

Research Article

Random-Parameter Multivariate Negative Binomial Regression for Modeling Impacts of Contributing Factors on the Crash Frequency by Crash Types

Chenzhu Wang,¹ Fei Chen ,¹ Jianchuan Cheng,¹ Wu Bo,² Ping Zhang,² Mingyu Hou,¹ and Feng Xiao ¹

¹School of Transportation, Southeast Univ.2 Sipailou, Nanjing, Jiangsu 210096, China

²Tibet University, No. 36 Jiangsu, Lhasa, Tibet 850000, China

Correspondence should be addressed to Fei Chen; cf@seu.edu.cn

Received 11 October 2020; Revised 6 November 2020; Accepted 12 November 2020; Published 28 November 2020

Academic Editor: Tingsong Wang

Copyright © 2020 Chenzhu Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Highways provide the basis for safe and efficient driving. Road geometry plays a critical role in dynamic driving systems. Contributing factors such as plane, longitudinal alignment, and traffic volume, as well as drivers' sight characteristics, determine the safe operating speed of cars and trucks. In turn, the operating speed influences the frequency and type of crashes on the highways. *Methods.* Independent negative binomial and Poisson models are considered as the base approaches to modeling in this study. However, random-parameter models reduce unobserved heterogeneity and obtain higher dimensions. Therefore, we propose the random-parameter multivariate negative binomial (RPMNB) model to analyze the influence of the traffic, speed, road geometry, and sight characteristics on the rear-end, bumping-guardrail, other, noncasualty, and casualty crashes. Subsequently, we compute the goodness-of-fit and predictive measures to confirm the superiority of the proposed model. Finally, we also calculate the elasticity effects to augment the comparison. *Results.* Among the significant variables, black spots, average annual daily traffic volume (AADT), operating speed of cars, speed difference of cars, and length of the present plane curve positively influence the crash risk, whereas the speed difference of trucks, length of the longitudinal slope corresponding to the minimum grade, and stopping sight distance negatively influence the crash risk. Based on the results, several practical and efficient measures can be taken to promote safety during the road design and operating processes. Moreover, the goodness-of-fit and predictive measures clearly highlight the greater performance of the RPMNB model compared to standard models. The elasticity effects across all the models show comparable performance with the RPMNB model. Thus, the RPMNB model reduces the unobserved heterogeneity and yields better performance in terms of precision, with more consistent explanatory power compared to the traditional models.

1. Introduction

The highway is a dynamic system that includes drivers, vehicles, roads, and environment. Drivers play important roles, and inattention, distraction, and misjudgment are the main causes of highway crashes. Nowadays, research studies mainly focus on drivers. Meanwhile, the road characteristics, which serve as an important basis of the dynamic system and influence safety, are often neglected [1–3]. As the basis of driving, road geometry has a very important impact on

safety. Almost 40% of all crashes attributed to driver mistake or vehicle failure are fundamentally caused by road geometry [4]. Babukov et al. studied the internal relationship between plane, profile parameters, and road safety in the design of highways and made significant suggestions for improving safety [4–7]. McGee et al. established statistical models between the curve length, curvature, traffic volume, and the crash frequency [8].

After studying 141,812 crashes, Pei and Ma focused on the relationship between road design factors (plane, vertical,

cross section, and intersection) and crashes and put forward corresponding countermeasures [9]. Many types of research have evaluated the safety of road geometry by visualization and computer simulation. Hassan et al. simulated the influence of the 3D environment on driver vision and analyzed the influence of a combination of the horizontal and vertical curve on road safety [10–14]. Then, the Federal Highway Administration developed the Interactive Highway Safety Design Model (IHSDM). Software used the lane width, lane number, horizontal alignment, vertical alignment, cross section, and superelevation as the evaluation parameters to check the consistency of the design [15–17]. Drivers often attempt to improve the efficiency of transportation when road and traffic conditions allow speeding, which causes many crashes on the highway. The Solomon and the Monash University Accident Research Center modeled the relationship between crash rate and vehicle speed and found that the higher the difference between the operating speed and the average speed, the higher the crash rate [18, 19]. The average speed is often exceeded because of the favorable geometry and environmental conditions of modern highways. Meanwhile, the speed difference between cars and trucks increases owing to the relatively poor power and braking performance of trucks, which lead to frequent rear-end crashes.

The initial road alignment design was based on the design speed, which was first proposed by the American Association of State Highway and Transportation Officials (AASHTO) as the key factor in road design. At the same time, designers and researchers also used design speed to evaluate the consistency of horizontal and longitudinal profiles. The expected speed would greatly exceed the design speed during favorable road conditions, and the traditional design method based on the design speed led to poor continuity and imbalance in actual driving. In addition, the drivers' visual perception expressed poor consistency with actual vehicle control [20]. Due to these limitations, AASHTO later adopted the operating speed to dominate the design, which is defined as the driving speed at the 85th percentile selected by the actual measurement of medium-skilled drivers under a free-flow state with good weather and road conditions [21]. Lamm et al. studied the influence of curve on the operating speed based on the design data, vehicle speed, traffic volume, and crash data of 261 inter-continental highway sections and determined the model of the curve radius and operating speed [22, 23]. Anderson et al. investigated the influence of curve radius on the operating speed according to the speed attenuation while passing the curve [24, 25]. Krammes et al. established the operation speed model by collecting the operating speed and design data of 138 plane curve sections based on 1,126 curve sections as samples for a preliminary evaluation to determine the causes of accidents in curve sections and select improvement methods [26]. Collins et al. evaluated the free-flow condition by measuring the actual driving speed at the midpoint of the curve [27–33]. Sil and Maji measured the operating speed through the curve on three sections and

established a model between the speed at the curve and the combined curves on the four-lane highway [34]. Based on the existing research abroad, the operating speed model of different vehicle types has been adopted in research in China. In 2004, the Guidelines for Safety Evaluation of Highway Projects proposed the highway speed model based on many research datasets. Specifications for the Highway Safety Audit published in 2015 optimized and improved the speed model of low-grade highways [35]. The evaluated sections are divided into tangent sections, longitudinal slope sections, horizontal curve sections, curved slope combination sections, tunnel sections, interchange sections, and other sections for calculation according to the radius of the horizontal curve and the slope of the profile. A sufficient driving sight distance is directly related to the safety and speed on the highway and is an important index of evaluation. The stopping sight distance enables drivers to stop, meet, and overtake smoothly, and the transverse clear distance measures the safe zone allowed for lateral offset during driving. Both distances are measurable representations of sight characteristics [35].

In summary, previous research studies focused on the contributing factors of road geometry or operating speed characteristics separately, and the sight characteristics determined by the road geometry alignments were rarely studied. This study takes the geometric, speed, and sight characteristics into consideration simultaneously under a single condition and then adopts the random-parameter approach to reduce the interactions between various variables to acquire a higher dimension and discover the contributing factors. The traditional negative binomial and Poisson models used in previous research studies become complicated when analyzing crash types. In contrast, this random-parameter multivariate negative binomial (RPMNB) model allows for a more simplified framework and reduces the unobserved heterogeneity with higher accuracy. In this study, we collect three-year (2009–2011) statistical crash data of the Beijing-Shanghai Highway as a sample to analyze whether the road geometry, operating speed, traffic volume, and sight characteristics affect the crash types (rear-end, bumping-guardrail, roll-over, etc.).

2. Methodology

Many methodological approaches, such as the multivariate Poisson (lognormal) model, zero-inflated negative binomial model, and Poisson lognormal spatial and/or temporal model typically address the crash rate considering the number of crashes occurring over a roadway segment or at an intersection [36–38].

Generally, some of the significant factors affecting the crash rate are missed in the collected data or difficult to analyze. These factors (called unobserved heterogeneity) make a variation in the influence on the observed factors, which may lead to misspecified parameters and erroneous inferences. To reduce the influence of the unobserved

heterogeneity, random parameters, latent-class (finite-mixture) models, and Markov switching models are considered in the multivariate models [39].

