

Research Article

A New GNB Model of Crash Frequency for Freeway Sharp Horizontal Curve Based on Interactive Influence of Explanatory Variables

Xiaofei Wang ¹, HuaQiao Pu,² Xinwei Li,³ Ying Yan ⁴, and Jiangbei Yao²

¹Associate Professor, School of Civil Engineering and Transportation, South China University of Technology, Guangzhou 510640, China

²Master, School of Civil Engineering and Transportation, South China University of Technology, Guangzhou 510640, China

³Senior Engineer, Guangzhou Highway Engineering Company, Guangzhou 510288, China

⁴Professor, Key Laboratory of Automobile Transportation Safety Support Technology, Chang'an University, Xi'an 710064, China

Correspondence should be addressed to Xiaofei Wang; xiaofeiw@scut.edu.cn and Ying Yan; yanying2199@chd.edu.cn

Received 2 April 2018; Revised 24 September 2018; Accepted 5 November 2018; Published 14 November 2018

Academic Editor: Antonio Comi

Copyright © 2018 Xiaofei 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.

Crash prediction of the sharp horizontal curve segment of freeway is a key method in analyzing safety situation of freeway horizontal alignment. The target of this paper is to improve predicting accuracy after considering the elastic influence of explanatory variables and interaction of explanatory variables on crash rate prediction. In the paper, flexibility and elasticity are defined to express the elastic influence of explanatory variables and interaction of explanatory variables on crash rate prediction. Thus, we proposed 6 types of models to predict crash frequency. These 6 types of models include 2 NB models (models 1 and 2), 2 GNB models (models 3 and 4), one NB model (model 5), and one GNB model (model 6) with flexibility and variable elasticity considered. The alignment and crash report data of 88 sharp horizontal curve segments from different institutions were surveyed to build the crash models. Traffic volume, highway horizontal radius, and curve length have been assigned as explanatory variables. Subsequently, statistical analysis is performed to determine the model parameters and conducted sensitivity analysis by AIC, BIC, and Pseudo R^2 . The results demonstrated the effective use of flexibility and elasticity in analyzing explanatory variables and in predicting freeway sharp horizontal curve segments. In six models, the result of model 6 is much better than those of the other models by fitting rules. We also compared the actual results from crashes of 88 sharp horizontal curve segments with those predicted by models 1, 3, and 6. Results demonstrate that model 6 is much more reasonable than the others.

1. Introduction

Accidents, and specifically highway-vehicle accidents, cost the lives of roughly one and a quarter million people worldwide every year. In addition, highway-traffic injuries are globally the leading cause of death among people 15 to 29 years old with over 300,000 deaths [1]. Compared with other highways, a freeway is often designed with relatively good driving environment characterized with high alignment indexes, good pavement, total enclosure, absence of pedestrians, no low speed interference, perfect traffic safety devices, and so on. Thus, the crash rate and death toll of freeways average 30%–51% and 43%–76%, respectively, compared with those of ordinary highways in developed countries. In China

however, the average crash number, death toll, injury toll, and the direct loss of property are 3.2, 8.4, 7.2, and 24.3 times more than those of the ordinary highways [2]. Therefore, it is important to determine the real law of crash occurring in freeways and how the different types of freeway environment influence the crash number based on reliable databases.

Over the past several decades, historical surveys covering the features and frequencies of crashes in freeways have been actively pursued [3, 4]. Researchers mainly focus on investigating the observed factors that affect the number of crashes for roadway segments or intersections over some fixed time periods [5]. However, in terms of freeway crashes within China, specialized crash databases and highway design databases are not available at present. Similarly,

investigations that could clarify China's current situation have not been performed. Thus, Liande et al. [6], Zhuanglin [7], and other researchers developed a crash prediction model with a relatively small number of samples. To improve on this effort, this paper attempts to establish a model with huge samples.

Mathematical statistics and regression analyses are common methods to predict highway crashes. Other methods, such as fuzzy mathematics, grey theory, nerve cell method, and clustering analysis, have also been used to establish the prediction models. American HSM2010 is an established prediction model based on statistical regression. IHSDM made a good simulation of the American two-lane highway crash prediction (U.S. federal highway data). Chengcheng [8] carried out two-lane highway crash prediction model research, which focused on low-grade highways in China.

