IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 11, 2020, accepted November 21, 2020, date of publication November 24, 2020, date of current version December 9, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3040211

Applying a Correlated Random Parameters Negative Binomial Lindley Model to Examine Crash Frequency Along Highway Tunnels in China

FENG TANG, XINSHA FU, MINGMAO CAI, YUE LU, SHIYU ZHONG, AND CHONGZHEN LU
School of Civil Engineering and Transportation, South China University of Technology, Guangzhou 510640, China
Corresponding author: Xinsha Fu (fuxinsha_scut@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 51978283 and Grant 51778242.

ABSTRACT Highway tunnels have a higher risk of crashing than open roads, which require a systematic approach to tunnel safety. However, previous research had the following problems: 1) Studies have largely focused on open roads, with very little research on tunnels. 2) The collected crash contributing factors involve narrow ranges, with very little tunnel crash data including both tunnel design features, traffic conditions and pavement conditions. 3) None of the studies considered both excess zero observations and unobserved heterogeneity with its interactions. To address these issues, this paper first established an appropriate tunnel dataset containing 3 to 5 years of crash data from several highways in China and the influence factors of tunnel design features, traffic conditions and pavement conditions. A correlated random parameters negative binomial Lindley (CRPNB-L) model that considers both excess zero observations and unobserved heterogeneity with its interaction effects was then proposed. Compared to the uncorrelated random parameters negative binomial Lindley (URPNB-L) model, fixed parameters negative binomial Lindley (FPNB-L) model and fixed parameters negative binomial (FPNB) model, the CRPNB-L model solves the deviation that arises from excess zero observations by introducing the Lindley distribution and considers the unobserved heterogeneity with its interactions by introducing correlated random parameters. In the comparisons, the CRPNB-L model achieves the best effects in the goodness-of-fit. Furthermore, the estimated results of the CRPNB-L model showed that segment length, traffic volume, proportion of class 5 vehicle (heavy trucks and trailers), tunnel entrance and exit segments, and steep uphill and downhill segments were associated with higher crash frequency, while curvature, tunnel length, pavement damage condition index (PCI) and skid resistance index (SRI) were associated with lower crash frequency. In addition, the random variables of the curvature, the steep downgrade indicator, the proportion of class 5 vehicle and SRI were identified and their intercorrelations were analyzed.

INDEX TERMS Tunnel safety, correlated random parameters, negative binomial Lindley model, tunnel design features, traffic conditions, pavement conditions.

I. INTRODUCTION

Over the past decade, highway safety has received increasing attention from transportation authorities and researchers around the world [1], [2]. As special structures of highways, tunnels have the characteristics of rapidly changing lighting at the entrance and exit, limited cross-section width and a closed field of vision, which make the driving environment more complicated and require more alertness of drivers. Therefore, tunnel segments are frequently prone to crashes, and their safety problems are particularly important [3]–[6]. Taking China as an example, by the end of 2019, the total mileage of tunnels in China reached 17236 kilometers and the crash frequency of tunnel segments was 1.44 times that of open segments [7], according to the Transportation Industry Development Statistics Bulletin (TIDSB). Therefore, it is very important to seek appropriate crash frequency models for highway tunnels and analyze factors that influence crashes for the design and management of these tunnels. The occurrence of a crash is a complex event involving the interactions of

The associate editor coordinating the review of this manuscript and approving it for publication was Rashid Mehmood.

drivers, vehicles, roads, and environments [1], [6]. In research related to crash models, the main problems to be overcomed are selecting appropriate influencing factors affecting crashes and considering the attributes of excess zero observations, unobserved heterogeneity and its interactions that exist in the crash panel data [8]–[10].

A. CRASH CONTRIBUTING FACTORS

A review of previous studies found that the indicators related to road design characters [11]–[22], traffic conditions [11]–[15] and pavement conditions [23]–[25] have significant impacts on traffic safety. First, the impacts of traffic conditions on crashes was reviewed. Studies [11]–[14] showed that the crash frequency increases with increases in traffic volume and the proportion of heavy trucks. Studies [11], [13]–[15] have found that speed limits also significantly affect crash frequency, where papers [11], [13] proved that higher speed limits are associated with a higher crash frequency, whereas other studies [14], [15] indicated that setting higher speed limits can reduce crash frequency. In terms of the impacts of road design characteristics on the crash frequency, some scholars found that the geometric shape of a highway cross-section is related to the crash frequency. For example, studies [11], [12], [16] verified that highway sections with narrow shoulders and a small offset of longitudinal separation columns have a higher crash frequency. The safety effects of adding lanes are debatable. Research [11], [16] asserted that adding lanes increases the chances of changing lanes, which leads to frequent crashes, while other research [17]–[19] believed that adding lanes is beneficial to safety because adding lanes not only reduces the traffic density but also reduces the potential dangerous interactions between vehicles under high-flow conditions. In addition, the impacts of horizontal and vertical alignment on safety are also crucial. Research [16], [20], [21] found that an increase in the longitudinal slope would sharply increase the crash frequency. However, the relationship between curves and safety remains a topic of discussion. Studies [11], [12], [15], [16] believed that curve segments increase the operating difficulty of drivers and may lead to a higher crash frequency. Other studies [13], [20], [22] showed that drivers pay more attention and drive at a comfortable speed in most cases when driving on a curved segments, which reduce the crash frequency of these segments. With regard to the impacts of pavement on crashes, relevant studies [23]–[25] deeply discussed this type of topics by establishing Safety Performance Functions (SPFs) between pavement conditions and crash frequency, which reached the following conclusions: (1) Good pavement conditions can reduce the crash frequency, so regular maintenance is an effective way to improve the safety performance of roads. (2) Pavement rutting had great influences on the crash frequency, especially when more heavy trucks are present. (3) Improvement in the friction coefficient can significantly reduce the crash frequency under wet pavement conditions. According to the above fundings, the selection of influencing factors of crash frequency has great impacts on estimation results, and even the opposite results may occur due to improper selection of influencing factors. Therefore, it is important to establish a appropriate database of crash influencing factors.

B. EXCESS ZERO OBSERVATIONS OF CRASH DATA

As for the crash modelling techniques, the negative binomial (NB) and its variant models [26]–[31], which can not only satisfy the nonnegative integer and randomness of crash frequency but also adapt to the attribute of overdiscreteness (variance greater than mean) of crashes, are the mainstream methods. Road crashes are, however, rare events, especially in highway tunnels, where excess zero observations result in a deviation between the actual distribution of crash frequency and the negative binomial distribution [32]. Therefore, the studies [33]–[35] employ the zero-inflated negative binomial (ZINB) model, where special attention is paid to the impacts of zero observations on crashes, as an alternative method. This kind of model first uses Logit or Probit models to divide the zero observations into an absolute safety state and a nonabsolute safety state, and then fits the zero observations of the nonabsolute safety state and nonzero observations by NB models, which effectively eliminates the counts of zero observations. However, the ZINB models have been criticized as an effort that seeks to maximize the statistical fit rather than as an explanation of the crash occurrence process [36]. To allow road segments to change in time under the two states of absolute safety and nonabsolute safety, Malyshkina *et al.* [37] and Malyshkina and Mannering [38] proposed two-state Markov switching count models whereby roadway segments are allowed to switch states over time across two unobserved but significantly different unsafe states. The specification of this model, however, is complex, and thus its application to a large dataset is computationally heavy [39]. Modeling efforts to deal with crash data with excess zeros have continued with the introduction of new distributions that are capable of handling observations with small counts and combining them with the parent distributions capturing the crash data generating process (NB distributions). These kinds of models mainly include the negative binomial Lindley (NB-L) model [40], [41], the negative binomial crack (NB-CR) model [42], the negative binomial generalized index (NB-GE) model [43], etc. Moreover, these studies also verify that the improvements of negative binomial distributions are more appropriate for actual crash frequency distributions. Concretely, Lord and Geedipally [41] compared the performance of Poisson, NB and NB-L models using two crash datasets containing zero observations of 89% and 90%. The results showed that the NB-L model has a better goodness-of-fit and prediction accuracy. Furthermore, Geedipally *et al.* [40] modeled crash data with zero observations of 36% in Indiana and zero observations of 70% in Michigan, which verified that the NB-L model has better performance than the NB model and the ZINB model. Vangala *et al.* [43] conducted a further analysis using the same data as that of Lord and Geedipally [41] and found that the performance of the NB-GE model was comparable with the NB-L model and significantly outperformed the traditional NB model. It should be emphasized that the above literatures, which attempt to address the excess zero observations of crash panel data, are all aimed at open roads. However, to the best of our knowledge, research to solve the problem of excess zero observations for crash data of highway tunnels has not occurred. Given the prior knowledge provided by the above research, it is reasonable to establish an appropriate improved model based on the NB-L model to fit the relationship between crash frequency and influence factors in view of the characteristics of excess zero observations of highway tunnel.

C. UNOBSERVED HETEROGENEITY AND ITS INTERACTION EFFECTS

In addition to excess zeros, another important challenge in crash modeling is to account for unobserved heterogeneity and its interaction effects resulting from other sources. Ideally, if we can obtain all the factors that cause crashes, the established models can show the most reasonable relationships between the influencing factors and crashes and achieve the best goodness-of-fit [44], [45]. However, based on the limitations of data collection methods, the data we collected represent only a part of all the influencing factors (road design characters, traffic conditions, pavement conditions, etc.). The influences of unobserved factors on observed variables or crashes vary in multiple dimensions (i.e., space, time, group and individual dimensions), which is called unobserved heterogeneity. For example, considering the impacts of the lighting indicator on the crash frequency of highway tunnels (assuming that the value of the variable is 1 if lighting exists; otherwise, it is 0), due to the different types of terrain, lighting conditions and degrees of fatigue of drivers in different highway tunnels, the variable of the lighting indicator has different influences on crash frequency, even if the values of lighting indicators are the same. Therefore, if the estimated parameters of the lighting indicator are limited to fixed values in all segments, biased estimation will inevitably occur [12], [45]. Ignoring unobserved heterogeneity and limiting the effects of observed variables to be the same in all segments, namely, fixed parameter (FP) models [15], [46], will usually lead to biased estimation, which in turn leads to incorrect inference and prediction. To address this limitation, studies [17], [44], [47] adopted random effect negative binomial (RENB) models to analyze crash frequency and verified better goodness-of-fit was achieved by the RENB model than that of a fixed parameter negative binomial (FPNB) model. The RENB models assume that the combined effects of unobserved variables obey a certain distribution (usually a normal distribution) regarding the intercept terms [17]. The constraint conditions assumed by an RENB are, however, too harsh, where the randomness of the intercept terms alone is not sufficient to explain the unobserved heterogeneity. Based on this situation, studies [48]–[50] proposed an uncorrelated random parameter negative binomial (URPNB) model as an alternative method, which does not treat the intercept term as the only random component but allows the regression parameters of each variable to change randomly in multiple dimensions (i.e., space, time, group and individual dimensions). The results also showed that the URPNB model achieved a better goodness-of-fit compared with RENB model.