2.1. Poisson (Negative Binomial) Regression Model. In statistics researches on crash frequency, Poisson regression is a generalized linear model analysis used to model crash data. The response variable Y is assumed yielding to Poisson distribution, with the expected value modeled by a linear combination of unknown parameters. In Poisson regression, the Poisson incidence rate μ is determined by X_k (the regressor variables) [40–42]:

$$\mu = te^{(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}. \quad (1)$$

The fundamental Poisson regression model (PRM) for an observation i is written as

$$\Pr(Y = y_i | \mu_i, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} \left(\frac{1}{1 + \alpha\mu_i}\right)^{\alpha^{-1}} \left(\frac{\alpha\mu_i}{1 + \alpha\mu_i}\right)^{y_i}, \quad (3)$$

$$\mu_i = e^{(\ln(t_i) + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})}. \quad (4)$$

When α is significantly different from 0, the negative binomial regression is appropriate. Otherwise, the Poisson model is better.

2.2. Zero-Inflated Poisson Model. When the crash data contain excess zero-count values in the model, the well-known zero-inflated Poisson model is adopted.

$$\begin{aligned} \Pr(Y = 0) &= \pi + (1 - \pi)e^{-\lambda}, \\ \Pr(Y = y_i) &= (1 - \pi) \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}, \quad y_i = 1, 2, 3, \dots, \end{aligned} \quad (5)$$

where y_i is the nonnegative integer value; λ is the i_{th} expected Poisson count; and π is the probability of extra zeros.

2.3. Random-Parameter Multivariate Negative Binomial Model

2.3.1. Random Parameters for Unobserved Heterogeneity. The PRM restricts the mean to be equal to the variance ($E = VAR$), and the PRM model does not fit well in some cases. When the model does not hold the equality, the data may be overdispersed ($E < VAR$) or underdispersed ($E > VAR$), and the standard errors of the estimated parameter of the PRM will be incorrect. To account for overdispersion in the crash count data, PRM is promoted and derived.

$$\lambda_{ik} = e^{(b'x_i + \varepsilon_{ik})}, \quad (6)$$

$$\Pr(Y_i = y_i | \mu_i, t_i) = \frac{e^{-\mu_i t_i} (\mu_i t_i)^{y_i}}{y_i!}, \quad (2)$$

$$\mu_i = t_i \mu(x_i' \beta) = t_i e^{(\beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})},$$

where t is the exposure of a period of time and $\beta_1, \beta_2, \dots, \beta_k$ are coefficients estimated as unknown parameters.

Negative binomial (NB) regression is a common generalization of Poisson regression including a gamma noise variable [43]. This model is very popular because it models the Poisson heterogeneity with a gamma distribution, and the variance is not equal to the mean restrictively. The negative binomial regression model (NBRM) meets the equation

where $e^{\varepsilon_{ik}}$ are error terms following the gamma distribution with mean 1 and variance α .

In response to the nonconstant explanatory variables in the models, we developed the random parameters in each estimated parameter to account for unobserved heterogeneity [39].

$$\beta_{lk} = b_k + \varphi_{lk}, \quad (7)$$

where β_{lk} denotes the l_{th} explanatory variable for observation l ; b_k are the mean parameter estimates; and φ_{lk} is a randomly distributed term capturing unobserved heterogeneity.

2.3.2. Multivariate Negative Binomial Model. In recent research studies, the general framework of the random-parameter multivariate negative binomial (RPMNB) model is mostly proposed with the expected number of crashes [44, 45], λ , in the i_{th} road segment and k_{th} crash types (in this paper, these types are rear-end, bumping-guardrail, and crash):

$$\lambda_{ik} = e^{(\beta_{ik} X_{ik}' + \varepsilon_{ik})}, \quad k = 1, 2, 3, \quad (8)$$

where $X_{ik} = (1, X_{1k}, X_{2k}, \dots, X_{Nk})$ is the independent variable vector; $\beta_{lk} = (\beta_{0k}, \beta_{1k}, \dots, \beta_{Nk})$ is the coefficient vector; and ε_{ik} is the multivariate error term distributed with zero mean and variance θ^2 .

In the RPMNB model, the correlation ρ follows the unstructured correlation covariance matrix, which represents the correlation between ε_{ik} of models for crash type a and b :

$$\mathbf{M} = \begin{bmatrix} \theta_1^2 & \theta_1\theta_2\rho_{12} & \theta_1\theta_2\rho_{13} \\ \theta_2\theta_1\rho_{21} & \theta_2^2 & \theta_2\theta_3\rho_{23} \\ \theta_3\theta_1\rho_{31} & \theta_3\theta_2\rho_{32} & \theta_3^2 \end{bmatrix}. \quad (9)$$

2.4. Model Comparison and Evaluation

2.4.1. Goodness of Fit. When analyzing the structure of the crash data model across different crash types, it is important and necessary to confirm the structure of the unobserved parameters. The likelihood ratio test is used to assess the models [46].

$$\chi^2 = -2[LL(\beta_R) - LL(\beta_U)], \quad (10)$$

where β_R and β_U denote the log-likelihood at the convergence of the restricted and unrestricted model, respectively.

Bayesian information criterion (BIC) is also used for model comparison, which is a generalized version of the Akaike information criterion (AIC) considering the Bayesian equivalent [47]:

$$\begin{aligned} \text{BIC} &= k \ln(n) - 2 \ln(L), \\ \text{AIC} &= 2k - 2 \ln(L), \end{aligned} \quad (11)$$

where k and L denote the number of model parameters and the likelihood function, respectively.

AIC introduces the penalty term to minimize the parameters of the model, which helps to reduce the possibility of overfitting and promote the degree of model fitting (maximum likelihood). BIC considers the number of samples, leading to a larger penalty term than AIC. When the number of samples is too large, BIC can effectively prevent the excessively high complexity caused by the precision of the model. So, the models with smaller AIC and BIC values perform better.

2.4.2. Prediction Accuracy. Other than BIC, we used root mean square error (RMSE) to evaluate the accuracy of the models. Similar to BIC, smaller RMSE values mean the model predicts more accurately.

$$\text{RMSE} = \sqrt{\frac{1}{n_0} \sum_{j=1}^{n_0} (O_j - P_j)^2}. \quad (12)$$

At the same time, the mean absolute error (MAE) and mean absolute percentage error (MAPE) are also used to estimate the accuracy of models:

$$\begin{aligned} \text{MAE} &= \frac{1}{n_0} \sum_{j=1}^{n_0} (O_j - P_j), \\ \text{MAPE} &= \frac{100\%}{n_0} \sum_{j=1}^{n_0} \left| \frac{O_j - P_j}{O_j} \right|, \end{aligned} \quad (13)$$

where O_j is the j_{th} observation value; P_j is the j_{th} predicted value; and n_0 is the number of observations.

3. Data Description

Yearly crash data from the Beijing-Shanghai Highway from Xinyi to Jiangdu, from 2009 to 2011, were collected from the traffic management department. Originally, the data were used to record the emergency vehicle on the highway. The crash data contained 3,293 crashes in detail, including crash type, location, time, vehicle type, climate, road surface condition, and casualty condition. Rear-end and bumping-guardrail crashes made up 52.9% and 30.3% of the crash data, respectively. The rear-end and bumping-guardrail crash type are the focus of this research, but the other crash types (roll-over, fire, scrub, and others) are also utilized in the models to better explore the correlations to the first two crash types. Table 1 summarizes descriptive statistics of the five crash types in each section on the highway.

Besides, roadway geometric data are collected by road design and construction drawings, which include road geometric features. The undivided four-lane highway from Xinyi to Jiangdu in this study is 259.5 km in total, with the design speed of 120 km/h, and includes 27 interchanges and rest areas. We divide the road into 426 different sections according to different horizontal alignment, vertical alignment, and interchange both-way. We obtain the AADT of 426 sections reported by roadway management agencies.

We adopt the modified crash frequency method to identify the crash black spots and calculate the average crash number of each section:

$$\lambda = \frac{\sum m_i}{n}. \quad (14)$$

The critical value of the crash number is determined with the confidence level of 95%:

$$R = \lambda + 1.96\lambda, \quad i = 1, 2, \dots, n, \quad (15)$$

where λ is the average crash number; m_i is the crash number in the i_{th} section; n is the total number of sections; and R is the critical value of the crash number.

The actual crash number of each section is compared with the critical value, and the section is identified as crash black spots if the actual number is greater than R .

Based on the models in the Specifications for Highway Safety Audit published in 2015, we calculated the operating speeds of cars and trucks by segments according to different geometric features [35].