The abovementioned methods have explained how a single factor influences the crashes. However, road crashes are complex events that involve a large variety of factors with multifaceted interactions, making it challenging to fully understand them. Advancing improved methodologies for road safety analysis and applying them to crash analysis continue to be investigated [9].

Many statistical models are applied to crash frequency analysis. Improved NB and GNB models were successfully used to predict the crash rate of freeway basic segment, tunnel entrance and exit, and so on [10, 11]. However most of these models with fixed parameters fail to reveal the true interrelationship between explanatory variables. Flexibility and elasticity are concepts introduced to analyze the degree of interaction between variables when the translog function is extended to the field of traffic accident prediction. In our research, flexibility and elasticity have been introduced to express the elastic influence of explanatory variables and interaction of explanatory variables on crash rate. To demonstrate the effective use of flexibility and elasticity in predicting freeway SHCS, we conducted sensitivity analysis by AIC, BIC, and Pseudo R^2 and compared our proposed models to traditional NB and GNB models. The results of the proposed prediction models were also compared with observed data.

Our research on the crash prediction models divided the freeway into several segments, namely, basic segment, general segment, and special segment. Since we have discussed the crash prediction model of the basic segments in the paper published in Journal of Southeast University (Wang et al. 2014), we take the freeway sharp horizontal curve segment (SHCS) as the research object in this paper. In the crash prediction model, segment length, curve radius, and traffic volume are assigned as explanatory variables and crash amount per year is determined as the dependent variable.

In the next section, relevant literature was reviewed. In Section 3, we presented the data and the glossaries as well as the basic model formulated and the discussion on the variables. In Section 4, we discussed the statistical analysis used to determine the model parameters and the sensitivity analysis. We also compared the actual crashes with the predicted ones. Moreover, we presented the final results in this section. In Section 5, we presented the conclusions.

2. Literature Review

A wide variety of advanced statistical count models are applied to crash frequency analysis over the past years, and the strengths and weaknesses were well summarized by Lord and Mannering [5], Mannering and Bhat [12], and Mannering et al. [13]. Mohammadi et al. [14] refined a good summary of prediction models based on several representative references (Gomes and Vieira, 2013; Ukkusuri et al., 2012; Hadayeghi et al., 2007; Kim et al., 2006; Qadeer Memon, 2006; Maher and Summersgill, 1996; Daniels et al., 2010; Lord et al., 2010, 2007, 2005). "Different formulations were used based on the purpose of study and nature of available data. When the data are overdispersed, the Negative Binomial (NB) model structure with a log-link function is the most favored. When the data are not overdispersed, the Poisson structure is the most favored" [14]. Then, they developed a series of aggregate crash prediction models that relate to the modal split step of the conventional four-step demand models [14]. Yajie (2014) investigated the effect of different functional forms on the estimation of the weight parameter as well as the group classification of the finite mixture of NB regression models, using crash data collected on rural roadway sections in Indiana.

The above review shows the general linear model or logarithm linear model by logarithmic transformation into linear equation is one of the most commonly used methods of building a highway traffic crash prediction model. Although there are several limitations [15], many of the crash prediction models of Highway Safety Manual 2010 are based on logarithm linear models and turned to be reliable to a certain degree. The conventional NB or GNB model with fixed parameters may fail to capture the possible variability of the individual effects associated with the variables across observations, which may lead to biased parameter estimation and incorrect inferences [16].

Analysis of the common traffic crash prediction models has resulted in the observation that, in the process of building the model, the basic assumption that all explanatory variables that are relatively independent are common does not consider the influence of each variable. This observation results in a situation where the relationship between explanatory variables and the traffic crash is not fully in accordance with the actual situation. Although a considerable number of recent highway safety studies [17–19] considered the interaction among explanatory variables, most are based on the analysis of the relationship between driver, vehicle, highway [20], and environment [21]. The results of these studies show the different dangers when driving in highways and the effect of division on the traffic flow, among others. Moreover, the results show that when the lengths of segments analyzed are different, the traffic flow prediction for the crash is also different.