It should be emphasized that the URPNB models assume that the distributions of random parameters are independent, which cannot capture the potential interaction impacts among random parameters. According to Conway and Kniesner [51], ignoring the correlation between random parameters may lead to biased estimation. Pavement damage conditions and the percentage of trucks, for example, have significant effects on tunnel crash frequency [26], but trucks are more likely to be involved in crashes when driving in segments with severe pavement damage. In other words, an interaction between the proportion of trucks and the pavement damage conditions jointly influences the crash frequency. Consequently, the assumption of the URPNB model that these two variables are independent does not reflect the actual situation, which may lead to biased estimation. Aiming to solve the problems existing in the URPNB model, some researchers tried to further capture the interactions of unobserved heterogeneity. Yu *et al.* [52] analyzed the influences of weather conditions on highway crashes by using a correlated random parameter model and found that a Tobit model with correlated random parameters was statistically superior to the corresponding uncorrelated random parameter model. Emine *et al.* [53] found that the goodness-of-fit of a correlated random parameter negative binomial (CRPNB) model was better than that of a URPNB model in analyzing crash frequency in multiple cities. Therefore, it is reasonable to believe that the application of correlated random-parameters methods to capture the unobserved heterogeneity and its interactions in expressway tunnels safety analysis will achieve a breakthrough effects.

D. SAFETY ANALYSIS OF TUNNELS

The studies about tunnel safety can divide into two categories. The first type of research divided the tunnel into different zones according to the driving environment and light conditions, which is aimed to deeply analyze the statistical characteristics of crashes among different zones and the qualitative relationship between crashes and some influence factors. For example, studies [58]–[61] divided a tunnel into three or four zones and investigated the crash rates of different tunnel zones to evaluate the tunnel features that could mainly affect safety. Literature [62] have investigated the characteristics of crashes of freeway tunnel groups in China, by adopting five-zone approach for safety analysis of tunnel groups, While the study [63] employed a seven-zone analytic approach for the safety investigation of 18 expressway tunnels with length ranging from 2 to 3 km. It should be pointed out that these studies focus on the characteristics of the temporal, spatial, and modality distributions of tunnel crashes, rather than quantitatively analyzing the influences of factors on crashes. The second type of research focuses on establishing Safety Performance Functions (SPFs) to understand the effects of crash contributing factors. Literature [5] applied, for example, the bivariate negative binomial model to fit the non-severe crash frequency and severe crash frequency, and used random effects binomial regression mode to analyze the year effect of severe crashes. Literature [64] provided an analysis of severe crashes (fatal and injury accidents only) that occurred in 260 Italian road tunnels on the basis of random-parameters regression models. Furthermore, research [54], [65] consider the interactions of unobserved heterogeneity, the research [54] used a random effects negative binomial model (RENB), an uncorrelated random parameters negative binomial model (URPNB), and a correlated random parameters negative binomial model (CRPNB) to fit crash frequency of freeway tunnels in China, which showed that the CRPNB model provided better goodness-of-fit and offered more insights into the factors that contribute to tunnel safety. Similarly, The research [65] provided an analysis of crash frequency, which occurred in 226 unidirectional motorway tunnels over a four-year monitoring period in Italy, based on the unrelated and correlated random-parameter Poisson models. The results still showed that the correlated random-parameter Poisson model obtained the optimal goodness-of-fit. The second type of research can effectively reveal the relationship between crash contribution factors and crash variables, which is also the purpose of this paper. However, the research related to this method (1) have not adequately collected the dadaset of crash contributing factors, where most of the tunnel crash contributing factors established in literature [5], [64], [65] are related to tunnel design features and the impacts of pavement conditions on tunnel crash are rarely reported, (2) have not considered both unobserved heterogeneity and excess zero observations, where studies [5], [64], [65] were all modeling only for unobserved heterogeneity.

E. OBJECTIVE AND SCOPE

By reviewing the above research, we summarize the issues that need to be addressed in the crash frequency modeling techniques of highway tunnels. (1) Studies have lacked relatively appropriate databases on the influencing factors of tunnel crashes. The occurrence of crashes involves various factors, such as drivers, vehicles, roads and the environment. Starting with unilateral factors to establish crash frequency models will lead to biased estimation and incorrect judgments. (2) Although some studies have proposed improved models to address the characteristics of excess zero observations existed in crash datasets, to our best knowledge, such crash modeling techniques have yet to address highway tunnels. (3) Few studies have used correlated random parameter models, which take into account the unobserved heterogeneity and its interaction effects, in the safety analysis of highway tunnels, especially for Chinese expressway tunnels.

In conclusion, this paper promoted three contributions for tunnel safety analysis (the pipeline of the proposed methods is shown in Fig. 1.): (1) We established an appropriate and accurate dataset for tunnel safety analysis by collecting 3–5 years of tunnel crash data from three typical highways in Guangdong Province, China (Lianzhou to Huaiji on the G55 highway, hereinafter referred to as G55-1; Huaiji to Sanshui on the G55 highway, hereinafter referred to as G55-2; and Maan to Hekou on the G80 highway, hereinafter referred to as G80), and the influencing factors involving tunnel design tunnel design features, traffic conditions and pavement conditions, with a total of 11 variables. (2) A tunnel-based correlated random parameters negative binomial Lindley (CRPNP-L) model, which simultaneously takes into account the excess zero observations by introducing the Lindley distribution and unobserved heterogeneity with its interactions by introducing correlated random parameters, was developed. Compared with the established uncorrelated random parameters negative binomial Lindley (URPNB-L) model, the fixed random parameters negative binomial Lindley (FPNB-L) model and the fixed random parameters negative binomial (FPNB) model, the Bayesian estimation results confirmed that the CRPNP-L model had a better goodness-of-fit based on the evaluation indexes of the Deviation information criterion (DIC), the Mean absolute deviance (MAD), and the Mean squared prediction error (MSPE). (3) Based on the established CRPNP-L model, an in-depth analysis of the interaction of variables and their influence principles on crash frequency was conducted, which provides new technical support for the design and management of highway tunnels in China.

II. DATA DESCRIPTION

In this study, a dataset related to multiyear crashes and the influencing factors of 84 one-way tunnels with a total length of 93.87 km, which are located on typical mountain highways G55-1, G55-2 and G80 in southern China, were collected (the routes of the three highways and the schematic of a tunnel are shown in Fig. 2). The three highways are all managed by the Guangdong Provincial Department of Communications (GPDC), where the data of the G55-1 highway were collected from 2012 to 2017, the data of the G55-2 highway were collected from 2014 to 2017, and the data of the G80 highway were collected from 2015 to 2018. In addition, these data involve a total of 545 crashes and a total of 11 influencing factors, including tunnel design features, traffic conditions and pavement conditions. The crash data were provided by the Guangdong Provincial Highway Administration (GDFA). The data on traffic condition indicators were collected from toll stations on each highway, which included the annual average daily traffic volume (AADT) and the proportion of class 5 vehicle. The data on the tunnel design features, which contain 7 variables, were provided by Jilin Transportation Construction Co., Ltd., and Guangdong Changda Highway Engineering Co., Ltd. Pavement condition indexes were provided at intervals of 20 m or 50 m by the Guangdong Transportation Group Testing Center (GTGTC), which included the pavement damage condition index (PCI) and skid resistance index (SRI). After the selection of variables,

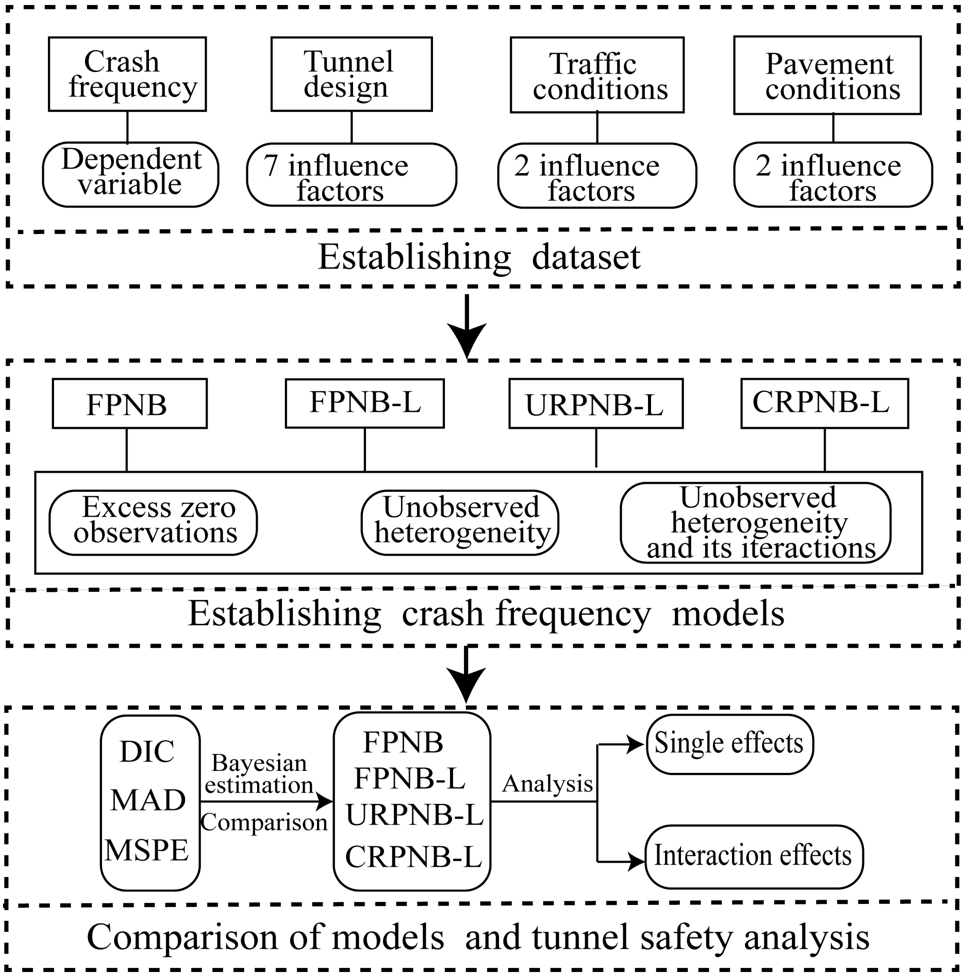


FIGURE 1. The pipeline of the proposed method. The top of the figure shows the process of creating a dataset that includes crashes and three types of explanatory variables. The middle of the figure presents the CRPNB-L model proposed in this paper and the contrasting URPNB-L, FPNB-L and FPNB models. The bottom of the figure shows the evaluation and safety analysis methods. The FPNB, FPNB-L, URPNB-L and CRPNB-L models represent the fixed parameters negative binomial model, the fixed parameters negative binomial Lindley model, the uncorrelated random parameters negative binomial Lindley model and the correlated random parameters negative binomial Lindley model, respectively. The DIC, MAD and MSPE represent the goodness-of-fit indexes of the Deviation information criterion, Mean absolute deviance and Mean square prediction error, respectively.