The stopping sight distance means the shortest distance required for one ordinary driver to react and slow down or stop when encountering obstacles while driving at a certain speed. Based on the Guidelines for Design of Highway Grade-separated Intersections [48], the stopping sight distances of the car and truck are equations (16) and (17), respectively.

$$S_{\text{car}} = \frac{v_{85}t}{3.6} + \frac{(v_{85}/3.6)^2}{2gf}. \quad (16)$$

The truck drivers can see the vertical plane of the obstacles at a considerable distance from the point of view with the low speed, but this advantage is not enough to make up

TABLE 1: Descriptive statistics of dependent variables.

Variable names	Definitions	Minimum	Maximum	Mean	SD
Rear-end crash	Rear-end crash occurs when a vehicle strikes the one in front of it	1	27	4.58	3.93
Bumping-guardrail crash	A vehicle strikes the guardrails in the middle of or outside the highways	1	14	3.08	2.23
Other crash	This type includes roll-over (a vehicle tips over onto its side or roof), fire (a vehicle caught fire due to mechanical failure), scrub (crashes caused by side scrub between vehicles), and other crashes	0	14	1.18	1.59
Noncasualty crash	Crashes leading to property damage only	0	42	6.70	6.41
Casualty crash	Crashes resulted in injury, disability, or fatality	0	5	0.39	0.75

for the poor braking performance. Despite the high viewpoint, truck drivers also lose their sight in places with limited lateral line of sight, especially

$$S_{\text{truck}} = \frac{v_{85}t}{3.6} + \frac{(v_{85}/3.6)^2}{2g(f+i)}, \quad (17)$$

where S_{car} and S_{truck} are the stopping sight distance of the car and truck, respectively; v_{85} is the operating speed (km/h); t is the reaction time, takes 2.5 s generally (judging time as 1.5 s and running time as 1.0 s); g is the gravitational acceleration, takes 9.8 m/s²; i is the longitudinal grade; and f is the longitudinal friction coefficient between truck tires and the road surface which takes 0.17 generally.

Corrugated beam guardrails are set in the middle and beside the road across all sections, which affect the drivers' sight. We consider the biggest transverse clear distance to confirm the sight safety, which means the distance between the curve of sight and the track. When the plane curve is sharp, the transverse clear distance should be judged on the inside lane. We calculate the required stopping sight distance of each section for safety following the equation

$$H = R_s \left(1 - \cos \frac{\gamma}{2} \right), \quad (18)$$

where H is the biggest transverse clear distance; R_s is the radius of the inside lane; and γ is the central angle of line of sight.

As shown in Table 2, we summarize the traffic, speed, geometric, and sight characteristics as the independent variables.

4. Results and Discussion

4.1. Model Specification. Estimation of traditional and proposed models is involved in the empirical analysis: (1) traditional model: independent (separate model for 5 different crash types) NB model and independent Poisson model are adopted as bases. Meanwhile, two models are estimated including the zero-inflated multivariate NB (ZIMNB) model and zero-inflated multivariate Poisson (ZIMP) model considering the excess zero-count data; (2) proposed model: random-parameter multivariate NB (RPMNB) model and random-parameter multivariate Poisson (RPMP) model are proposed to estimate. The results of the proposed models are shown in Tables 3 and 4, and Appendix provides the results of corresponding traditional models.

Subsequently, each base effect is estimated for common exogenous variables across five crash types, and we estimate the deviation of variables versus the base for each crash type. The corresponding t -statistic will present statistically significant if the deviation term offers a difference from the base effect. Based on the t -statistic, the parameter does not reveal differential sensitivity for the base crash type if the variable is statistically insignificant.

4.2. Model Estimation Results. In this section, we conduct a detailed discussion of the significant factors affecting crash count components on different crash types. The model estimation results for the RPMP model and RPMNB model are shown in Tables 3 and 4, respectively.

Similar to the traditional approach, we presented the individual effect of each exogenous variable accommodating to the crash propensity. Taken the constants estimated in various crash type propensity equations as an example, the effect of the casualty crash in Table 3 can be computed as 1.867 (base of the rear-end crash) \pm 1.000 (casualty crash deviation) = 0.867. At the same time, we identify the number of significant parameters estimated across the five crash types and estimate one individual parameter across 5 crash types. The positive value of the variable in Table 3 indicates there will be more crashes with the increase of the variable, and less crashes, otherwise. As shown in Tables 3 and 4, the significant factors in the two models are not exactly the same. Then, we focus on the results of the RPMNB model while analyzing the effects of factors, and the differences in the RPMP model are appended.

4.2.1. Constant. The constant represents the intercept of the crash type with exogenous variables, without any substantive interpretation.

4.2.2. Traffic Characteristics. Whether there is an interchange in the section is found to negatively influence other, noncasualty, and casualty crashes indicating a lower likelihood of crash propensity for these three crash types in interchange sections. Compared with the general sections, more vehicles accelerate, decelerate, and change lanes leading to more interweaving areas in the interchange sections, getting more attention from drivers and causing fewer crashes to some extent. The results are consistent with the earlier research on the highway interchange [48]. And the impact is also found not significantly different for rear-

TABLE 2: Summary statistics of independent variables.

Variable names	Definition	Min.	Max.	Mean	SD
<i>Traffic characteristics</i>					
Interchange	1, interchange in the section (25.8%); 0, otherwise (74.2%)	0.00	1.00	0.25	0.44
Black spots	1, the crash number is higher than 18 (7.0%); 0, otherwise (93.0%)	0.00	1.00	0.07	0.26
AADT	Average annual daily traffic volume	31158	68836	46685.94	11969.59
<i>Speed characteristics</i>					
V_{O-car}	Operating speed of cars	88.92	134.37	121.91	21.70
ΔV_{O-car}	Speed difference of cars with the adjacent segment	-85.93	78.03	0.00	34.07
$V_{O-truck}$	Operating speed of trucks	60.38	104.86	78.80	12.34
$\Delta V_{O-truck}$	Speed difference of trucks with the adjacent segment	-34.01	25.72	0.01	19.89
ΔV_O	Speed difference between cars and trucks	12.64	73.99	43.11	16.99
<i>Geometric characteristics</i>					
R_{front}	Radius of the plane curve of the front section	5597.0	1000000.0	370713.6	478894.40
L_{front}	Length of the plane curve of the front section	45.0	3677.0	1218.48	703.48
$R_{present}$	Radius of the present plane curve	5597.0	1000000.0	370717.10	478891.70
$L_{present}$	Length of the present plane curve	45.0	3677.0	1218.48	703.48
R_{back}	Radius of the plane curve of the back section	5597.0	1000000.0	373043.70	479536.80
L_{back}	Length of the plane curve of the back section	45.0	3677.0	1218.48	703.48
i_{min}	Minimum longitudinal grade of the current section	-1.63	1.63	0.00	0.47
Ls_{min}	Length of the longitudinal slope corresponding to the minimum grade	240.00	1740.00	773.33	296.03
i_{max}	Maximum longitudinal grade of the current section	-2.50	2.50	0.00	0.97
Ls_{max}	Length of the longitudinal slope corresponding to the maximum grade	362.00	1740.00	652.63	248.00
<i>Sight characteristics</i>					
S_{car}	Stopping sight distance of cars	244.00	1009.90	439.06	142.23
S_{truck}	Stopping sight distance of trucks	52.00	279.50	87.48	29.92
H_{car}	Transverse clear distance of cars	0.02	6.59	1.66	1.71
H_{truck}	Transverse clear distance of trucks	0.00	0.47	0.06	0.06

TABLE 3: Random-parameter multivariate Poisson model estimation results.

