of implemented highway safety laws for traffic safety across US states. *Transportmetrica A: Transport science* 13 (2): 91-107.

Dong, N., H. Huang, J. Lee, M. Gao, and M. Abdel-Aty. 2016b. Macroscopic hotspots identification: a Bayesian spatio-temporal interaction approach. *Accident Analysis and Prevention* 92: 256-264.

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

Gross, F., and P. P. Jovanis. 2007. Estimation of the safety effectiveness of lane and shoulder width: Case-control approach. *Journal of Transportation Engineering* 133 (6): 362-369.

Hauer E., B. K. Allery, J. Kononov, and M. S. Griffith. 2004. How best to rank sites with promise. *Transportation Research Record* 1897: 48-54.

Huang, H., H. Chin, and M. Haque. 2009. Empirical evaluation of alternative approaches in identifying crash hot spots: Naive ranking, empirical Bayes, and full Bayes methods. *Transportation Research Record* 2103: 32-41.

Huang, H., B. Song, P. Xu, Q. Zeng, J. Lee, and M. Abdel-Aty. 2016a. Macro and micro models for zonal crash prediction with application in hot zones identification. *Journal of Transport Geography* 54: 248-256.

Huang, H., Q. Zeng, X. Pei, S. C. Wong, and P. Xu. 2016b. Predicting crash frequency using an optimised radial basis function neural network model. *Transportmetrica A: Transport Science* 12 (4): 330-345.

Jovanis, P. P., and H. L. Chang. 1986. Modeling the relationship of accidents to miles traveled. *Transportation Research Record* 1068: 42-51.

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

Lord, D., and B. Persaud. 2000. Accident prediction models with and without trend: application of the generalized estimating equations procedure. *Transportation Research Record* 1717: 102-108.

Ma, J., and K. M. Kockelman. 2006. Bayesian multivariate Poisson regression for models of injury count, by severity. *Transportation Research Record* 1950: 24-34.

Ma, J., K. M. Kockelman, and P. Damien. 2008. A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. *Accident Analysis and Prevention* 40 (3): 964-975.

Ma, X., F. Chen, and S. Chen. 2015a. Modeling crash rates for a mountainous highway by using refined-scale panel data. *Transportation Research Record* 2515: 10-16.

Ma, L., X. Yan, and J. Weng. 2015b. Modeling traffic crash rates of road segments through a lognormal hurdle framework with flexible scale parameter. *Journal of Advanced Transportation* 49 (8): 928-940.

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

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

NHTSA, 2012. *2010 motor vehicle crashes: Overview*.

Noland, R. B., M. A. Quddus, and W. Y. Ochieng. 2008. The effect of the London congestion charge on road casualties: An intervention analysis. *Transportation* 35: 73-91.

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

Pei, X., S. C. Wong, and N. N. Sze. 2012. The roles of exposure and speed in road safety analysis. *Accident Analysis and Prevention* 48: 464-471.

Quddus, M. A. 2008. Time series count data models: an empirical application to traffic accidents. *Accident Analysis and Prevention* 40 (5): 1732-1741.

Song, J. J., M. Ghosh, S. Miaou, and B. Mallick. 2006. Bayesian multivariate spatial models for roadway traffic crash mapping. *Journal of Multivariate Analysis* 97:

246-273.
Spiegelhalter, D., A. Thomas, N. Best, and D. Lunn. 2005. *WinBUGS user manual*. MRC Biostatistics Unit, Cambridge, United Kingdom.
Wang, X., and M. Abdel-Aty. 2006. Temporal and spatial analyses of rear-end crashes at signalized intersections. *Accident Analysis and Prevention* 38 (6): 1137-1150.
Wang, J., H. Huang, and Q. Zeng. 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.
Washington, S. P., M. G. Karlaftis, and F. L. Mannering. 2011. *Statistical and econometric methods for transportation data analysis*, 2nd edition. CRC press, New York.
Wu, C. Y., and B. P., Loo. 2017. Changes in novice motorcyclist safety in Hong Kong after the probationary driving license scheme. *Transportmetrica A: Transport Science* 13 (5): 1-14.
Xu, X., S. C. Wong, and K. Choi. 2014. A two-stage bivariate logistic-Tobit model for the safety analysis of signalized intersections. *Analytic Methods in Accident Research* 3-4: 1-10.
Ye, X., R. M. Pendyala, S. P. Washington, K. Konduri, and J. Oh. 2009. A simultaneous equations model of crash frequency by collision type for rural intersections. *Safety Science* 47 (3): 443-452.
Yu, R., Y. Xiong, and M. Abdel-Aty. 2015. A correlated random parameter approach to investigate the effects of weather conditions on crash risk for a mountainous freeway. *Transportation Research Part C: Emerging Technologies* 50: 68-77.
Zeng, Q., and H. Huang. 2014. Bayesian spatial joint modeling of traffic crashes on an urban road network. *Accident Analysis and Prevention* 67: 105-112.
Zeng, Q., H. Huang, X. Pei, and S. C. Wong. 2016a. Modeling nonlinear relationship between crash frequency by severity and contributing factors by neural networks. *Analytic Methods in Accident Research* 10: 12-25.
Zeng, Q., H. Huang, X. Pei, S. C. Wong, and M. Gao. 2016b. Rule extraction from an optimized neural network for traffic crash frequency modeling. *Accident Analysis and Prevention* 97: 87-95.
Zeng, Q., H. Wen, H. Huang, X. Pei, and S. C. Wong. 2017a. A multivariate random-parameters Tobit model for analyzing highway crash rates by injury severity. *Accident Analysis and Prevention* 99: 184-191.
Zeng, Q., H. Wen, H. Huang, and M. Abdel-Aty. 2017b. A Bayesian spatial random parameters Tobit model for analyzing crash rates on roadway segments. Accident *Analysis & Prevention* 100: 37-43.

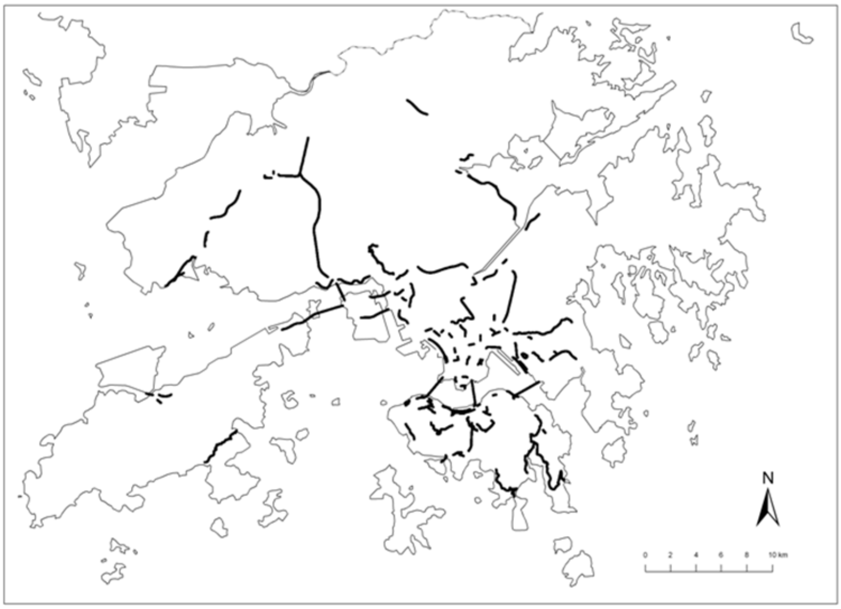

Fig. 1. Selected roadway segments in Hong Kong for the analysis.

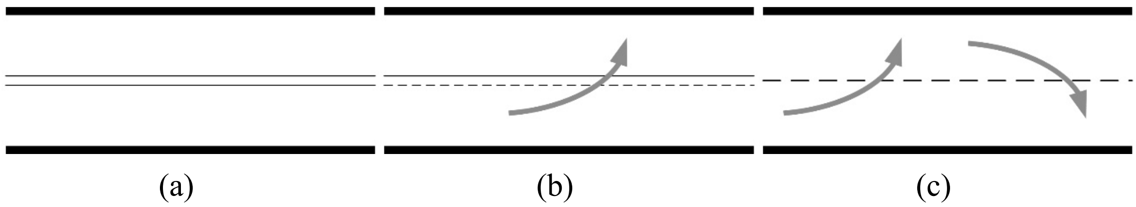

Fig. 2. Lane changing opportunities for different road section configurations.

Table 1. Descriptive statistics of the variables.