Thus, we can find that the traditional log-linear model has two limitations when used to analyze and predict the frequency of road traffic crashes. Firstly, the assumption that the elastic coefficient is constant is not compatible with common logic [22]. Secondly, exploratory variables are simplified or idealized as independent variables, which is hard to reveal

TABLE 1: Sample source and size.

Source	observation period	Freeway length (Km)	Accident
MCGP	2008-2012	2200.779	135498
NSG freeway TPD & FAMC	2008-2012	72.00	1428
GZJC freeway TPD & FAMC	2008-2012	50.74	3115
JZN freeway TPD & FAMC	2006-2012	109.84	11209
GH freeway TPD & FAMC	2008-2012	155.306	3441
KY freeway TPD & FAMC	2008-2012	125.20	2351
GZBH freeway TPD & FAMC	2007-2012	21.652	12850
SM freeway TPD & FAMC	2008-2011	58.361	719

TABLE 2: Crash data.

Name of variable	Mean value	Standard deviation	MIN	MAX
Crash amount (/km/year)	3.307	8.814	0	15
AADT (10thousands pcu)	0.714	1.037	0.25	4.78
Length of segment (km)	0.860	0.017	0.56	1.18
Radius of segment (km)	0.624	0.041	0.31	1
Operation Time (years)	8.020	2.251	3	14

the true interrelationship between variables [22]. Thus, the idea of flexibility has been introduced in our models to overcome the above limitations. Flexibility is often used in the manufacturing industry to explain the variation environment or the probabilistic ability from the variation. Cobb–Douglas production function, linear production function, Leontief production function, Variable Elasticity of Substitution (VES) production function, and translog production function are often used to analyze flexibility. Among these methods, the translog production function is the most popularly used to analyze traffic problems [11]. Wei Huang (2007) and Li Li (2011) studied the generalized beyond the logistic cost function (GTCF). António (2011), Lurong Wu (2010), Juan Zeng [23], Rong Li (2013), and Xiang Liu [24] introduced the translog function to analyze the traffic crash of loss and frequency. Using the logarithmic function NB model, Xiang Liu [24] and Rong Li (2013) established the frequency forecast model of the highway traffic crash in Ontario, Canada. Compared with the log-linear NB model, it was proven to be more credible. Other results also show that the above limitation can be well overcome [22, 25, 26].

3. Materials and Methods

3.1. Materials. To acquire enough samples for a meaningful statistical analysis, four major sources were used: *National Statistics Annual Report of Highway Traffic Crashes* (NSAR-RTA, 2013), *Statistical Bulletin of Transportation Industry Development* (SBTID, 2013), *Traffic Management Committee of Guangdong Province* (MCGP), different Traffic Police Detachments (TPD, 7 freeways, 593.099 km total), different Freeway Administration and Maintenance Centers (FAMC, 7 freeways, 593.099 km total), and additional results provided by other scholars. Table 1 presents the sample size.

In paper, 88 SHCSs from eight four-lane highways of Guangdong Province covering the period from 2008 to 2012

and their crash data of five years were selected for analysis. The statistics are shown in Table 2.

3.2. Definition. In the study, the following segments are defined as sharp horizontal curve prediction segments:

- (1) Radius of horizontal curve: less than 1000 m
- (2) Lane number: two-way 4-lane
- (3) Lane width: 3.75 m
- (4) Hard shoulder: on both sides
- (5) Median: yes
- (6) Crash barrier: on both sides
- (7) Lighting: none
- (8) AADT (two directions): no more than 5.76×10^4 pcu/day
- (9) Open to traffic duration: no less than 2 years and no reconstruction in 2 years

3.3. Modeling Method. When building a road traffic accident frequency prediction model, in order to keep the model stable (that is, to keep variance as low as possible), model transformation is performed to the NB, which is commonly used method. Thus, NB model will be transferred into log-linear model. Equations (1)~(5) and all NB models in this paper are results of transformation. Thus, it looks to be not a common NB formulation of the form $Y = \exp(a + \sum \beta_i x_i)$. In the paper, we still name these formulations as NB regression models or GNB models.