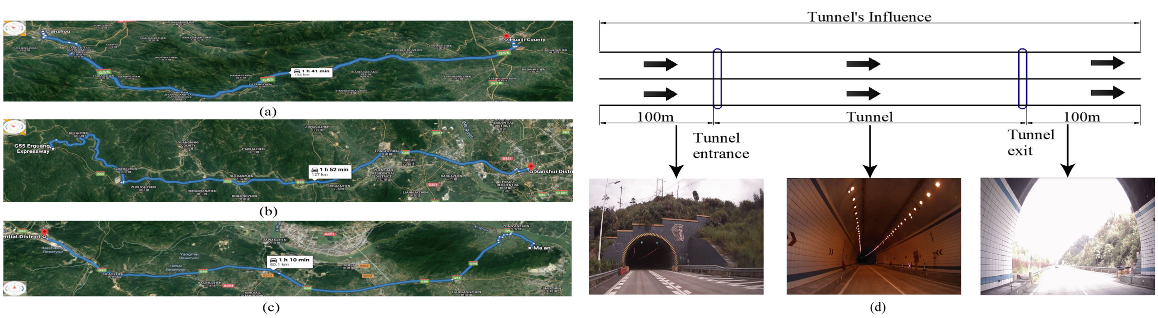


FIGURE 2. The routes of the three highways where the studied tunnels are located and the schematic of a tunnel. (a) represents the route of the G55-1 highway. (b) represents the route of the G55-2 highway. (c) represents the route of the G80 highway. (d) represents the schematic diagram of a tunnel, in which also present the actual pictures at the entrance, inside and exit of the tunnel, respectively.

the division of tunnel segments is a key step in modeling techniques, which usually includes fixed-length methods and homogeneous methods. Relevant studies [54], [55] showed that the fixed-length methods have defects in the average treatment of indicators that significantly affect the model's performance. Therefore, this paper used the homogeneous

TABLE 1. Descriptive statistics of the variables.

Variables	Continuous variables				Discrete variables	
	Mean	Std. Dev.	Min	Max	Count	Percentage
DEPENDENT VARIABLE						
Crash counts	0.96	1.39	0	12	–	–
EXPOSURE VARIABLES						
Segment length (km)	0.607	0.579	0.106	4.811	–	–
AADT (veh/day)	7107	4143	3753	17809	–	–
TUNNEL DESIGN FEATURES						
Length of tunnel						
0, Shorter than 500 m	–	–	–	–	207	26.437%
1, Longer than 500 m	–	–	–	–	576	73.563%
Tunnel entrance indicator						
0, Including tunnel entrance	–	–	–	–	432	55.172%
1, Do not including tunnel entrance	–	–	–	–	351	44.828%
Tunnel exit indicator						
0, Including tunnel exit	–	–	–	–	483	61.686%
1, Do not including tunnel exit	–	–	–	–	345	44.061%
Steep upgrade indicator						
0, Less than 2% of largest grade	–	–	–	–	648	82.759%
1, Greater than 2% of largest grade	–	–	–	–	135	17.241%
Steep downgrade indicator						
0, Greater than -2% of largest grade	–	–	–	–	660	84.291%
1, Less than -2% of largest grade	–	–	–	–	123	15.709%
Curvature (1/km)	0.299	0.317	0	1.163	–	–
TRAFFIC CONDITIONS						
Proportion of class 5 vehicle (%)	12.325	4.426	5.279	17.44	–	–
PAVEMENT CONDITIONS						
Pavement damage condition index (PCI)	90.36	7.04	61.795	100	–	–
Skid resistance index (SRI)	86.136	7.706	63.75	98.8	–	–

method to classify segments on the basis of curvature and longitudinal slope indexes, where the tunnels were divided into 783 observations with consistent alignment. The length distribution is shown in Fig. 3, where the lengths of the tunnel segments are concentrated in the interval [300 m, 1000 m]. The shortest length of a segment is only 106 m, and the maximum length reaches 4811 m.

On this basis, the statistical values of the variables in each tunnel segment were sorted in terms of years, and a dataset was established for tunnel crash frequency models (as shown in Table 1). Notably, since the sampling frequency of the pavement conditions is not aligned with the segment length, the mean values of the pavement condition indicators are used as a proxy. From Table 1, the following explanations are required: (1) The AADT is calculated for highway sections between adjacent toll stations according to the method provided by [54]. (2) The literature [54], [55] provides clues that it is reasonable to design segment length and AADT as exposure variables, which participate in the parameter estimation process by taking their logarithmic forms in the models. (3) The vehicles are categorized into five classes for the Guangdong Freeway Network Toll System (GFNTS)

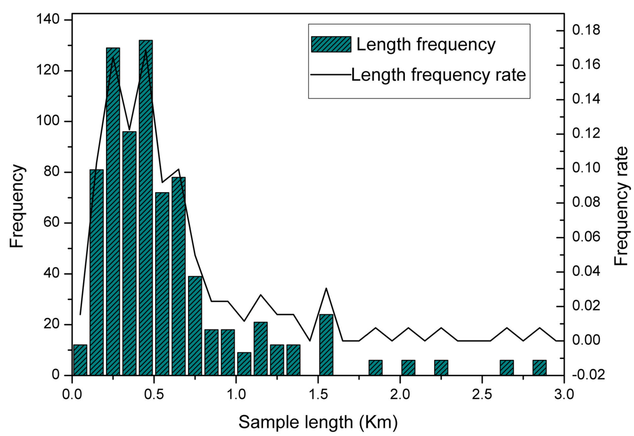


FIGURE 3. Length distribution of homogeneous segments.

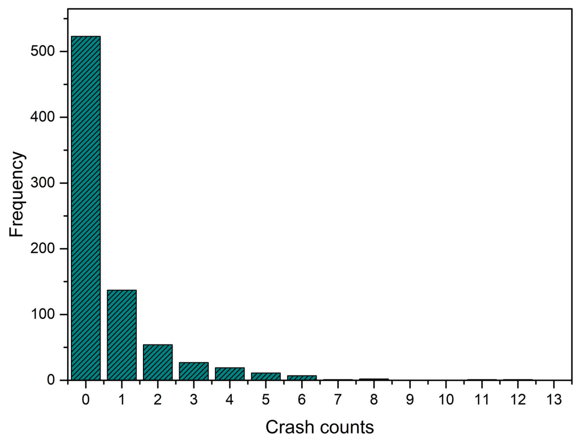


FIGURE 4. The distribution of crash frequency.

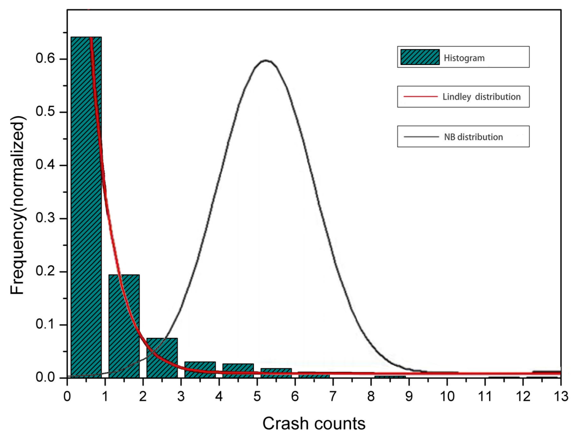


FIGURE 5. Resemblance of Lindley distribution to the frequency distribution of observed tunnel crash counts.

according to the height, number of axles, number of wheels and wheelbase. The class 5 vehicle are represented by heavy trucks, heavy trailers and 40-foot container trucks. (4) According to the Chinese technical standards for road engineering (JTG B01-2014), a tunnel with a length greater than or equal to 500 m is defined as a medium-long tunnel; otherwise, the tunnel is defined as a short tunnel. (5) A tunnel's influence zone starts at the tunnel exit/entrance points and is 100 m long [54]. (6) Literature [54] defines tunnel segments with a longitudinal slope degree of more than 2% as steep segments. Research [55], [56] found that the crash frequency of tunnel segments with a longitudinal slope degree of more than 2% is significantly higher than that of other tunnel segments. Therefore, whether the slope is greater than 2% is used as one of the influencing factors of crash frequency in this paper.

Furthermore, the Variance Inflation Factor (VIF) is used to test whether multicollinearity exists among variables (as shown in Table 2). Note that according to the clues provided in the literature [56], a VIF value of a variable greater than 5 indicates the multicollinearity exists between this variable and other variables, and such a variable should be eliminated. Table 2 clearly shows that the VIF values of the

TABLE 2. VIF values of the explanatory variables.

Variables	VIF	Variables	VIF
Log (Segment length)	1.81	Steep downgrade indicator	1.57
Log (AADT)	1.41	Curvature	2.32
Length of tunnel	1.8	Proportion of class 5 vehicle	1.33
Tunnel entrance indicator	1.59	PCI	1.40
Tunnel exit indicator	1.65	SRI	1.68
Steep upgrade indicator	2.7		

variables selected in this paper are all less than 5, so these influencing factors used for modeling are reasonable.

In addition, we drew a distribution diagram of the crash frequency (as shown in Fig. 4) and a figure that shows the probability density of the Lindley distribution against NB distributions overlaid with the histogram of tunnel crash frequency data used in this study (as shown in Fig. 5). As shown in Fig. 4, the crash counts were basically concentrated between 0 and 4, and the segments of 0 crash count accounts for 64.879% of the total segments, which verified that the dataset in this paper has the characteristics of excess zero observations. It can be seen from Fig. 5 that the density of the Lindley distribution around zero fits the histogram of tunnel crashes quite well, and thus a combination of NB and Lindley distributions can shift the sole NB distribution to the left and can capture excess zeros more effectively.