Variables ¹	No. ²	Rear-end estimate (<i>t</i> -stat)	Bumping-guardrail estimate (<i>t</i> -stat)	Other estimate (<i>t</i> -stat)	Noncasualty estimate (<i>t</i> -stat)	Casualty estimate (<i>t</i> -stat)
<i>Constant</i>	5	0.0432 (0.27)	0.501 (3.17)	-0.527 (-2.11)	-0.0332 (-0.21)	-0.430 (-1.13)
<i>Traffic characteristics</i>						
Interchange	3	— ³	—	-0.422 (-2.91)	-0.164 (-1.95)	-0.601 (-2.55)
Black spots	5	0.0986 (1.24)	0.859 (9.77)	1.276 (7.17)	1.196 (9.54)	1.322 (4.88)
AADT	4	1.057 (11.06)	0.0000936 (3.00)	0.0000126 (2.52)	0.0000365 (11.55)	—
<i>Speed characteristics</i>						
V_{O-car}	5	-0.0131 (-3.63)	-0.00812 (-2.38)	-0.0140 (-2.63)	-0.0177 (-4.58)	-0.016 (-1.85)
ΔV_{O-car}	5	0.0114 (3.31)	0.00828 (2.45)	0.0109 (2.09)	0.0119 (3.25)	0.0237 (2.45)
$V_{O-truck}$	2	0.0210 (3.14)	—	—	-0.0272 (3.82)	—
$\Delta V_{O-truck}$	4	-0.0190 (-4.09)	-0.00148 (0.22)	-0.0182 (-2.67)	-0.0221 (-4.40)	—
<i>Geometric characteristics</i>						
L_{front}	1	—	—	—	—	0.000251 (2.13)
$R_{present}$	2	0.000000361 (3.69)	—	—	0.000000375 (3.63)	—
$L_{present}$	5	0.000588 (10.32)	0.000468 (8.12)	0.000755 (8.09)	0.000753 (11.64)	0.000748 (6.25)
L_{back}	1	—	-0.000133 (-2.48)	—	—	—
Ls_{min}	2	-0.000357 (-2.51)	—	—	-0.000423 (-2.83)	—
<i>Sight characteristics</i>						
S_{truck}	5	-0.00540 (-2.78)	-0.00633 (-2.86)	-0.0111 (-3.20)	-0.00706 (-3.44)	-0.0135 (-2.51)
H_{car}	3	-0.0860 (-2.13)	-0.105 (-2.57)	—	-0.0878 (-2.06)	—

Note: variables not shown here attribute to insignificant. ¹Variable definitions and units can be seen in Table 2. ²Number of parameters estimated. ³Insignificant at the 95% significance level. Total number of parameters: 52; log-likelihood: -14353.079; AIC: 28996.159; BIC: 29567.814; χ^2 : 21264.957.

TABLE 4: Random-parameter multivariate negative binomial model estimation results.

Variables ¹	No. ²	Rear-end estimate (<i>t</i> -stat)	Bumping-guardrail estimate (<i>t</i> -stat)	Other estimate (<i>t</i> -stat)	Noncasualty estimate (<i>t</i> -stat)	Casualty estimate (<i>t</i> -stat)
<i>Constant</i>	5	1.867 (1.77)	15.393 (15.21)	11.763 (0.02)	-0.604 (-3.49)	0.867 (0.85)
<i>Traffic characteristics</i>						
Interchange	3	— ³	—	-0.422 (-2.91)	-0.165 (-2.05)	-0.636 (-2.58)
Black spots	5	1.048 (11.25)	0.859 (9.77)	1.276 (7.17)	1.052 (12.37)	1.274 (4.97)
AADT	5	0.0000254 (7.90)	0.00000936 (3.00)	0.0000126 (2.52)	0.0000322 (11.11)	-0.0000128 (-1.63)
<i>Speed characteristics</i>						
V_{O-car}	5	-0.0120 (-3.37)	-0.00795 (-2.32)	-0.0161 (-3.10)	-0.0180 (-5.34)	-0.0165 (-1.89)
ΔV_{O-car}	5	0.00999 (2.76)	0.00788 (2.29)	0.0120 (2.22)	0.0103 (3.07)	0.0231 (2.30)
$V_{O-truck}$	2	0.0200 (2.92)	—	—	-0.0326 (4.93)	—
$\Delta V_{O-truck}$	4	-0.0189 (-4.02)	-0.00948 (-2.12)	-0.020 (-3.18)	-0.0243 (-5.44)	—
<i>Geometric characteristics</i>						
R_{front}	2	0.000000173 (1.98)	—	—	0.00000017 (1.91)	—
L_{front}	1	—	—	—	—	0.000278 (2.44)
$R_{present}$	2	0.000000359 (3.72)	—	—	0.000000365 (3.75)	—
$L_{present}$	5	0.000557 (9.86)	0.000468 (8.12)	0.000672 (7.32)	0.000685 (11.07)	0.000747 (6.44)
L_{back}	1	—	-0.000133 (-2.48)	—	—	—
LS_{min}	3	-0.000319 (-2.34)	—	—	-0.000453 (-3.19)	-0.000608 (-1.85)
LS_{max}	1	—	—	—	-0.000420 (-2.17)	—
<i>Sight characteristics</i>						
S_{truck}	5	-0.00493 (-2.29)	-0.0059 (-2.61)	-0.0107 (-3.01)	-0.00706 (-3.44)	-0.0135 (-2.51)
H_{car}	2	—	-0.100 (-2.40)	—	-0.0878 (-2.06)	—

Total number of parameters: 56; log-likelihood: -14326.558; AIC: 28985.114; BIC: 29610.316; χ^2 :23043.735.

end and bumping-guardrail crashes, which shows that these two crash types are not associated with the interchange areas.

The corresponding parameter of black spots offers a positive impact on crash occurrence for rear-end, bumping-guardrail, other, noncasualty, and casualty crashes revealing a higher likelihood of crashes with the increased proportion of black spots. More crashes across all the crash types occurred in the black spots than other sections. Similarly, the AADT offers a positive influence across the five crash types in the RPMNB model indicating more crashes occur with more traffic. Greater traffic volume results in higher crash risk [49]. Then, the AADT indicates no significance on the casualty crashes in the RPMP model.

4.2.3. Speed Characteristics. In terms of operating speed, the estimated results of both RPMP model and RPMNB model show that the higher operating speed of cars is likely to result in less crash risk across five crash types. The finding is expected because drivers of cars are likely to drive at a higher speed under good traffic conditions in the sections with great geometric alignment.

The estimated results show that the speed difference of cars with the adjacent segment shows the same positive effect on all the crash types. The higher speed difference of cars between adjacent segments leads to higher crash risk, and this finding is corresponding to the Solomon curve [18]. Oppositely, the speed difference of trucks with the adjacent segment shows the same negative effect on all the crash types except the casualty crash.

The result of trucks is contrary, and the estimated variable of the speed difference of trucks shows a negative influence on the rear-end, bumping-guardrail, other, and noncasualty crashes and an insignificant effect on the casualty crashes. The reasonable explanation is that the poor power and braking performance of trucks limit the operating speed. When the design speed reaches a certain value, the operating speed will not increase with the improvement of the geometric alignment. And the operating speed restricted by the poor geometric alignment of the sections reduces the crash risk. The operating speed of trucks has a positive influence on the risk of rear-end crashes and a negative influence on the noncasualty crashes. The higher operating speed of trucks leads to more rear-end crashes and less casualty in the crashes.

Relative to the cars, the trucks show poor power and braking performance. The greater operating speed results in a higher propensity of rear-end crashes and a smaller difference between cars and trucks, which lead to less collision impact and damage in the crashes.

4.2.4. Geometric Characteristics. The length of the present plane curve presents a positive effect on the five crash types, which is perhaps indicating that the longer length of the plane curve results in more crash risk propensity. The reasonable explanation for this result is that the drivers will adapt to the curvature of the curve gradually, causing the fatigue or distraction of drivers. Otherwise, this adaption causes the drivers to accelerate to the speed not suitable for the geometric alignment, leading to more crashes. The

positive effect of the radius of the present plane curve on the rear-end and casualty crashes shows the same tendency.

Concerning the radius of the plane curve of the front section, the variable is found to have a positive influence on the rear-end and noncasualty crashes with relatively little influence coefficients (0.000000359 and 0.000000365) in the RPMNB model. However, the estimated results are not found in the RPMP model. The vehicles will run at an uncoordinated speed when transiting to a smaller radius and increasing the rear-end crash propensity. Similarly, the coefficient of the length of the plane curve of the front section shows a positive influence on the casualty crashes. The length of the longitudinal slope corresponding to the minimum grade is also found to have a negative influence on the probability of rear-end, noncasualty, and casualty crashes in the RPMNB model. And in the RPMP model, only the rear-end and noncasualty crashes show significance. Among the road sections with the small longitudinal grade, the trucks show greater speed with less speed difference from cars, which promotes safety. In the RPMNB model, the length of the longitudinal slope corresponding to the maximum grade also presents a negative influence on the noncasualty crashes. However, the results are not present in the RPMP model.