The basic expression form of the translog function model is as shown in formula (1) [27, 28].

$$\ln Y = \alpha_0 + \alpha_k \ln K + \alpha_L \ln L + \alpha_{kk} (\ln K)^2 + \alpha_{ll} (\ln L)^2 + \alpha_{kl} \ln K \ln L \quad (1)$$

TABLE 3: Models and statistics parameters.

Model	The basic function form of model	Estimated parameters *
1 (NB)	$\ln \mu_i = \alpha_0 + \alpha_1 \ln(T_i) + \alpha_2 \ln(L_i) + \alpha_3 \ln(R_i)$	α_{0-3}, β
2 (NB)	2.1 $\ln \mu_i = \alpha_0 + \alpha_1 \ln(T_i) + \alpha_2 \ln(L_i) + \ln(R_i)$ 2.2 $\ln \mu_i = \alpha_0 + \alpha_1 \ln(T_i) + \ln(L_i) + \alpha_3 \ln(R_i)$ 2.3 $\ln \mu_i = \alpha_0 + \ln(T_i) + \alpha_2 \ln(L_i) + \alpha_3 \ln(R_i)$	α_{0-3}, β
3 (GNB)	$\ln \mu_i = \alpha_0 + \alpha_1 \ln(T_i) + \alpha_2 \ln(L_i) + \alpha_3 \ln(R_i)$ $\beta_i = e^{(\lambda_0 + \lambda_1 \ln(V_i))}$ $\beta_i = e^{(\lambda_0 + \lambda_1 \ln(V_i))}$	α_{0-3} λ_{0-1}
4 (GNB)	4.1 $\ln \mu_i = \alpha_0 + \alpha_1 \ln(T_i) + \alpha_2 \ln(L_i) + \ln(R_i)$ 4.2 $\ln \mu_i = \alpha_0 + \alpha_1 \ln(T_i) + \ln(L_i) + \alpha_3 \ln(R_i)$ 4.3 $\ln \mu_i = \alpha_0 + \ln(T_i) + \alpha_2 \ln(L_i) + \alpha_3 \ln(R_i)$	α_{0-3} λ_{0-1}
5 (NB+translog +flexibility)	$\ln \mu_i = \alpha_0 + \alpha_1 \ln(T_i) + \alpha_2 \ln(L_i) + \alpha_3 \ln(R_i) + \alpha_4 [\ln(T_i)]^2 + \alpha_5 [\ln(L_i)]^2 + \alpha_6 [\ln(R_i)]^2 + \alpha_7 \ln(T_i) \ln(L_i) + \alpha_8 \ln(T_i) \ln(R_i) + \alpha_9 \ln(L_i) \ln(R_i)$	α_{0-9}, β
6 (GNB +translog +flexibility)	$\ln \mu_i = \alpha_0 + \alpha_1 \ln(T_i) + \alpha_2 \ln(L_i) + \alpha_3 \ln(R_i) + \alpha_4 [\ln(T_i)]^2 + \alpha_5 [\ln(L_i)]^2 + \alpha_6 [\ln(R_i)]^2 + \alpha_7 \ln(T_i) \ln(L_i) + \alpha_8 \ln(T_i) \ln(R_i) + \alpha_9 \ln(L_i) \ln(R_i)$ $\beta_i = e^{(\lambda_0 + \lambda_1 \ln(V_i))}$	α_{0-9} λ_{0-1}

* β : overdispersion parameter. The higher β is, the more scattered the distribution is. V_i represent T, L, R, TR, TL, RL, and TRL of segment i . The determining method is discussed in Section 4.

where Y is the dependent variable, K and L are the explanatory variables, and $\alpha_0, \alpha_k, \alpha_L, \alpha_{kk}, \alpha_{ll}, \alpha_{kl}$ are the estimated parameters.