Variable	Description	Mean	SD	Min.	Max.
Response variable					
Slight	Slightly injured crash count per million vehicle-kilometers traveled	1.72	2.24	0	24.35
KSI	Killed and seriously injured crash count per million vehicle-kilometers traveled	0.48	1.00	0	9.86
Risk factors					
AADT	Average annual daily traffic (10^3 vehicles)	22.51	20.04	1.16	101.63
Width	Average width of each lane (m)	3.62	0.62	2.60	7.30
SL	Posted speed limit (km/h)	60.62	14.85	50	110
Merge	Number of merging ramps	0.85	1.01	0	4
Diverge	Number of diverging ramps	1.75	2.31	0	17
Inter	Number of intersections	1.87	2.38	0	16
Gradient	Average segment gradient (10^{-2})	0.04	2.67	-9	9
Curvature	Average segment curvature	22.2	17.6	0	85
LCO	Lane changing opportunity	2.43	1.63	0	7.85
Median	Presence of median barrier: yes = 1, no = 0	0.70	0.46	0	1
BS	Presence of bus stop: yes = 1, no = 0	0.63	0.48	0	1
Rainfall	Annual precipitation (m)	2.28	0.56	1.24	3.22

Table 2. Model comparison results.

	Base model	Temporal model I	Temporal model II	Temporal model III
DIC	4573	3348	3171	2951
R_1^2	0.771	0.923	0.917	0.932
R_2^2	0.766	0.951	0.951	0.971
R_T^2	0.771	0.945	0.945	0.963
σ_{11}	**0.26(0.18, 0.35)**[a]	**0.10(0.06, 0.14)**	**0.10(0.06, 0.15)**	**0.08(0.05, 0.13)**
$\sigma_{21}(=\sigma_{12})$	**047(0.33, 0.62)**	**-0.06(-0.11, -0.01)**	**-0.07(-0.12, -0.02)**	-0.01(-0.04, 0.03)
σ_{22}	**1.30(0.81, 1.91)**	**0.28(0.11, 0.51)**	**0.27(0.12, 0.52)**	**0.17(0.07, 0.34)**
ρ [b]	**0.82(0.73, 0.90)**	**-0.37(-0.61, -0.08)**	**-0.43(-0.64, -0.18)**	-0.06(-0.35, 0.25)
γ	—	**0.99(0.97, 1.00)**	**0.98(0.97, 1.00)**	—
σ_δ	—	**0.11(0.05, 0.09)**	0.05(0.03, 0.09)	—
v_1	—	—	**2.50(1.59, 3.66)**	—
v_2	—	—	**5.58(3.61, 7.30)**	—
γ_1	—	—	—	**0.99(0.97, 1.00)**
σ_{δ_1}	—	—	—	0.10(0.04, 0.18)
γ_2	—	—	—	**0.96(0.92, 0.99)**
σ_{δ_2}	—	—	—	**0.40(0.22, 0.57)**

[a] Estimated mean (95% Bayesian credible interval) for the parameter. Boldface indicates statistical significance at the 95% credibility level.

[b] $\rho = \sigma_{12} / \sqrt{\sigma_{11}\sigma_{22}}$.

Table 3. Parameter estimation in the multivariate random-parameters Tobit model with independent temporal effect (Temporal model III)[a].

Variable	Slight injury			Killed and serious injury		
	Mean	95% credible interval		Mean	95% credible interval	
		2.5%	97.5%		2.5%	97.5%
Constant	0.438	-0.074	1.135	**2.895**[b]	**1.673**	**4.128**
AADT	**-0.010**	**-0.018**	**-0.001**	**-0.025**	**-0.041**	**-0.010**
SL	**-0.011**	**-0.019**	**-0.002**	**-0.037**	**-0.055**	**-0.017**
Merge	-0.084	-0.200	0.019	**-0.246**	**-0.490**	**-0.030**
Diverge	**0.068**	**0.004**	**0.126**	0.005	-0.114	0.125
Inter	**-0.090**	**-0.153**	**-0.028**	-0.045	-0.173	0.083
BS	**0.296**	**0.066**	**0.539**	**0.759**	**0.171**	**1.215**
SD of BS	**0.472**	**0.388**	**0.551**	**0.685**	**0.466**	**0.873**
Curvature	-0.004	-0.013	0.004	**-0.018**	**-0.030**	**-0.005**
LCO	0.053	-0.019	0.121	**0.230**	**0.076**	**0.393**
Width	**0.224**	**0.079**	**0.356**	**0.392**	**0.158**	**0.617**
SD of Width	**0.092**	**0.067**	**0.115**	**0.231**	**0.183**	**0.272**

[a] Median, Gradient, and Rainfall are excluded, as none of their effects on crash rates at the two severity degrees is significant at the 95% credibility level.

[b] Boldface indicates statistical significance at the 95% credibility level.