4.2.5. Visibility Characteristics. As expected, the coefficient associated with the stopping sight distance of trucks offers a negative effect on the crash risk of all the crash types. The likelihood of all single-vehicle crashes will reduce with higher stopping sight distance of trucks.

The visibility field and distance in front of the vehicle are important for safe and effective driving on the highway. The speed and direction of the vehicles depend on the visibility of the ahead road and surrounding environment. Therefore, the higher stopping sight distance of the truck facilitates the drivers to control the direction more accurately, leading to fewer crashes. Meanwhile, the transverse clear distance of cars indicates negative effects on the rear-end, bumping-guardrail, and noncasualty crashes in the RPMP model. And the transverse clear distance of cars only shows a negative influence on bumping-guardrail and noncasualty crashes in the RPMNB model. The more transverse safe zone allows for a higher tolerance for vehicles, which leads to fewer crashes.

4.3. Model Comparisons. Statistical measures such as log-likelihood, AIC, and BIC values are calculated to measure the goodness of fit in the estimated models, and the results are shown in Table 5.

Based on the comparison of AIC and BIC values, we firstly find the models with zero-inflated effects are not suitable for the crash data in this study. Secondly, the models considering effects caused by unobserved heterogeneity perform better than the independent models. Finally, the proposed models estimate efficient and accurate parameters in the parsimonious model systems. To assess the predictive accuracy of the estimated models, we used RMSE (root mean square error), MAE (mean absolute error), and MAPE

(mean absolute percentage error) for discussion. Table 6 presents the values of these measures.

Despite the difference in the total number of parameters between the two models (56 vs. 52), the performance of the two models across five crash types is quite similar. From the total values of the three measures, we can observe that the RPMNB model performs better than the RPMP model. From the perspective of different crash types, the RPMP model performs marginally better than the RPMNB model for the measures with respect to bumping-guardrail and noncasualty crashes, and RPMNB performs better in terms of rear-end, other, noncasualty, and casualty crashes. The difference in the number of parameters across the two models makes no effects on the deviation measures.

4.4. Elasticity Effects. To provide more insight and explain the marginal effects of the exogenous variables, the elasticity effect is computed for RPMNB and RPMP across all the crash types. Elasticity effect denotes an estimate of the effect of a variable on the expected frequency assuming all the other variables take the average values. The elasticity effect is the effect on the expected frequency λ_i of a 1% change in the variable, which is as follows:

$$E_{x_{ik}}^{\lambda_i} = \frac{\partial \lambda_i}{\lambda_i} \times \frac{x_{ik}}{\partial x_{ik}} = \beta_k x_{ik}, \quad (19)$$

where x_{ik} is the value of the k_{th} independent variable for observation i ; β_k is the estimated parameter for the k_{th} independent variable; and λ_i is the expected frequency for observation i .

As shown in Figure 1, there is not any large difference in the elasticity effects of the two models across five crash types. And almost half of the number of the variables make very little effects (41 out of 79).

We observe that there are significant differences in the elasticity effects for different crash types across different models. With respect to the length of the present plane curve for the rear-end crashes, the RPMNB model predicts a 27.4% increase, while a 52.7% increase from the RPMP. Most likely, the nonlinearity of the two models results in such differences.

5. Conclusions

The empirical analysis was based on the traffic crash data from the Beijing-Shanghai Highway for the year 2009–2011 and the road design and construction drawings including the road geometric features. The road geometric alignment, operating speed, traffic volume, and sight distance are considered as the exogenous variables across the traditional and proposed models.

The traditional negative binomial and Poisson models are commonly adopted in the analysis on crash count frequency to analyze the impacts of various factors. To reduce the influence of the unobserved heterogeneity and get higher dimensions, we adopt the random-parameter multivariate NB (RPMNB) model and random-parameter multivariate Poisson (RPMP) model to analyze the impacts on different

TABLE 5: Goodness-of-fit measure of six models.

Model	Total no.	Log-likelihood	AIC	BIC
Independent NB model	55	-14,352.787	29,007.581	29,602.802
Independent Poisson model	61	-16,595.848	33,441.697	33,934.501
ZIMNB model	53	-14,352.888	29,131.770	29,674.168
ZIMP model	53	-16,194.117	32,662.229	33,203.686
RPMNB model	52	-14,326.558	28,985.114	29,610.316
RPMP model	56	-14,353.079	28,996.159	29,567.814

TABLE 6: Predictive performance measure of two models (RPMNB and RPMP).

Crash type	RMSE		MAE		MAPE	
	RPMNB	RPMP	RPMNB	RPMP	RPMNB	RPMP
Rear-end	7.413	8.374	6.551	6.941	1.433	1.638
Bumping-guardrail	11.070	10.833*	9.781	9.551*	3.639	2.919*
Other crash	12.222	12.983	10.661	11.051	5.349	5.989
Noncasualty crash	1.643	1.597*	1.253	1.252*	0.136	0.128*
Casualty crash	14.351	15.105	12.811	13.298	10.372	11.116
Total	46.699	48.892	41.055	42.092	20.929	21.789

Note: RPMNB = random-parameter multivariate negative binomial model. RPMP = random-parameter multivariate Poisson model. *Measures where the RPMP model performs better.

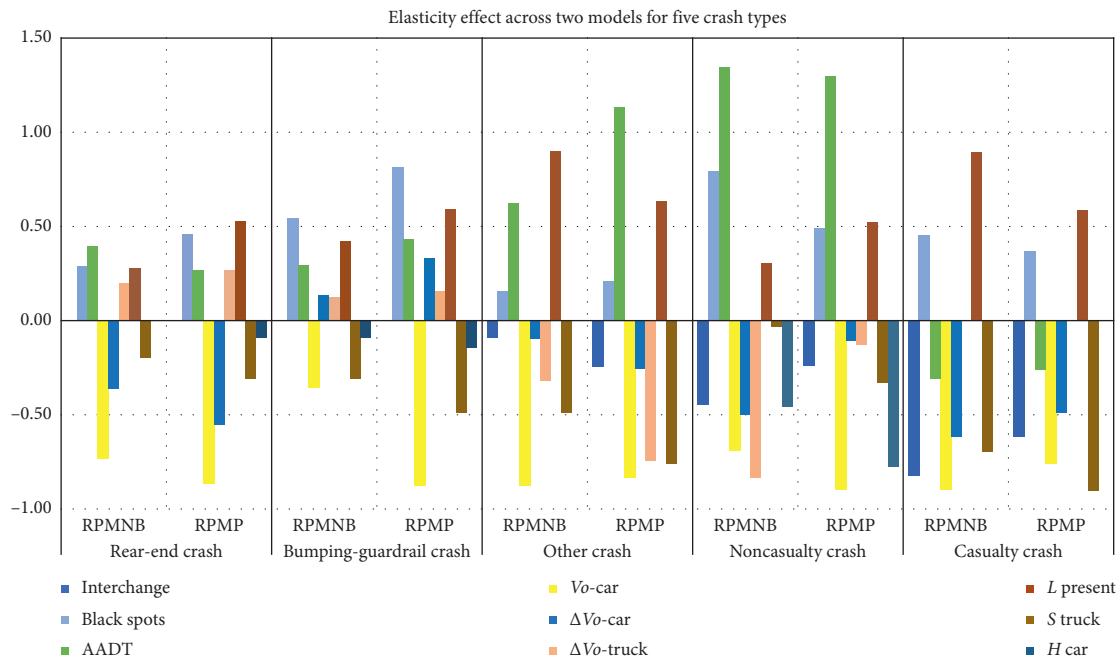


FIGURE 1: Elasticity effects across two models (RPMNB and RPMP) for 5 crash types.

crash types including rear-end, bumping-guardrails, non-casualty, and casualty crashes, with the traditional models as the base.