III. METHODOLOGY

A. DESCRIPTION OF THE FPNB AND FPNB-L MODELS

In this paper, an FPNB model was used to match the nonnegative integer and over-discreteness (variance greater than the mean) attributes of the crash frequency in highway tunnels. Specifically, the probability of tunnel segment i experiencing n_i crashes within a specified period of time (1 year in this paper) is calculated by the following equation.

$$P(n_i) = \left[\frac{1/\alpha}{1/\alpha + \lambda_i}\right]^{1/\alpha} \frac{\Gamma[1/\alpha + n_i]}{\Gamma(1/\alpha)n_i!} \left[\frac{\lambda_i}{1/\alpha + \lambda_i}\right]^{n_i} \tag{1}$$

where α is a dispersion parameter used to illustrate the relationship between the mean and variance of the crash frequency. When the value of α is 0, that is, the mean and variance of the crash frequency are equal, the NB model degenerates into a Poisson model. $\Gamma(.)$ is the Gamma distribution function, and λ_i represents the mean number of crashes for segment i, which is generally designated as an exponential function of independent variables and logarithmic forms of exposure variables. The calculation formula is as follows:

$$\lambda_i = \exp(\beta X_i + \varepsilon_i) \tag{2}$$

where β is a vector of the estimable parameters, and X_i is a vector of the independent variables and exposure variables for segment i. It is important to note that the exposure variables of segment length and AADT participate in the regression parameter estimation process by taking their logarithmic forms in the models (independent variables and logarithmic forms of exposure variables are referred to as explanatory variables below). $\exp(\varepsilon_i)$ is a gamma-distributed error term with a mean of 1 and variance α.

The appealing characteristic of the FPNB-L model stems from the property of its core distribution, the Lindley distribution, whose mean is close to zero and has a long tail for observations, which is extremely similar to the actual distribution of the crash frequency. The FPNB-L is a combination of NB and Lindley distributions, which can also be expressed as a hierarchical representation of negative binomial, Bernoulli and gamma distributions [40]. For the FPNB-L model, the probability of tunnel segment i experiencing n_i crashes within a specified period of time (1 year in this paper) is calculated by the following formula:

$$P(n_i) = \left[\frac{1/\alpha}{1/\alpha + \theta_i}\right]^{1/\alpha} \frac{\Gamma[1/\alpha + n_i]}{\Gamma(1/\alpha) n_i!} \left[\frac{\theta_i}{1/\alpha + \theta_i}\right]^{n_i} \tag{3}$$

where θ_i is a Lindley parameter, which is calculated as follows:

$$\begin{aligned} \theta_i &= \lambda_i \varphi_i, \\ \varphi_i &\sim Gamma(1 + \gamma, \chi) \\ \gamma &\sim Bernoulli(1/(1 + \chi)) \end{aligned} \tag{4}$$

B. DESCRIPTION OF THE RPNB-L MODEL

The parameters β of equation (2) are fixed in all tunnel segments, which does not deeply reveal the unobserved heterogeneity of crash frequency. In an actual situation, due to the existence of unobserved heterogeneity, the influence degree of explanatory variables on the crash frequency varies in different segments. Therefore, the unobserved heterogeneity through the random parameters in the RPNB model is allowable by introducing a random component δ_i to the parameter β:

$$\beta_i = \beta + \delta_i \tag{5}$$

where β_i is a regression parameter vector of explanatory variables for segment i, and δ_i is a random variable with a deterministic probability density function, which is subject to a normal distribution with a mean of 0 and variance of σ.

A regression parameter is defined as random if the estimated standard deviation of the parameter is significantly different from zero; otherwise, it is defined as a fixed parameter. Therefore, the probability density function for the RPNB-L model is expressed as equation (6), where $g(.)$ is the probability density function of δ_i

$$\begin{aligned} P(n_i \mid \theta_i, \alpha, \sigma) = \int_{\delta_i} \left[\frac{1/\alpha}{1/\alpha + \theta_i}\right]^{1/\alpha} \frac{\Gamma[1/\alpha + n_i]}{\Gamma(1/\alpha) n_i!} \\ \times \left[\frac{\theta_i}{1/\alpha + \theta_i}\right]^{n_i} g(\delta_i) d(\delta_i) \end{aligned} \tag{6}$$

C. DESCRIPTION OF THE URPNB-L AND CRPNB-L MODELS

There is no correlation among the random parameters in an RPNB-L model, namely, a URPNB-L model, which means that the random parameters are independent. In reality, there are likely correlations among the parameters due to the possible interactive effects of the explanatory variables on the dependent variable. To explore the potential correlation effects among the random parameters, β_i was assumed to follow a multivariate normal distribution in this study:

$$\beta_i = \beta + W\zeta_i \tag{7}$$

$$W = \begin{bmatrix} (\sigma_1)^2 & \sigma_{1,2} & \cdots & \sigma_{1,j-1} & \sigma_{1,j} \\ \sigma_{2,1} & (\sigma_2)^2 & \cdots & \sigma_{2,j-1} & \sigma_{2,j} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{j-1,1} & \sigma_{j-1,2} & \cdots & (\sigma_{j-1})^2 & \sigma_{j-1,j} \\ \sigma_{j,1} & \sigma_{j,2} & \cdots & \sigma_{j,j-1} & (\sigma_j)^2 \end{bmatrix} \tag{8}$$

where W is the variance-covariance matrix of the multivariate normal distribution, which is a lower triangular matrix that engenders the correlation among the elements of the random parameter vector β_i. j is the number of random parameters, and ζ_i is the randomly and independently distributed uncorrelated vector term.

As previously mentioned, the URPNB-L model assumes that the off-diagonal elements in the variance-covariance matrix W are zero (i.e., no correlation exists among the random parameters), which to some extent seems to be a rather strong restriction. In this study, a CRPNB-L model was developed by using an unrestricted variance-covariance matrix to investigate the safety and the heterogeneous effects of various factors as well as the possible interactive effects of correlated variables.

D. EVALUATION INDEXES OF MODEL PERFORMANCE

Three common standards, Deviation information criterion (DIC), Mean absolute deviance (MAD) and Mean square prediction error (MSPE) were used in this study to measure the goodness-of-fit of the models. According to the definition in the literature [10], the calculation formula of DIC is as:

$$DIC = \bar{D} + PD \tag{9}$$

in which $\bar{D}$ is the posterior mean deviance for assessing the model fit, and PD is the effective number of parameters in the model for measuring model complexity. Generally, a model with a lower DIC value is preferred. DIC differences between 5 and 10 are deemed substantial, while differences of more than 10 indicate the significant outperformance of the model with a lower DIC [10].

MAD and MSPE reflect the deviation degree between the crash frequency fitted by the model and the actual crash frequency. Lower values of MAD and MSPE represent better goodness-of-fit. According to the clues in the literature [10], MAD and MSPE are defined by the following equations, respectively.

$$MAD = \frac{1}{N}\sum_{i=1}^{i=N}|K_i' - K_i| \tag{10}$$

$$MSPE = \sqrt{\sum_{i=1}^{i=N}(K_i' - K_i)^2/N} \tag{11}$$

where N is the number of all tunnel segments, K_i' represents the crash counts fitted by the models for segment i, and K_i represents the crash counts observed for segment i.

In addition, to understand the influences of observable factors on crash frequency, the Incidence Rate Ratio (IRR), which represents the change scope of the expected crash frequency when an independent variable x_m is increased by one unit [56], is calculated as follows.

$$IRR_m = \frac{E(n_i|\overline{X}_i, x_m + 1)}{E(n_i|\overline{X}_i, x_m)} = \exp(\beta_m) \tag{12}$$

where $\overline{X}_i$ are the independent variables other than x_m, and β_m is the regression coefficient of x_m. For the exposure variables, the IRR values represent the change of expected crash frequency when an exposure variable x_e increases by 1%, whose equation is as follows:

$$IRR_e = \frac{E(n_i|\overline{X}_i, 1.01x_e)}{E(n_i|\overline{X}_i, x_e)} = 1.01^{x_e} \tag{13}$$

where β_e is the regression coefficient of x_e.

E. MODEL ESTIMATION

With the continuous progress of crash modeling technology, the Bayesian method has become the mainstream parameter estimation method in the field of road safety statistical analysis [9], [10], [56], [57]. The Bayesian method does not need the closed likelihood function, which can solve the parameter estimation problem of the complex model. WinBUGS is one of the most commonly software approaches for Bayesian inference, which adopts Markov chain Monte Carlo (MCMC) simulation and the Metropolit-Hastings algorithm to infer the posterior distribution of parameters based on the prior distribution and observation data. Therefore, this paper uses WinBUGS software to implement the Bayesian estimation process of all models.

Without sufficient prior knowledge, noninformative priors are specified for the parameters and hyperparameters. Specifically, we use a diffused normal distribution, Normal(0, 10^4), for the priors of the regression coefficients (i.e., the elements of β and β_i) [10]. A diffused gamma distribution, Gamma(0.001, 0.001), is used to obtain the priors of the precision parameters $1/\sigma$ and $1/\alpha$ [9], [10]. The prior distribution of χ is set as Bernoulli $(1/(1 + e^{-1})$. A Wishart prior, W (P, r), is used for parameter W, where P is the identity matrix with 4 rows and 4 columns, and $r = 4$ is the degrees of freedom [10]. For each model, a chain of 130000 iterations of the MCMC simulation is constructed, and the first 20000 iterations are excluded as a burn-in. The ratios of the Monte Carlo Errors are monitored relative to the standard deviations of the estimates, and the MCMC trace plots for the model parameters are visually inspected to assess the MCMC convergence.

IV. RESULTS AND DISCUSSION

This section includes the comparison of the goodness-of-fit of each model, the interpretation of estimated regression parameters, and the in-depth analysis of the impacts of significant variables and interaction effects of random variables on the crash frequency based on the CRPNB-L model.

A. COMPARISON OF GOODNESS-OF-FIT

The superparameters and goodness-of-fit indexes estimated by candidate models in this paper are shown in Table 3.

In order to test the advantages of introducing the Llindley distribution, the performance of the FPNB model and FPNB-L model is compared. It can be seen from Table 3 that the DIC, MAD and MSPE values of the FPNB-L model are 1185, 0.33 and 0.3, respectively, which are all lower than that of the FPNB model with DIC, MAD and MSPE values of 1212, 0.35 and 0.31, respectively. These results are similar to those in paper [40], [41], which demonstrated that the FPNB-L model is more adaptable to the tunnel crash dataset of excess zero observations and thus achieves a better goodness-of-fit. In addition, the Lindley parameter θ of the FPNB-L model is significantly 1.108 at the 95% Bayesian credibility level, indicating that it is reasonable to use the NB-Lindley distribution to fit the tunnel crash data in this study. Furthermore, the dispersion parameter α of the FPNB-L model with a value of 3.16 is much larger than the that of FPNB model with a value of 1.914, which also reflected that the FPNB-L model considers more zero-value segments, leading to the increase of discrete parameters (the relative magnitude of the mean and variance of crashes).