All the basic formulas are formulated based on formula (1) by introducing second cross variables and using trans logarithmic (TCF) cost function form to extend NB and the Poisson model. Consequently, the interaction between the variables can be reflected. Couto et al. [25] established the logarithmic function model based on AADT, segment length, density of access, and time trend variables. The logarithmic functional formula is expressed as

$$\begin{aligned}
\ln \mu_{it} = & \beta_0 + \beta_1 \ln(F_{it}) + \beta_2 \ln(Leng_i) + \beta_3 IntDen_i \\
& + \ln(\gamma) TimeT_t + \beta_4 [\ln(F_{it})]^2 \\
& + \beta_5 [\ln(Leng_i)]^2 + \beta_6 (IntDen_i)^2 + \beta_7 \\
& \times \ln(F_{it}) \ln(Leng_i) + \beta_8 \ln(F_{it}) IntDen_i \\
& + \beta_9 \ln(Leng_i) IntDen_i
\end{aligned} \quad (2)$$

where μ_{it} is the average accidents per year, F_{it} , $Leng_i$, Den_i , and T_t are the explanatory variables, which are referred to as AADT, segment length, density of access, and time trend variables, and β_k ($k = 0 \sim 9$) and γ are the estimated parameters.

In our paper, the freeway crash prediction model is built by selecting AADT, length of sharp horizontal curve segments, and curve radius as explanatory variables. The

NB crash prediction model is set up based on the constant elasticity and flexibility of variables; see formulas (3) and (4).

$$\ln \mu_i = \alpha_0 + \alpha_1 \ln(T_i) + \alpha_2 \ln(L_i) + \alpha_3 \ln(R_i) \quad (3)$$

$$\begin{aligned}
\ln \mu_i = & \alpha_0 + \alpha_1 \ln(T_i) + \alpha_2 \ln(L_i) + \alpha_3 \ln(R_i) \\
& + \alpha_4 [\ln(T_i)]^2 + \alpha_5 [\ln(L_i)]^2 \\
& + \alpha_6 [\ln(R_i)]^2 + \alpha_7 \ln(T_i) \ln(L_i) \\
& + \alpha_8 \ln(T_i) \ln(R_i) + \alpha_9 \ln(L_i) \ln(R_i)
\end{aligned} \quad (4)$$

where μ_i is the estimate of crash amount for a specific year of segment i , T_i is AADT for a specific year of segment i , R_i is radius of the curve i , L_i is the length of segment i , and α_k ($k = 0, 1, 2, \dots, 5$) is estimated parameter.

Then, flexibility and elasticity are defined to express the elastic influence of explanatory variables and interaction of explanatory variables on crash rate. Totally, we proposed 6 types of models to predict crash frequency; see Table 3. These 6 types of models include 2NB models (models 1 and 2), 2GNB models (models 3 and 4), one NB model (model 5), and one GNB model (model 6) with flexibility and variable elasticity considered. As discussed above (literature review), there are 2 main limitations of NB or GNB. Translog transformation and variables flexibility are expected to solve the limitations. Thus, first 4 models (2 NB models and 2 GNB models) are used as comparisons to check whether translog transformation and variables flexibility can improve the prediction. Akaike Information Criteria (AIC criterion), Bayesian Information Criteria (BIC) rule, and Pseudo R2 test were used as criteria to evaluate the imitative effect of the crash CPM of SHCS.

For NB distribution, the overdispersion parameter is a constant greater than zero. The higher the value is, the more dispersed the observations are. GNB distribution also follows this rule but allows the overdispersion parameter to change with other variables. The overdispersion parameter was selected and its expression equation was determined by analyzing the fit goodness of overdispersion parameter of different parameters. Then, we compared the fit goodness of the six models below and ultimately determined the CPM of SHCS. The corresponding forms of each model and estimated parameters are shown in Table 3.

4. Statistical Analysis

4.1. Overdispersion Parameter. The difference between the NB model and the GNB model is whether overdispersion parameter β is a constant or not.

The first 2 types are NB models and the overdispersion parameter is a constant greater than zero. The difference of models 2.1, 2.2, and 2.3 is variables' coefficients. In model 2.1, we set a fixed value 1 to R_i as relative coefficient to find the relative influence of T_i and L_i on R_i . Purposes are similar in models 2.2, 2.3, and 4.1~4.3. The comparison gives us a clear understanding about relationship between 3 variables. This method has also been applied in some other literatures [10, 11].