Viewing on the traffic characteristic, more rear-end, bumping-guardrail, other, and noncasualty crashes occurred with more black spots and AADT, while other, noncasualty, and casualty crashes happened in interchange sections less likely. For speed characteristics, the crash risk of five types is greater with the higher operating speed of cars, while greater speed difference of cars results in more crashes across five types. Oppositely, the greater speed difference of trucks

results in less rear-end, bumping-guardrail, other, and noncasualty crashes. As for the road geometry, more crashes occurred on the longer length of the present plane curve, while the greater length of the longitudinal slope corresponding to the minimum grade reduces the crash propensity. Considering the sight characteristics, truck drivers with greater stopping sight distance perform safer on the highways.

From the perspective of road design, to avoid the long length of a plane curve with big curvature and reduce the length of the longitudinal slope with large grade are

TABLE 7: Independent negative binomial model estimation results.

Variables ¹	No. ²	Rear-end estimate (<i>t</i> -stat)	Bumping-guardrail estimate (<i>t</i> -stat)	Other estimate (<i>t</i> -stat)	Noncasualty estimate (<i>t</i> -stat)	Casualty estimate (<i>t</i> -stat)
<i>Constant</i>	5	0.043 (0.27)	0.502 (3.07)	-0.528 (-2.11)	-0.033 (-0.21)	-0.431 (-1.13)
<i>Traffic characteristics</i>						
Interchange	3	— ³	—	-0.422 (-2.91)	-0.165 (-1.95)	-0.601 (-2.55)
Black spots	5	1.057 (11.06)	0.859 (9.77)	1.276 (7.17)	1.196 (9.54)	1.322 (4.88)
AADT		0.00003 (8.66)	0.00001 (3.00)	0.00001 (2.52)	0.00004 (11.55)	—
<i>Speed characteristics</i>						
V_{O-car}	5	-0.013 (-3.63)	-0.008 (-2.38)	-0.014 (-2.63)	-0.017 (-4.58)	-0.016 (-1.85)
ΔV_{O-car}	5	0.011 (3.31)	0.008 (2.45)	0.010 (2.09)	0.011 (3.25)	0.023 (0.014)
$V_{O-truck}$	2	0.021 (3.14)	—	—	0.027 (3.82)	—
$\Delta V_{O-truck}$	4	-0.019 (-4.09)	-0.009 (-2.11)	-0.018 (-2.67)	-0.022 (-4.40)	—
<i>Geometric characteristics</i>						
L_{front}	1	—	—	—	—	0.000251 (2.13)
$R_{present}$	2	0.00000036 (3.63)	—	—	0.000000379 (3.60)	—
$L_{present}$	5	0.0005888 (10.28)	0.0004658 (8.04)	0.0007505 (8.02)	0.0007521 (11.57)	0.000750 (6.23)
L_{back}	1	—	-0.0001341 (-2.49)	—	—	—
LS_{min}	4	-0.000356 (-2.5)	—	-0.0004069 (-1.88)	-0.0004248 (-2.83)	-0.000586 (-1.76)
<i>Sight characteristics</i>						
S_{truck}	5	-0.054 (-2.78)	-0.006 (-2.86)	-0.011 (-3.20)	-0.007 (-3.44)	-0.013 (-2.51)
H_{car}	3	-0.086 (-2.13)	-0.105 (-2.57)	—	-0.087 (-2.06)	—

Note: ¹variable definitions and units can be seen in Table 2. Variables not shown here attribute to insignificant. ²Number of parameters estimated. ³Insignificant at the 95% significance level. Total number of parameters: 55; log-likelihood: -14352.787; AIC: 29007.581; BIC: 29602.802; χ^2 :21274.146.

TABLE 8: Independent Poisson model estimation results.

Variables ¹	No. ²	Rear-end estimate (<i>t</i> -stat)	Bumping-guardrail estimate (<i>t</i> -stat)	Other estimate (<i>t</i> -stat)	Noncasualty estimate (<i>t</i> -stat)	Casualty estimate (<i>t</i> -stat)
<i>Constant</i>	5	0.177 (1.41)	0.515 (3.27)	-0.460 (-2.31)	0.228 (2.51)	-0.348 (-1.07)
<i>Traffic characteristics</i>						
Interchange	3	— ³	—	-0.3482 (-3.01)	-0.088 (-1.93)	-0.529 (-2.54)
Black spots	5	1.020 (16.68)	0.859 (10.40)	1.227 (10.68)	1.083 (23.20)	1.316 (6.10)
AADT	5	0.0000236 (10.04)	0.00000909 (3.03)	0.000011 (2.77)	0.0000312 (18.14)	-0.000014 (0.047)
<i>Speed characteristics</i>						
V_{O-car}	5	-0.0134 (-5.96)	-0.00807 (-2.81)	-0.0144 (-3.73)	-0.0181 (-10.98)	-0.0154 (-2.05)
V_{O-car}	5	0.0119 (5.27)	0.00829 (2.85)	0.0112 (2.80)	0.0124 (7.53)	0.0219 (2.70)
$V_{O-truck}$	2	0.0209 (4.86)	—	—	0.0285 (8.90)	—
$\Delta V_{O-truck}$	4	-0.0196 (-6.81)	-0.00939 (-2.54)	-0.0190 (-3.90)	-0.0234 (-11.05)	—
<i>Geometric characteristics</i>						
R_{front}	2	0.000000154 (2.54)	—	—	0.000000148 (3.36)	—
L_{front}	3	—	—	—	0.0000506 (1.94)	0.000272 (2.62)
$R_{present}$	2	0.000000410 (6.12)	—	—	0.000000375 (7.71)	—
$L_{present}$	5	0.000567 (15.62)	0.000460 (9.02)	0.000659 (10.59)	0.000630 (24.12)	0.00075 (7.16)
L_{back}	1	—	-0.000131 (-2.71)	—	—	—
i_{min}	1	—	—	—	-0.108 (-2.44)	—
LS_{min}	4	-0.0003483 (-3.67)	—	-0.0004866 (-2.91)	-0.000437 (-6.18)	-0.000644 (-2.12)
<i>Sight characteristics</i>						
S_{truck}	5	-0.00644 (-4.78)	-0.00641 (-3.35)	-0.0118 (-4.28)	-0.00847 (-8.14)	-0.0142 (-3.03)
H_{car}	4	-0.104 (-3.93)	-0.106 (-3.06)	-0.110 (-2.37)	-0.1232 (-6.11)	—

Total number of parameters: 61; log-likelihood: -16595.848; AIC: 33441.697; BIC: 33934.501; χ^2 :50507.432.

beneficial for highway safety. To reduce the interference between cars and trucks, it is suggested to adopt the lane distribution mode of different vehicle types. Set up door-frame traffic signs along with pavement markings to remind the drivers of trucks to drive in the outside lane and the ones of cars to drive in the inside lane.

With respect to the data fit, the comparison exercise including log-likelihood, AIC, and BIC value highlights the superiority of the multivariate models over the traditional models. Based on the exercise, the RPMNB model shows identical performance even with more numbers of significant parameters compared with the RPMP model (56 vs. 52).

TABLE 9: Zero-inflated multivariate negative binomial model estimation results.

Variables ¹	No. ²	Rear-end estimate (<i>t</i> -stat)	Bumping-guardrail estimate (<i>t</i> -stat)	Other estimate (<i>t</i> -stat)	Noncasualty estimate (<i>t</i> -stat)	Casualty estimate (<i>t</i> -stat)
<i>Constant</i>	5	0.0433 (0.27)	0.501 (3.07)	-0.527 (-2.11)	-0.0191 (-0.12)	-0.430 (-1.13)
<i>Traffic characteristics</i>						
Interchange	3	— ³	—	-0.422 (-2.91)	-0.164 (-1.95)	-0.601 (-2.55)
Black spots	5	1.057 (11.06)	0.859 (9.77)	1.276 (7.17)	1.196 (9.54)	1.322 (4.88)
AADT	4	0.0000263 (8.66)	0.00000936 (3.00)	0.0000126 (2.52)	0.0000365 (11.55)	—
<i>Speed characteristics</i>						
V_{O-car}	5	-0.0131 (-3.63)	-0.00812 (-2.38)	-0.0140 (-2.63)	-0.0177 (-4.58)	-0.0161 (-1.85)
ΔV_{O-car}	5	0.0114 (3.31)	0.00828 (2.45)	0.0109 (2.09)	0.0119 (3.25)	0.0237 (2.45)
$V_{O-truck}$	2	0.0210 (3.14)	—	—	0.0272 (3.82)	—
$\Delta V_{O-truck}$	4	-0.0190 (-4.09)	-0.00934 (-2.11)	-0.0182 (-2.67)	-0.0221 (-4.40)	—
<i>Geometric characteristics</i>						
L_{front}	1	—	—	—	—	0.000251 (2.13)
$R_{present}$	2	0.000000361 (3.69)	—	—	0.000000375 (3.63)	—
$L_{present}$	5	0.000588 (10.32)	0.000468 (8.12)	0.000755 (8.09)	0.000753 (11.64)	0.000748 (6.25)
L_{back}	1	—	-0.000133 (-2.48)	—	—	—
LS_{min}	3	-0.000357 (-2.51)	—	—	-0.000423 (-2.83)	-0.000588 (-1.76)
<i>Sight characteristics</i>						
S_{truck}	5	-0.00540 (-2.78)	-0.00633 (-2.86)	-0.0111 (-3.20)	-0.00706 (-3.44)	-0.0135 (-2.51)
H_{car}	3	-0.0860 (-2.13)	-0.105 (-2.57)	—	-0.0878 (-2.06)	—