Second, we discuss the effects of unobserved heterogeneity and its interactions by comparing the performance of FPNB-L, URPNB-L and CRPNB-L models. Similarly, it can be seen from Table 3 that the DIC value of the URPNB-L model (a value of 1163) is far lower than that of FPNB-L model (a value of 1185). Although the MAD value of the URPNB-L model is the same as that of the FPNB-L model with a value of 0.33, the MSPE value of the URPNB-L model (a value of 0.28) is lower than that of the FPNB-L model (a value of 0.3). In general, the URPNB-L model has a better goodness-of-fit. This result is consistent with previous research [55], [57], which explained that randomizing the regression parameters is helpful to capture the unobserved heterogeneity, reducing occurrences of an incorrect definition and improving the goodness-of-fit. When adding the CRPNB-L model to the discussion, we found that the CRPNB-L model has the best goodness-of-fit among the candidate models due to its lowest DIC value of 1128, MAD value of 0.3 and MSPE value of 0.27. Such findings are expected because the CRPNB-L model captures not only

TABLE 3. Superparameters and goodness-of-fit indexes of candidate models in this paper.

Evaluation indexes	FPNB		FPNB-L		URPNB-L		CRPNB-L	
	Estimates	95% BCI	Estimates	95% BCI	Estimates	95% BCI	Estimates	95% BCI
Dispersion parameter α	1.914	[1.550,2.278]	3.16	[2.654,3.666]	4.083	[3.226,4.940]	3.699	[2.996,4.402]
Lindley parameter θ	–	–	1.108	[0.909,1.296]	0.843	[0.717,0.961]	0.743	[0.646,0.840]
$\bar{D}$	1202		1172		1137		1086	
PD	10		13		26		42	
DIC	1212		1185		1163		1128	
MAD	0.35		0.33		0.33		0.3	
MSPE	0.31		0.3		0.28		0.27	

95% BCI represents the 95% Bayesian credibility interval.

TABLE 4. Estimation results of regression coefficients in the candidate models.

Variables	FPNB		FPNB-L		URPNB-L		CRPNB-L	
	Estimates	95% BCI	Estimates	95% BCI	Estimates	95% BCI	Estimates	95% BCI
Constant	-6.542	[-7.654,-5.364]	-8.274	[-9.515,-7.033]	-9.864	[-11.541,-8.088]	-9.056	[-10.414,-7.698]
Log (Segment length)	1.033	[0.795,1.291]	1.112	[0.901,1.334]	1.075	[0.892,1.269]	1.194	[0.967,1.433]
Log (AADT)	1.374	[1.154,1.580]	1.425	[1.140,1.739]	1.684	[1.381,2.004]	1.564	[1.251,1.908]
Length of tunnel	–	–	-0.156	[-0.190,-0.122]	-0.188	[-0.231,-0.144]	-0.173	[-0.211,-0.135]
Tunnel entrance indicator	0.159	[0.122,0.194]	0.209	[0.178,0.240]	0.224	[0.172,0.273]	0.275	[0.234,0.316]
Tunnel exit indicator	–	–	–	–	0.042	[0.032,0.052]	0.051	[0.042,0.060]
Curvature$^{\sharp\sharp}$	-0.031	[-0.037,-0.025]	-0.061	[-0.076,-0.046]	-0.058	[-0.068,-0.048]	-0.052	[-0.063,-0.039]
Steep upgrade indicator	0.184	[0.160,0.209]	0.143	[0.120,0.163]	0.169	[0.147,0.193]	0.176	[0.148,0.201]
Steep downgrade indicator$^{\sharp\sharp}$	0.222	[0.189,0.251]	0.274	[0.227,0.323]	0.251	[0.213,0.284]	0.259	[0.215,0.306]
Proportion of class 5 vehicle$^{\sharp\sharp}$	0.103	[0.082,0.124]	0.109	[0.088,0.131]	0.104	[0.087,0.121]	0.102	[0.084,0.121]
PCI	–	–	-0.081	[-0.100,-0.062]	-0.063	[-0.075,-0.052]	-0.067	[-0.076,-0.058]
SRI$^{\sharp\sharp}$	-0.108	[-0.127,-0.089]	-0.106	[-0.127,-0.085]	-0.107	[-0.122,-0.888]	-0.101	[-0.123,-0.079]

(1) 95% BCI represents the 95% Bayesian credibility interval.
(2) $\sharp\sharp$ means that the parameters of these variables are identified as random for the URPNB-L and CRPNB-L models.
(3) – means the parameters are insignificant at the 95% Bayesian credibility level, and the corresponding variable do not participate in the estimation process of the model.

unobserved heterogeneity but also further explains the interactions between heterogeneity [54], [55].

B. ANALYSIS OF ESTIMATED REGRESSION PARAMETERS

The estimation results of the posterior means of parameters in each model are shown in Table 4. As can been seen from Table 4, 8 variables are significant at the 95% Bayesian credibility level in the FPNB model, while the FPNB-L model added 2 significant variables based on the 8 significant variables determined by the FPNB model, namely, length of tunnel and PCI (i.e., 10 significant variables). The URPNB-L and CRPNB-L models added a significant variable of the tunnel exit indicator based on the FPNB-L model (i.e., 11 significant variables).

Furthermore, the URPNB-L model detected four variables related to random parameters, namely, curvature, steep downgrade indicator, proportion of class 5 vehicle and SRI, whose mean and standard deviation were statistically significant. Table 5 and Fig. 6 present the standard deviations and distribution curves of random parameters in the URPNB-L model, respectively. According to Table 5, the standard deviations of random parameters follow the order of the proportion of class 5 vehicle > steep downgrade indicator > curvature > SRI. Among them, the random parameters related to the proportion of class 5 vehicle and the steep downgrade indicator exhibit strong fluctuations in each segment, indicating that the unobserved heterogeneity has prominent interactions with the above two influencing factors. By contrast, the standard deviations of random parameters related to curvature and SRI were small, indicating that the unobserved heterogeneity has some influences on these two influencing factors, but their degree was relatively weak. Furthermore, we can analyze the impacts of each random variable on crash frequency based on the URPNB-L model and conclude with the following conclusions from Fig. 6: (1) The parameter of curvature is positive in 20.96% of segments, that is, setting the curved

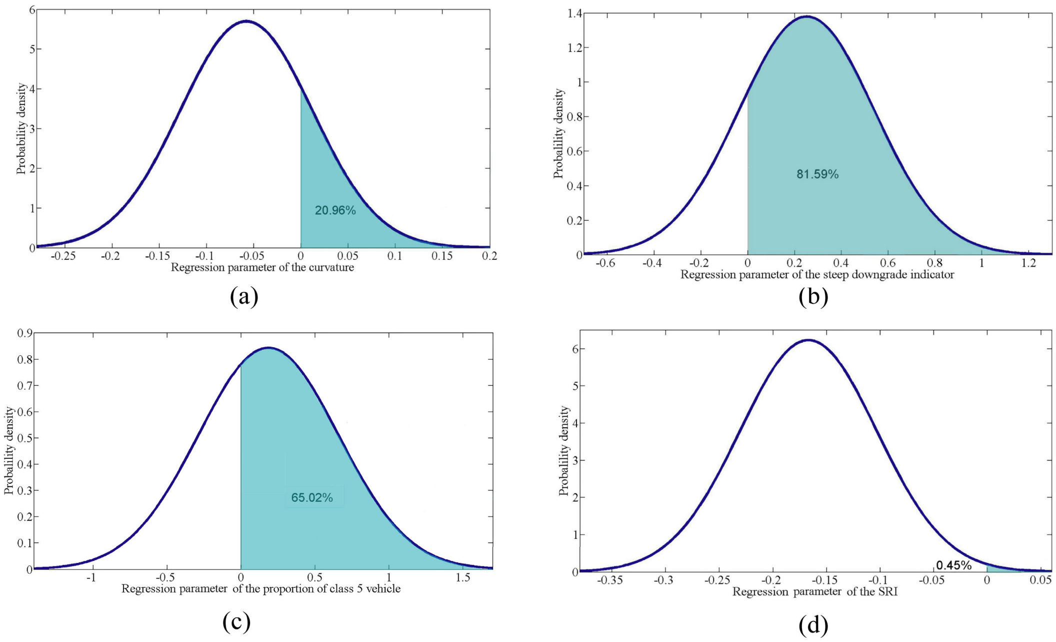


FIGURE 6. Distributions of random parameters in the URPNB-L model. (a) represents the random parameter distribution of the curvature. (b) represents the random parameter distribution of steep downgrade indicator. (c) represents the random parameter distribution of the proportion of class 5 vehicle. (d) represents the random parameter distribution of the SRI.

TABLE 5. Estimated standard deviations of random parameters in the URPNB-L model.

Variables	Standard deviation	95% BCI
Curvature	0.07	[0.057,0.083]
Steep downgrade indicator	0.289	[0.243,0.332]
Proportion of class 5 vehicle	0.437	[0.341,0.529]
SRI	0.054	[0.042,0.066]

95% BCI represents the 95% Bayesian credibility interval.

segments will increase the crash frequency, while the parameter of curvature is negative in 79.04% of segments, indicating that the curved segments will be beneficial to reduce the crash frequency. (2) 81.59% of segments indicated that the crash frequency of segments with a downhill slope of more than 2% was higher than that of other segments, while 18.41% of segments indicated that tunnel sections with a downhill slope of more than 2% was conducive to reducing the crash frequency. (3) The parameter of proportion of class 5 vehicle is positive for 65.02% of segments and is negative for 34.98% of segments, indicating that most of the segments justify that the proportion of class 5 vehicle is positively correlated with crash frequency. (4) The parameter of SRI was only positive in 0.45% of segments, indicating that most segments showed a negative correlation between SRI and crash frequency.

The random parameters detected by the CRPNB-L model were consistent with the random parameters detected by the URPNB-L model. However, the CRPNB-L model further explained the degree of correlation of each random parameter. Table 6 and Table 7 list the variance-covariance matrix and the correlation matrix of random parameters in the CRPNB-L model, respectively. Two main pieces of information can be obtained from Table 6: (1) The variance order of the four random parameters was proportion of class 5 vehicle > SRI > steep downgrade indicator > curvature. (2) The parameter covariance between the proportion of class 5 vehicle and the steep downgrade indicator was the largest, reaching 0.284, while the parameter covariance between the curvature and SRI was the smallest, at only 0.012. The following two main pieces of information can be obtained from Table 7: (1) All the random parameters were positively correlated. (2) The correlation coefficients of parameters between the proportion of class 5 vehicle and the steep downgrade indicator and between the curvature and the steep downgrade indicator reached 0.666 and 0.432, respectively, proving that the interaction effects between the above two pairs of random variables were obvious.

C. THE INFLUENCES OF VARIABLES AND INTERACTION EFFECTS OF RANDOM VARIABLES

In this section, we specifically analyzed the influences of significant variables and the interactions of random variables on the crash frequency for the CRPNB-L model according to the posterior mean values of parameters (as shown in Table 4 and Fig. 7.) and IRR values of significant variables (as shown in Table 8 and Fig. 7.).