Total number of parameters: 53; log-likelihood: -14352.888; AIC: 29131.777; BIC: 29674.168; χ^2 :19426.561.

TABLE 10: Zero-inflated multivariate Poisson model estimation results.

Variables ¹	No. ²	Rear-end estimate (<i>t</i> -stat)	Bumping-guardrail estimate (<i>t</i> -stat)	Other estimate (<i>t</i> -stat)	Noncasualty estimate (<i>t</i> -stat)	Casualty estimate (<i>t</i> -stat)
<i>Constant</i>	5	0.177 (1.41)	0.515 (3.27)	-0.228 (-1.02)	0.426 (4.48)	-0.0568 (-0.16)
<i>Traffic characteristics</i>						
Interchange	2	— ³	—	-0.314 (-2.48)	—	-0.552 (-2.46)
Black spots	5	1.020 (16.68)	0.859 (10.40)	1.004 (7.92)	1.050 (22.56)	1.230 (5.18)
AADT	4	0.0000236 (10.04)	0.00000909 (3.03)	0.0000114 (2.62)	0.0000282 (15.83)	—
<i>Speed characteristics</i>						
V_{O-car}	4	-0.0134 (-5.96)	-0.00807 (-2.81)	-0.00999 (-2.21)	-0.0155 (-9.21)	—
ΔV_{O-car}	4	0.0119 (5.27)	0.00829 (2.85)	—	0.0113 (6.76)	0.0235 (2.48)
$V_{O-truck}$	2	0.0209 (4.86)	—	—	0.0228 (6.93)	—
$\Delta V_{O-truck}$	4	-0.0196 (-6.81)	-0.00939 (-2.54)	-0.0142 (-2.62)	-0.202 (-9.33)	—
<i>Geometric characteristics</i>						
R_{front}	2	0.000000154 (2.54)	—	—	0.000000110 (2.45)	—
L_{front}	1	—	—	—	—	0.000239 (2.13)
$R_{present}$	2	0.00000041 (6.12)	—	—	0.000000336 (6.75)	—
$L_{present}$	5	0.000567 (15.62)	0.000460 (9.02)	0.000606 (8.71)	0.000595 (21.66)	0.000735 (6.54)
R_{back}	1	—	—	—	0.000000093 (2.08)	—
L_{back}	1	—	-0.000131 (-2.71)	—	—	—
LS_{min}	3	-0.000348 (-3.67)	—	-0.000545 (-3.07)	-0.000423 (-2.83)	—
<i>Sight characteristics</i>						
S_{truck}	5	-0.00644 (-4.78)	-0.00641 (-3.35)	-0.0106 (-3.51)	-0.00696 (-6.65)	-0.0137 (-2.61)
H_{car}	3	-0.104 (-3.93)	-0.106 (-3.06)	—	-0.107 (-5.39)	—

Total number of parameters: 53; log-likelihood: -16194.117; AIC: 32662.229; BIC: 33203.686; χ^2 :41460.088.

To quantify the predictive performance measure, we calculate the elasticity effects of significant variables including interchange, black spots, AADT, V_{O-car} , ΔV_{O-car} , $\Delta V_{O-truck}$, $L_{present}$, S_{truck} , and H_{car} for both RPMNB and RPMP models. Significant differences in the elasticity effects exist for different crash types across the two models, which are attributed to the nonlinearity of the two models.

Really, more efforts can be done on further research. With the respect to the variables, more detailed information can be added to reduce the parameter explosion, such as the proportion of trucks in each section and distinguishing the rear-end crash between different vehicle types. Otherwise, the spatial and temporal terms of the crashes are beneficial for the models to reduce the potential unobserved effects,

which can be considered in the future. Furthermore, the other methodological approaches such as latent-class, finite-mixture, and Markov switching models can be used to confirm the finding of this study.

Appendix

The estimation results of the independent negative binomial (INB) model, independent Poisson (IP) model, zero-inflated multivariate negative binomial (ZIMNB) model, and zero-inflated multivariate Poisson (ZIMP) model are shown in Tables 7–10. And the goodness-of-fit measure of the four models is shown in Table 5.

Based on the comparison in Table 5, the ZIMNB model has higher AIC and BIC values than the independent NB model and RPMNB model. The length of the longitudinal slope corresponding to the minimum grade indicates a negative influence on the casualty crashes in the ZIMNB model, while the same result is not shown in the RPMNB model.

As shown in Table 8, the number of parameters estimated in the ZIMP model is 53, which is the least in the three Poisson models. For example, the common significant factors such as interchange show no significance on the noncasualty crashes in the ZIMP model. However, the bigger AIC and BIC values indicate that the ZIMP model performs worse than the RPMP model, even with fewer parameters.

Data Availability

Yearly crash data and roadway geometric data from the Beijing-Shanghai Highway from Xinyi to Jiangdu, from 2009 to 2011, were collected from the traffic management department in this article.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

All authors reviewed the results and approved the final version of the manuscript.

Acknowledgments

This study was supported by the Project of the National Natural Science Foundation of China (Grant nos. 51768063 and 51868068).

References

- [1] M. A. Rupp, M. D. Gentzler, and J. A. Smither, "Driving under the influence of distraction: examining dissociations between risk perception and engagement in distracted driving," *Accident Analysis & Prevention*, vol. 97, pp. 220–230, 2016.
- [2] M. A. Regan, C. Hallett, and C. P. Gordon, "Driver distraction and driver inattention: definition, relationship and taxonomy," *Accident Analysis and Prevention*, vol. 43, no. 5, pp. 1771–1781, 2011.