1) THE INFLUENCES OF SIGNIFICANT VARIABLES ON CRASH FREQUENCY

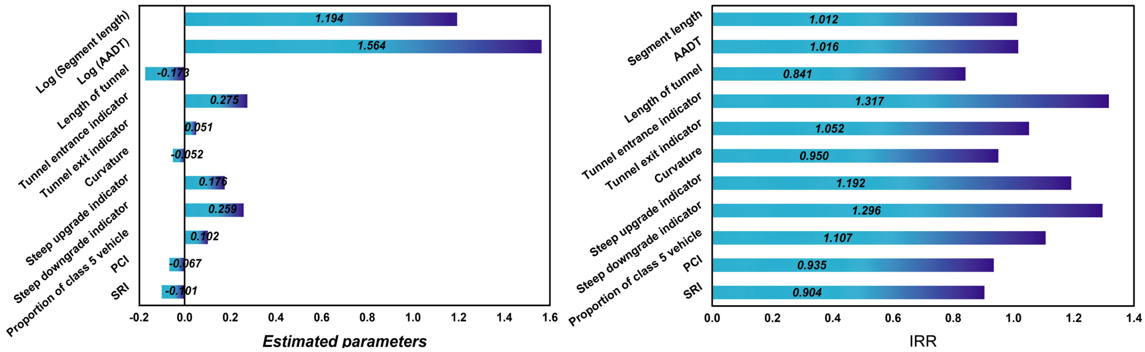
First, the influences of exposure variables on crash frequency were analyzed. Regression parameters of the Log (segment length) and the Log (AADT) were significantly positive at

TABLE 6. Variance-covariance matrix for the CRPNB-L model, the values in brackets represent the 95% Bayesian credibility interval.

Variables	Curvature	Steep downgrade indicator	Proportion of Class 5 vehicles	SRI
Curvature	0.154 [0.120,0.188]	0.084 [0.068,0.099]	0.105[0.081,0.128]	0.012 [0.010,0.014]
Steep downgrade indicator	0.084 [0.068,0.099]	0.245 [0.198,0.292]	0.284 [0.219,0.346]	0.065 [0.054,0.076]
Proportion of Class 5 vehicle	0.105 [0.081,0.128]	0.284 [0.219,0.346]	0.743 [0.580,0.906]	0.078 [0.066,0.090]
SRI	0.012 [0.010,0.014]	0.065 [0.054,0.076]	0.078 [0.066,0.090]	0.505 [0.439,0.576]

TABLE 7. Correlation matrix of random parameters for the CRPNB-L model.

Variables	Curvature	Steep downgrade indicator	Proportion of Class 5 vehicles	SRI
Curvature	1	0.432	0.310	0.043
Steep downgrade indicator	0.432	1	0.666	0.185
Proportion of Class 5 vehicle	0.310	0.666	1	0.127
SRI	0.043	0.185	0.127	1



(a) (b)

FIGURE 7. The estimated parameters and IRR values of significant variables based on the CRPNB-L models. (a) represents the estimated parameters of the CRPNB-L model. (b) represents the IRR values of the CRPNB-L model.

the 95% Bayesian credibility level with a posterior mean of 1.194 and a posterior mean of 1.564, respectively. These results indicated that larger traffic volume and longer tunnel segments have more crash frequencies, which was consistent with the results of relevant research [8], [9], [33], [34]. The IRR values of segment length and AADT are 1.012 and 1.016, respectively, which was insignificantly different from 1, meaning that the crash frequency increased almost linearly with the increase of the segment length or the AADT. Concretely, the crash frequency will increase by 1.2% if the segment length increases by 1%, while the crash frequency will increase by 1.6% if the AADT increases by 1%.

Then, the influences of tunnel design features on crash frequency were analyzed. The effects of the length of tunnel on crash frequency were significantly negative at the 95% Bayesian credibility level with a parameter value of −0.173, indicating that the increase of tunnel length could effectively reduce the crash frequency. According to experience, drivers in a closed tunnel environment for a long time will be more alert and less prone to traffic crashes. The IRR value (0.841) of the length of tunnel makes it clear that the crash frequency on the tunnels with a length longer than 500 m was reduced by 15.9% compared with the tunnels with a length shorter than 500 m. Similarly, the impacts of the curvature on crash frequency are significantly negative at the 95% Bayesian credibility level, with a parameter value of −0.052 and an IRR

TABLE 8. The IRR values of each significant variable in candidate models.

Variables	FPNB	FPNB-L	URPNB-L	CRPNB-L
Segment length	1.010	1.011	1.011	1.012
AADT	1.014	1.014	1.017	1.016
Length of tunnel	–	0.856	0.829	0.841
Tunnel entrance indicator	1.172	1.232	1.251	1.317
Tunnel exit indicator	–	–	1.043	1.052
Curvature	0.969	0.941	0.944	0.950
Steep upgrade indicator	1.202	1.154	1.184	1.192
Steep downgrade indicator	1.249	1.315	1.285	1.296
Proportion of class 5 vehicle	1.108	1.115	1.110	1.107
PCI	–	0.922	0.939	0.935
SRI	0.897	0.899	0.898	0.904

value of 0.950. The results provide evidence that the crash frequency will be nonlinearly reduced with the increase of curvature, where a 1% increase in curvature results in a 5% decrease in crash frequency. These estimation results are in line with the actual situation of highway tunnels in China. Since the curvature indexes of highway tunnels are strictly set in accordance with the design specifications, and even far better than the critical value in the specifications, there are usually no crashes caused by an unreasonable design of curvature, but caused by driver fatigue or distraction due to a curve's failure to reach the alert level or an insufficient warning level. A driver may thus exhibit an appropriate visual impact when driving on sharp curvature segments in a tunnel, which forces the driver's attention to be highly concentrated and the driver to drive in a relatively safe state. The tunnel entrance indicator and tunnel exit indicator are both significantly positive at the 95% Bayesian credibility level, with posterior means of parameters at 0.275 and 0.051, respectively, indicating that the crash frequency at the tunnel entrance and exit is higher than that at other tunnel segments. The IRR values of the tunnel entrance indicator and tunnel exit indicator (values of 1.317 and 1.052) also clearly proved that the crash frequency at the tunnel entrance/exit increased by 31.7%/5.2% than that of other tunnel segments, respectively. Apparently, The sudden changes in environment at a tunnel entrance or exit, including the light changing and the visual changes, have great impacts on the physiology of the drivers, increasing the driver's safe operation standard and easily leading to crashes. In addition, tunnel segments with upper and lower longitudinal slopes greater than 2% are also prone to crashes, due to the parameter value of the steep upgrade indicator at 0.176 and the parameter value of the steep downgrade indicator at 0.259. It can be concluded from IRR values that the crash frequency of tunnel segments with an upper or lower longitudinal slope degree greater than 2% increases by 19.2% and 29.6%, respectively, compared with other tunnel segments. This phenomenon is easy to explain because the longitudinal slope has certain influences on the vehicle braking performance, which can easily cause vehicle braking failures and occurrences of crashes.

In terms of traffic conditions, the proportion of class 5 vehicle is significantly positively correlated with the crash frequency at the 95% Bayesian credibility level with the posterior mean of the parameter at 0.102, indicating that the greater the proportion of class 5 vehicle there is, the greater is the crash frequency of this tunnel segment. In reality, class 5 vehicles, comprising heavy trucks, heavy trailers and 40-foot container trucks limit the driving vision of standard vehicles in their proximity, so class 5 vehicles are more likely to have crashes with standard vehicles in their proximity in a closed tunnel environment. The IRR value of the proportion of class 5 vehicle is 1.107, explaining that there is a nonlinear relationship between class 5 vehicles and the crash frequency. Specifically, if the proportion of class 5 vehicle increases by 1%, the crash frequency will increase by 10.7%.

Finally, the influences of pavement conditions on tunnel safety were analyzed. The estimation results showed that PCI and SRI had significantly negative impacts on crash frequency at the 95% Bayesian credibility level, with parameter posterior means of −0.067 and −0.101, respectively. The results are also easily understood since drivers have a high safety rate when driving on tunnel segments with good pavement integrity and better anti-skid performance in most cases. It is worth noting that the IRR values of PCI and SRI are 0.935 and 0.904, respectively, indicating their nonlinear relationship with crash frequency. Specifically, crash frequency decreased by 6.5% when PCI increased by 1 unit. While crash frequency decreased by 9.6%, when SRI increased by 1 unit.

2) THE INTERACTION EFFECTS OF RANDOM VARIABLES ON CRASH FREQUENCY

According Table 7, the interaction effects of random variables on crash frequency can be analyzed. These interactions of all random variables were positive correlations, where the interaction between the proportion of class 5 vehicle and the steep downgrade indicator and between the curvature and the steep downgrade indicator were most obvious, with correlation coefficients of 0.666 and 0432, respectively. The interaction effects of each group of random variables are explained as follows.

- The interaction effects between the proportion of class 5 vehicle and the steep downgrade indicator were positively correlated, with a correlation coefficient of 0.666 (close to 1), which indicated that when steep downgrade tunnel segments existed a high proportion of class 5 vehicles, the crash frequency of these segments is significantly increased. The reason is that a large number of class 5 vehicle driving on steep downhill tunnel segments are prone to brake failure and crashes with standard cars around them.
- The presence of a high proportion of class 5 vehicle in curved tunnel segments led to a higher crash frequency, with a correlation coefficient of 0.310. According to experience, the tunnel segments with a high proportion of 5 vehicle have limited vision, especially on the curved tunnel segments. When standard vehicles drive on the curved tunnel segments with class 5 vehicles at the same time, the narrow vision will hinder drivers' correct operation and increase the crash frequency.
- The combined effects of the proportion of class 5 vehicle and the SRI were positively correlated with a correlation coefficient of 0.127, which indicated that the combined effects of these two factors increased the driving risk in tunnel segments to some extent. These results are also easy to understand. Due to the large volume and heavy weight of class 5 vehicles, the anti-skid performance of the pavement can easily affect the operation stability of the class 5 vehicles. Therefore, when class 5 vehicles drive on the segments with a low SRI value, the vehicles are more likely to lose control and lead to crashes.

- The combined effects of the curvature and the steep downgrade indicator were positive with a correlation coefficient of 0.432, which meant that a greater crash frequency will exist in tunnel segments that combine curved and downhill segments. The tunnel segments combined curved and longitudinal slope sections improves the operation difficulty of drivers, it is difficult for drivers to turn smoothly and quickly on the tunnel segments that combined with curved and steep downslope sections.
- Steep downslope tunnel segments with poor skid resistance had a higher crash frequency, with a correlation coefficient of 0.185. The reason is also clear given that vehicles are prone to brake failure when driving on the steep tunnel segments with poor skid resistance.
- Curved tunnel segments with poor skid resistance had a higher crash frequency. The main reason is that it may lead to the failure of the vehicle braking performance when turning on the curved tunnel segments with poor skid resistance. It should be noted that the correlation coefficient between the curvature and SRI is only 0.043, indicating that the interaction effects of the above two random variables have weak influences on crash frequency.

V. CONCLUSION

Based on the safety status of highway tunnels in China, this paper discusses crash frequency models of highway tunnels and analyses the influences of factors and the interactions of random variables. First, we collected a total of 545 tunnel crashes on typical highways in Guangdong for 3 to 5 years and three types of influence factors of tunnel design features, traffic conditions and pavement conditions to establish an appropriate dataset. Based on the established dataset, we proposed a CRPNB-L model for fitting crash frequency, which simultaneously considers excess zero observations, unobserved heterogeneity and its interaction effects existed in tunnel crash dataset. The CRPNB-L model was compared with its corresponding FPNB, FPNB-L and URPNB-L models in terms of goodness-of-fit with DIC, MAD and MSPE, and the results verified that the CRPNB-L model can deeply reveal the influences of various factors and their interactions on crash frequency in highway tunnels, resulting in a better effectiveness in terms of goodness-of-fit. Finally, we took the CRPNB-L model as the object to analyze the influencing factors of tunnel safety. The CRPNB-L model detected 11 significant variables at the 95% Bayesian credibility level, among which the estimated parameters of 7 variables were fixed and those of 4 variables were random parameters, namely, curvature, steep downgrade indicator, proportion of class 5 vehicle and SRI. The impacts of length of tunnel, curvature, PCI and SRI on crash frequency were significantly negative, while the other significant variables were positively correlated with crash frequency. The IRR values indicated that the tunnel entrance indicator, the steep upgrade indicator, the steep downgrade indicator and the proportion of class 5 vehicle had more prominent impacts on crash frequency. On the other hand, the CRPNB-L model explained the expected interaction effects of random variables as follows: (1) Interaction effects between the proportion of class 5 vehicle and the steep downgrade indicator and between the steep downgrade indicator and the curvature on crash frequency were relatively strong with correlation coefficients of 0.666 and 0.432, respectively. (2) When three cases exist, namely, curved tunnel segments with numerous class 5 vehicle, downslope tunnel segments that have poor anti-sliding performance of the pavement, and class 5 vehicles that are traveling on pavement with anti-sliding performance, the crash frequency of these tunnel segments will be increased to a certain extent. (3) The interaction effects between curvature and SRI on crash frequency are positive, but their influence is relatively weak with a correlation coefficient of 0.043.

The results from this study are expected to provide a better understanding of how tunnel design features, traffic conditions, and pavement conditions influence crash frequency in freeway tunnels and provide some suggestions on tunnel safety measures. For example, the degree of longitudinal slope in the tunnel segments should be less than 2% as far as possible, especially when the proportion of class 5 vehicle is expected to be high. The entrance and exit of tunnels should be equipped with driving warning signs, especially in the tunnel entrance. In addition, this study addressed the importance of road surface maintenance from the aspects of PCI and SRI. Maintaining road surfaces in good condition could significantly decrease the crash frequency in tunnels.

REFERENCES

[1] P. C. Anastasopoulos, F. L. Mannering, V. N. Shankar, and J. E. Haddock, "A study of factors affecting highway accident rates using the random-parameters tobit model," *Accident Anal. Prevention*, vol. 45, pp. 628–633, Mar. 2012.

[2] M. D. Luca and G. Dell'Acqua, "Freeway safety management: Case studies in Italy," *Transport*, vol. 27, no. 3, pp. 320–326, Sep. 2012.

[3] P. Ntzeremes, K. Kirytopoulos, and V. Leopoulos, "Development of a risk-based method for predicting the severity of potential fire accidents in road tunnels based on real-time data," *Environ. Res.*, vol. 189, Oct. 2020, Art. no. 109895.

[4] Y. Yan, Y. Dai, X. Li, J. Tang, and Z. Guo, "Driving risk assessment using driving behavior data under continuous tunnel environment," *Traffic Injury Prevention*, vol. 20, no. 8, pp. 807–812, Nov. 2019.

[5] C. Caliendo, M. L. De Guglielmo, and M. Guida, "A crash-prediction model for road tunnels," *Accident Anal. Prevention*, vol. 55, pp. 107–115, Jul. 2013.

[6] (2008). *Human Factors and Road Tunnel Safety Regarding Users*. [Online]. Available: http://www.piarc.org

[7] *Ministry of Transport of the People's Republic of China*, Road Traffic Accidents Statistics of PRC, Nov. 2019. [Online]. Available: http://www.mot.gov.cn/

[8] R. Rusli, M. M. Haque, A. P. Afghari, and M. King, "Applying a random parameters negative binomial lindley model to examine multi-vehicle crashes along rural mountainous highways in malaysia," *Accident Anal. Prevention*, vol. 119, pp. 80–90, Oct. 2018.

[9] Q. Zeng, H. Wen, H. Huang, and M. Abdel-Aty, "A Bayesian spatial random parameters tobit model for analyzing crash rates on roadway segments," *Accident Anal. Prevention*, vol. 100, pp. 37–43, Mar. 2017.

[10] HY. Wen, X. Zhang, Q. Zeng, and NN. Sze, "Bayesian spatial-temporal model for the main and interaction effects of roadway and weather characteristics on freeway crash incidence," *Accid. Anal. Prev.*, vol. 132, pp. 44–51, Nov. 2019.

[11] MA. Abdel-Aty and AE. Radwan, "Modeling traffic accident occurrence and involvement," *Accid. Anal. Prev.*, vol. 32, no. 5, pp. 633–642, Sep. 2000.

[12] B. Wali, A. J. Khattak, J. Waters, D. Chimba, and X. Li, "Development of safety performance functions: Incorporating unobserved heterogeneity and functional form analysis," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2672, no. 30, pp. 9–20, Dec. 2018.

[13] P. C. Anastasopoulos and F. L. Mannering, "A note on modeling vehicle accident frequencies with random-parameters count models," *Accident Anal. Prevention*, vol. 41, no. 1, pp. 153–159, Jan. 2009.

[14] Y. Lao, G. Zhang, Y. Wang, and J. Milton, "Generalized nonlinear models for rear-end crash risk analysis," *Accident Anal. Prevention*, vol. 62, pp. 9–16, Jan. 2014.

[15] M. Hosseinpour, A. S. Yahaya, and A. F. Sadullah, "Exploring the effects of roadway characteristics on the frequency and severity of head-on crashes: Case studies from malaysian federal roads," *Accident Anal. Prevention*, vol. 62, pp. 209–222, Jan. 2014.

[16] M. M. Ahmed, M. Abdel-Aty, J. Lee, and RJ. Yu, "Real-time assessment of fog-related crashes using airport weather data: A feasibility analysis," *Accid. Anal. Prev.*, vol. 62, pp. 309–317, Jan. 2014.

[17] Z. Ma, H. Zhang, S. I.-J. Chien, J. Wang, and C. Dong, "Predicting expressway crash frequency using a random effect negative binomial model: A case study in China," *Accident Anal. Prevention*, vol. 98, pp. 214–222, Jan. 2017.

[18] B. Naik, L.-W. Tung, S. Zhao, and A. J. Khattak, "Weather impacts on single-vehicle truck crash injury severity," *J. Saf. Res.*, vol. 58, pp. 57–65, Sep. 2016.

[19] K. El-Basyouny and T. Sayed, "A full Bayes multivariate intervention model with random parameters among matched pairs for before–after safety evaluation," *Accident Anal. Prevention*, vol. 43, no. 1, pp. 87–94, Jan. 2011.

[20] R. Yu, M. Abdel-Aty, and M. Ahmed, "Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors," *Accident Anal. Prevention*, vol. 50, pp. 371–376, Jan. 2013.

[21] R. Fu, Y. Guo, W. Yuan, H. Feng, and Y. Ma, "The correlation between gradients of descending roads and accident rates," *Saf. Sci.*, vol. 49, no. 3, pp. 416–423, Mar. 2011.

[22] R. Rusli, M. M. Haque, M. King, and W. S. Voon, "Single-vehicle crashes along rural mountainous highways in malaysia: An application of random parameters negative binomial model," *Accident Anal. Prevention*, vol. 102, pp. 153–164, Mar. 2017.

[23] J. Tang, L. Zheng, C. Han, W. Yin, Y. Zhang, Y. Zou, and H. Huang, "Statistical and machine-learning methods for clearance time prediction of road incidents: A methodology review," *Anal. Methods Accident Res.*, vol. 27, Sep. 2020, Art. no. 100123.

[24] E. T. Donnell and J. M. Mason, "Predicting the frequency of median barrier crashes on pennsylvania interstate highways," *Accident Anal. Prevention*, vol. 38, no. 3, pp. 590–599, May 2006.

[25] P. Buddhavarapu, A. Banerjee, and J. A. Prozzi, "Influence of pavement condition on horizontal curve safety," *Accident Anal. Prevention*, vol. 52, pp. 9–18, Mar. 2013.

[26] VN. Shankar, RB. Albin, JC. Milton, and F. Mannering, "Evaluating median crossover likelihoods with clustered accident counts-An empirical inquiry using the random effects negative binomial model," in *Proc. 77th Annu. Meeting Transp.-Res.-Board*, Washington, DC, USA, 1998, vol. 1635, pp. 44–48.

[27] A. Pirdavani, T. Brijs, T. Bellemans, B. Kochan, and G. Wets, "Evaluating the road safety effects of a fuel cost increase measure by means of zonal crash prediction modeling," *Accid. Anal. Prev.*, vol. 50, pp. 186–195, Jan. 2013.

[28] A. Montella and L. L. Imbriani, "Safety performance functions incorporating design consistency variables," *Accident Anal. Prevention*, vol. 74, pp. 133–144, Jan. 2015.

[29] L. Garach, J. de Oña, G. López, and L. Baena, "Development of safety performance functions for spanish two-lane rural highways on flat Terrain," *Accident Anal. Prevention*, vol. 95, pp. 250–265, Oct. 2016.

[30] M. J. T. L. Gomes, F. Cunto, and A. R. da Silva, "Geographically weighted negative binomial regression applied to zonal level safety performance models," *Accident Anal. Prevention*, vol. 106, pp. 254–261, Sep. 2017.

[31] SP. Miaou, "The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions," *Accid. Anal. Prev.*, vol. 26, no. 4, pp. 471–482, Aug. 1994.

[32] V. Shankar, J. Milton, and F. Mannering, "Modeling accident frequencies as zero-altered probability processes: An empirical inquiry," *Accid. Anal. Prev.*, vol. 29, no. 6, pp. 829–837, Nov. 1997.

[33] D. Lord, S. Washington, and J. N. Ivan, "Further notes on the application of zero-inflated models in highway safety," *Accident Anal. Prevention*, vol. 39, no. 1, pp. 53–57, Jan. 2007.