- [3] T. Horberry, J. Anderson, M. A. Regan, T. J. Triggs, and J. Brown, "Driver distraction: the effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance," *Accident Analysis & Prevention*, vol. 38, no. 1, pp. 185–191, 2006.
- [4] T. J. Babukov, *Road Condition and Traffic Safety*, Tongji University Press, Shanghai, China, 1990.
- [5] C. V. Zegeer, D. Reinfurt, T. Neuman et al., "Safety improvements on horizontal curves for two-lane rural roads-informational guide (FHWA-RD-90-074)," *Curves*, 1991.
- [6] C. V. Zegeer, J. R. Stewart, F. M. Council, D. W. Reinfurt, and E. Hamilton, "Safety effects of geometric improvements on horizontal curves," *Transportation Research Record*, vol. 1356, 1992.
- [7] J. McFadden and L. Elefteriadou, "Evaluating horizontal alignment design consistency of two-lane rural highways: development of new procedure," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1737, no. 1, pp. 9–17, 2000.
- [8] H. W. McGee, W. E. Hughes, and K. Daily, *Effect of Highway Standards on Safety*, Transportation Research Board, Washington, DC, USA, 1995.
- [9] Y. L. Pei and J. Ma, "Research on countermeasures for road condition causes of traffic accidents," *China Journal of Highways*, vol. 16, no. 4, pp. 77–82, 2003.
- [10] Y. Hassan, S. M. Easa, and A. El Halim, "Analytical model for sight distance analysis on three-dimensional highway alignments," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1523, no. 1, pp. 1–10, 1996.
- [11] S. M. Easa, T. R. Strauss, Y. Hassan, and R. R. Souleyrette, "Three-dimensional transportation analysis: planning and design," *Journal of Transportation Engineering*, vol. 128, no. 3, pp. 250–258, 2002.
- [12] S. M. Easa, Y. Hassan, and Z. A. Karim, "Establishing highway vertical alignment using profile field data," *ITE Journal*, vol. 68, no. 8, pp. 81–86, 1998.
- [13] Y. Hassan, S. M. Easa, and A. O. Abd El Halim, "State-of-the-art of three-dimensional highway geometric design," *Canadian Journal of Civil Engineering*, vol. 25, no. 3, pp. 500–511, 1998.
- [14] M. Hasan, T. Sayed, and Y. Hassan, "Influence of vertical alignment on horizontal curve perception: effect of spirals and position of vertical curve," *Canadian Journal of Civil Engineering*, vol. 32, no. 1, pp. 204–212, 2005.
- [15] J. A. Reagan, "The interactive highway safety design model: designing for safety by analyzing road geometrics," *Public Roads*, vol. 58, no. 1, 1994.
- [16] W. Levison, "Interactive highway safety design model: issues related to driver modeling," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1631, no. 1, pp. 20–27, 1998.
- [17] FHWA, *Interactive Highway Safety Design Model*, IHSDM, Chicago, IL, USA, 2017.
- [18] D. Solomon, *Accidents on Main Rural Highways Related to Speed, Driver, and Vehicle*, Bureau of Public Roads, U.S. Department of Transportation, Washington, DC, USA, 1964.
- [19] R. Elvik, P. Christensen, and A. H. Amundsen, "Speed and road accidents: an evaluation of the power model," Report 740, Institute of Transport Economics, Oslo, Norway, 2004.
- [20] X. S. Fu, *Study on Theoretical Speed and Highway Alignment Design and Evaluation Method*, Chang'an University, Xi'an, China, 2008.

- [21] AASHTO, *A Police on Geometric Design of Highways and Streets*, American Association of State Highway and Transportation Officials, Washington, DC, USA, 2011.
- [22] R. Lamm and E. M. Choueiri, *Recommendations for Evaluating Horizontal Design Consistency Based on Investigations in the State of New York*, Transportation Research Board, Washington, DC, USA, 1987.
- [23] R. Lamm, B. Psarianos, and T. Mailaender, *Highway Design and Traffic Safety Engineering Handbook*, McGraw-Hill, New York, NY, USA, 1999.
- [24] I. Anderson and R. Krammes, "Speed reduction as a surrogate for accident experience at horizontal curves on rural two-lane highways," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1701, no. 1, pp. 86–94, 2000.
- [25] R. A. Krammes, K. D. Tyer, D. R. Middleton et al., "An alternative to post-mounted delineators at horizontal curves on two-lane highways (FHWA/TX-90/1145-1F)," *Geometric Design*, 1990.
- [26] R. A. Krammes, R. Q. Brackett, M. A. Shafer et al., "Horizontal alignment design consistency for rural two-lane highways," *Geometric Design (FHWA-RD-94-034)*, 1995.
- [27] J. Collins, K. Fitzpatrick, K. Bauer et al., "Speed variability on rural two-lane highways," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1658, no. 1, pp. 60–69, 1999.
- [28] K. Fitzpatrick, "Speed prediction for two-lane rural highway," Federal Highway Administration, US Department of Transportation, Washington, DC, USA, FHWA-RD-99-171, 2000.
- [29] R. Lamm, E. M. Choueiri, and T. Mailaender, "Comparison of operating speeds on dry and wet pavements of two-lane rural highways," *Transportation Research Record*, vol. 1280, no. 1, pp. 199–207, 1990.
- [30] R. Lamm, B. Psarianos, E. M. Chourieri, and G. Soilemezoglou, "A practical safety approach to highway geometric design. International case studies: Germany, Greece, Lebanon and the United States," in *Proceedings of the International Symposium on Highway Geometric Design Practices*, Transportation Research Board, Boston, MA, USA, August 1995.
- [31] A. Jacob and M. V. L. R. Anjaneyulu, "Operating speed of different classes of vehicles at horizontal curves on two-lane rural highways," *Journal of Transportation Engineering*, vol. 139, no. 3, pp. 287–294, 2012.
- [32] A. Maji, G. Sil, and A. Tyagi, "85th and 98th percentile speed prediction models of car, light, and heavy commercial vehicles for four-lane divided rural highways," *Journal of Transportation Engineering, Part A: Systems*, vol. 144, no. 5, Article ID 04018009, 2018.
- [33] A. Maji and A. Tyagi, "Speed prediction models for car and sports utility vehicle at locations along four-lane median divided horizontal curves," *Journal of Modern Transportation*, vol. 26, no. 4, pp. 278–284, 2018.
- [34] G. Sil and A. Maji, "Video based data collection process for geometric design consistency evaluation of four-lane median divided horizontal curves," *Transportation Research Procedia*, vol. 27, pp. 672–679, 2017.
- [35] The Ministry of Transportation of the People's Republic of China, *Specifications for Highway Safety Audit: JTG B05-2015*, People's Communications Press, Beijing, China, 2016.
- [36] M. A. Mohammadi, V. A. Samaranayake, and G. H. Bham, "Crash frequency modeling using negative binomial models: an application of generalized estimating equation to longitudinal data," *Analytic Methods in Accident Research*, vol. 2, pp. 52–69, 2014.
- [37] X. Wang and M. Abdel-Aty, "Temporal and spatial analyses of rear-end crashes at signalized intersections," *Accident Analysis & Prevention*, vol. 38, no. 6, pp. 1137–1150, 2006.
- [38] E. Chen and A. P. Tarko, "Modeling safety of highway work zones with random parameters and random effects models," *Analytic Methods in Accident Research*, vol. 1, pp. 86–95, 2014.
- [39] F. L. Mannering, V. Shankar, and C. R. Bhat, "Unobserved heterogeneity and the statistical analysis of highway accident data," *Analytic Methods in Accident Research*, vol. 11, pp. 1–16, 2016.
- [40] D. Lambert, "Zero-inflated Poisson regression, with an application to defects in manufacturing," *Technometrics*, vol. 34, no. 1, pp. 1–14, 1992.
- [41] C. Gouriéroux, "The econometrics of discrete positive variables: the Poisson model," *Econometrics Of Qualitative Dependent Variables*, pp. 270–283, Cambridge University Press, New York, NY, USA, 2000.
- [42] R. Winkelmann, "Seemingly unrelated negative binomial regression," *Oxford Bulletin of Economics and Statistics*, vol. 62, no. 4, pp. 553–560, 2000.
- [43] J. M. Hilbe, *Negative Binomial Regression*, Cambridge University Press, Cambridge, UK, 2007.
- [44] R. Winkelmann, *Econometric Analysis of Count Data*, Springer, Berlin, Germany, 2008.
- [45] P. C. Anastasopoulos, V. N. Shankar, J. E. Haddock, and F. L. Mannering, "A multivariate Tobit analysis of highway accident-injury-severity rates," *Accident Analysis & Prevention*, vol. 45, pp. 110–119, 2012.
- [46] P. W. Simon, G. K. Matthew, and L. M. Fred, *Statistical and Econometric Methods for Transportation Data Analysis*, CRC Press, Boca Raton, FL, USA, 2004.
- [47] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde, "Bayesian measures of model complexity and fit," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 64, no. 4, pp. 583–639, 2002.
- [48] B. H. Pang, B. Liu, H. Y. Zhou, Y. F. Hou, and S. Q. Wu, "Design index of transition section of ramp lane number change," *Journal of Chang'an University (Natural Edition)*, vol. 38, no. 1, pp. 82–88, 2018.
- [49] J. B. Hu, L. C. He, and R. H. Wang, "Review of Safety evaluation of freeway interchange," *China Journal of Highway Transportation*, vol. 33, no. 7, pp. 17–27, 2020.
- [50] Y. Zhang, Y. Xie, and L. Li, "Crash frequency analysis of different types of urban roadway segments using generalized additive model," *Journal of Safety Research*, vol. 43, no. 2, pp. 107–114, 2012.