[34] K. K. W. Yau, K. Wang, and A. H. Lee, "Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros," *Biometrical J.*, vol. 45, no. 4, pp. 437–452, Jun. 2003.

[35] C. Dong, J. Shi, B. Huang, X. Chen, and Z. Ma, "A study of factors affecting intersection crash frequencies using random-parameter multivariate zero-inflated models," *Int. J. Injury Control Saf. Promotion*, vol. 24, no. 2, pp. 208–221, Jun. 2017.

[36] D. Lord, SP. Washington, and JN. Ivan, "Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory," in *Proc. 83rd Annu. Meeting Transp.-Res.-Board*, Washington, DC, USA, Jan. 2004, vol. 37, no. 1, pp. 35–46.

[37] N. V. Malyshkina, F. L. Mannering, and A. P. Tarko, "Markov switching negative binomial models: An application to vehicle accident frequencies," *Accident Anal. Prevention*, vol. 41, no. 2, pp. 217–226, Mar. 2009.

[38] N. V. Malyshkina and F. L. Mannering, "Zero-state Markov switching count-data models: An empirical assessment," *Accident Anal. Prevention*, vol. 42, no. 1, pp. 122–130, Jan. 2010.

[39] A. Behnood and F. L. Mannering, "An empirical assessment of the effects of economic recessions on pedestrian-injury crashes using mixed and latent-class models," *Analytic Methods Accident Res.*, vol. 12, pp. 1–17, Dec. 2016.

[40] S. R. Geedipally, D. Lord, and S. S. Dhavala, "The negative binomial-lindley generalized linear model: Characteristics and application using crash data," *Accident Anal. Prevention*, vol. 45, pp. 258–265, Mar. 2012.

[41] D. Lord and S. R. Geedipally, "The negative binomial–Lindley distribution as a tool for analyzing crash data characterized by a large amount of zeros," *Accident Anal. Prevention*, vol. 43, no. 5, pp. 1738–1742, Sep. 2011.

[42] P. Saengthong and W. Bodhisuwan. (2013). *Negative Binomial-Crack (NB-CR) Distribution*. [Online]. Available: https://arxiv.org/abs/10.12732.

[43] P. Vangala, D. Lord, and S. R. Geedipally, "Exploring the application of the negative Binomial–Generalized exponential model for analyzing traffic crash data with excess zeros," *Analytic Methods Accident Res.*, vol. 7, pp. 29–36, Jul. 2015.

[44] Y. Yan, Y. Zhang, X. Yang, J. Hu, J. Tang, and Z. Guo, "Crash prediction based on random effect negative binomial model considering data heterogeneity," *Phys. A, Stat. Mech. Appl.*, vol. 547, Jun. 2020, Art. no. 123858.

[45] F. L. Mannering, V. Shankar, and C. R. Bhat, "Unobserved heterogeneity and the statistical analysis of highway accident data," *Analytic Methods Accident Res.*, vol. 11, pp. 1–16, Sep. 2016.

[46] SP. Washington, MG. Karlafitis, and FL. Mannering, *Statistical and Econometric Methods of Transportation Data Analysis*. New York, NY, USA: Chapman-Hall, Nov. 2011.

[47] F. Naznin, G. Currie, D. Logan, and M. Sarvi, "Application of a random effects negative binomial model to examine tram-involved crash frequency on route sections in melbourne, australia," *Accident Anal. Prevention*, vol. 92, pp. 15–21, Jul. 2016.

[48] E. Chen and A. P. Tarko, "Modeling safety of highway work zones with random parameters and random effects models," *Analytic Methods Accident Res.*, vol. 1, pp. 86–95, Jan. 2014.

[49] P. Xu and H. Huang, "Modeling crash spatial heterogeneity: Random parameter versus geographically weighting," *Accident Anal. Prevention*, vol. 75, pp. 16–25, Feb. 2015.

[50] P. Xu, H. Huang, N. Dong, and S. C. Wong, "Revisiting crash spatial heterogeneity: A Bayesian spatially varying coefficients approach," *Accident Anal. Prevention*, vol. 98, pp. 330–337, Jan. 2017.

[51] K. S. Conway and T. J. Kniesner, "The important econometric features of a linear regression model with cross-correlated random coefficients," *Econ. Lett.*, vol. 35, no. 2, pp. 143–147, Feb. 1991.

[52] R. Yu, Y. Xiong, and M. Abdel-Aty, "A correlated random parameter approach to investigate the effects of weather conditions on crash risk for a mountainous freeway," *Transp. Res. C, Emerg. Technol.*, vol. 50, pp. 68–77, Jan. 2015.

[53] E. Coruh, A. Bilgic, and A. Tortum, "Accident analysis with aggregated data: The random parameters negative binomial panel count data model," *Analytic Methods Accident Res.*, vol. 7, pp. 37–49, Jul. 2015.

[54] Q. Hou, A. P. Tarko, and X. Meng, "Analyzing crash frequency in freeway tunnels: A correlated random parameters approach," *Accident Anal. Prevention*, vol. 111, pp. 94–100, Feb. 2018.

[55] Q. Hou, X. Huo, and J. Leng, "A correlated random parameters tobit model to analyze the safety effects and temporal instability of factors affecting crash rates," *Accident Anal. Prevention*, vol. 134, Jan. 2020, Art. no. 105326.

[56] M. Ahmed, H. Huang, M. Abdel-Aty, and B. Guevara, "Exploring a Bayesian hierarchical approach for developing safety performance functions for a mountainous freeway," *Accident Anal. Prevention*, vol. 43, no. 4, pp. 1581–1589, Jul. 2011.

[57] Q. Zeng, H. Wen, H. Huang, X. Pei, and S. C. Wong, "Incorporating temporal correlation into a multivariate random parameters tobit model for modeling crash rate by injury severity," *Transportmetrica A, Transp. Sci.*, vol. 14, no. 3, pp. 177–191, Mar. 2018.

[58] F. H. Amundsen and G. Ranes, "Studies on traffic accidents in norwegian road tunnels," *Tunnelling Underground Space Technol.*, vol. 15, no. 1, pp. 3–11, Jan. 2000.

[59] R. Brandt, M. Schubert, and N. P. Hoj, "On risk analysis of complex road tunnel system.," in *Proc. 6th Int. Conf. Tunnel Saf. Ventilation*, Graz, Austria, 2012, p.4.

[60] L. Lu, J. Lu, Y. Xing, C. Wang, and F. Pan, "Statistical analysis of traffic accidents in shanghai river crossing tunnels and safety countermeasures," *Discrete Dyn. Nature Soc.*, vol. 2014, pp. 1–7, Jan. 2014.

[61] J. S. Yeung and Y. D. Wong, "Road traffic accidents in singapore expressway tunnels," *Tunnelling Underground Space Technol.*, vol. 38, pp. 534–541, Sep. 2013.

[62] J. Wang, A. Pervez, Z. Wang, C. Han, L. Hu, and H. Huang, "Crash analysis of chinese freeway tunnel groups using a five-zone analytic approach," *Tunnelling Underground Space Technol.*, vol. 82, pp. 358–365, Dec. 2018.

[63] A. Pervez, H. Huang, J. Lee, C. Han, J. Wang, and X. Zhang, "Crash analysis of expressway long tunnels using a seven-zone analytic approach," *J. Transp. Saf. Secur.*, pp. 1–15, Jun. 2019.

[64] C. Caliendo, M. L. De Guglielmo, and M. Guida, "Comparison and analysis of road tunnel traffic accident frequencies and rates using random-parameter models," *J. Transp. Saf. Secur.*, vol. 8, no. 2, pp. 177–195, Apr. 2016.

[65] C. Caliendo, M. L. De Guglielmo, and I. Russo, "Analysis of crash frequency in motorway tunnels based on a correlated random-parameters approach," *Tunnelling Underground Space Technol.*, vol. 85, pp. 243–251, Mar. 2019.

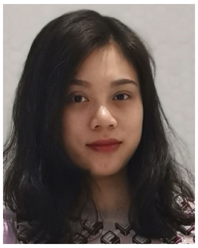
FENG TANG was born in Huaihua, China, in 1992. He received the B.S. degree in civil engineering from the Changsha University of Science and Technology, Changsha, China, in 2015. He is currently pursuing the Ph.D. degree with the School of Civil and Transportation Engineering, South China University of Technology, Guangzhou, China. His research interests include traffic safety analysis, and intelligent transportation systems (ITSs).

XINSHA FU was born in Changsha, China, in 1955. He received the B.S. degree in highway and bridge engineering from the Changsha University of Science and Technology, Changsha, China, in 1981, and the M.S. degree in highway engineering and the Ph.D. degree from Chang'an University, Xi'an, China, in 1987 and 2008, respectively.

From 1981 to 2000, he was a Professor with the Changsha University of Science and Technology. Since 2000, he has been a Professor with the South China University of Technology, Guangzhou, China. His research interests include theoretical research on highway alignment design, research on highway alignment and computer-aided design, road safety audit, and intelligent transportation systems (ITSs). He served for China Highway Society Youth Expert Committee, China's institutions of higher learning civil engineering professional guidance committee, the Director of the Institute of Computer Application, an Academic Committee Member of road of China Highway Society, China faces the transport in the 21st century version of the Institutions of Higher Learning Materials (Highway) Editorial Committee, the computer-aided engineering staff, and central south highway project staff.

MINGMAO CAI was born in Fuzhou, China, in 1997. He received the B.S. degree in traffic engineering from the Qingdao University of Technology, Qingdao, China, in 2019. He is currently pursuing the M.S. degree with the School of Civil and Transportation Engineering, South China University of Technology, Guangzhou, China. His research interests include traffic safety analysis and intelligent transportation systems (ITSs).

YUE LU was born in Shijiazhuang, China, in 1992. He received the B.S. degree in civil engineering from the Changsha University of Science and Technology, Changsha, China, in 2015, and the M.S. degree in highway and railway engineering from the South China University of Technology, Guangzhou, China, in 2018, where he is currently pursuing the Ph.D. degree with the School of Civil and Transportation Engineering. His research interests include traffic safety analysis and highway route design theory.



SHIYU ZHONG was born in Ganzhou, China, in 1992. She received the B.S. degree from the Southwest China University of Science and Technology, Mianyang, China, in 2013. She is currently pursuing the Ph.D. degree in civil engineering and transportation with the South China University of Technology, Guangzhou, China. Her research interests include transportation data quality issues, intelligent transportation systems, and transportation safety.

CHONGZHEN LU was born in Shangrao, China, in 1997. He received the B.S. degree in civil engineering from Nanchang Hangkong University, Nanchang, China, in 2013. He is currently pursuing the M.S. degree with the School of Civil and Transportation Engineering, South China University of Technology, Guangzhou, China. His research interests include traffic safety analysis and intelligent transportation systems (ITSs).