Overdispersion parameter β of GNB is not a constant value. We separately tried T, L, R, TR, TL, RL, and TR as explanatory variables of β . Thus, each GNB model has 7 specific models.

AIC, BIC, and Pseudo R^2 are used to select the best specific model for each of the six main models with the best goodness of fit. The overdispersion parameter of each model and its AIC, BIC, and Pseudo R^2 coefficient are listed in Table 4.

The following standards are used to examine and verify the goodness of fit of parameters of β :

(1) The Pseudo R^2 statistical magnitude should be used to test the goodness of fit of the models. The bigger it is, the better the model is.

(2) AIC is used to evaluate whether the model is useful or not. The smaller it is, the better the model is.

(3) BIC states that any given problem can find the smallest error probability by the likelihood ratio test of decision rules. Thus, the smaller it is, the better the model is.

As shown in Table 4, among 3 NB models, overdispersion parameter of model 5 has relatively small AIC and BIC and big Pseudo R^2 . It proves that flexibility and elasticity are beneficial to improve the models' overdispersion parameter. Then, among 3GNB models, although the value of the models is quite close for some models, when T is selected as explanatory variable of β , the AIC and BIC values of models tend to be smaller, and the Pseudo R^2 value tends to be larger than the others. The results indicate that the fitting effect of these models is better than those of others. Thus, we determined T as explanatory variable of β . That is, $\beta_i = e^{(\lambda_0 + \lambda_1 \ln(T_i))}$.

4.2. Model Result. Based on the collected data mentioned in Section 2, we calibrated the estimated parameters of the six main models and the specific models cited above. The goodness of fit was also calculated. The results are shown in Table 5.

Table 5 indicates the obvious interactive influence between two variables. Model 5 and model 6, which take the interactive influence into consideration, have a good fitting effect, particularly when compared to models 2 and 4. Thus, we ignored models 2 and 4 directly. According to 3 evaluation criteria, model 6 is better than model 5. By contrast, we found that the Pseudo R^2 of model 6 is larger than those of models 1 and 3, indicating that model 6 is much better than models 1 and 3.

Based on the above analysis, we determined model 6 as CPM and expressed it as follows:

$$N = e^{(3.18+0.60\ln(T_i)-11.70\ln(L_i)+7.85\ln(R_i)+0.015[\ln(T_i)]^2+8.49[\ln(L_i)]^2-2.41[\ln(R_i)]^2+0.64\ln(T_i)\ln(L_i)-1.09\ln(T_i)\ln(R_i)-5.80\ln(L_i)\ln(R_i))} \quad (5)$$

The overdispersion parameter is

$$\beta_i = e^{[1.17-6.35\ln(T_i)]} \quad (6)$$

where N is estimate of crash amount for every year of SHCS, T_i is the basic segment of the annual average daily traffic, L_i is the length of the SHCS, and R_i is the radius of the SHCS.

4.3. Prediction Analysis. To demonstrate the effectiveness of the prediction, we performed prediction of a certain freeway with model 1, model 3, and model 6. Then, we compared the results with the real crash data we collected from the institutions. See Table 6.

As shown in Table 6, the crash averages of the 3 models are all close to the real crash value. However, there are some

differences when referring to standard deviation as shown in Table 6. Due to overfitting through the translog specification and the large number of interactions, the result of model 6 is much closer to the statistics value of the real cash data than those of the other two models. For the maximum and minimum values, the forecast range of model 6 is very close to the actual situation. Based on the above discussion, model 6 is the best among the six models.

5. Conclusion and Discussions

The traditional log-linear model has two limitations when used to analyze and predict the frequency of road traffic crashes. One is the constant elastic coefficient and the other is

TABLE 4: AIC, BIC, and Pseudo R2 of 6 main models and their specific models.

Model	Explanatory Variable of β	AIC	BIC	Pseudo-R ² s
1	--	401.12	413.51	0.03
2.1		406.32	416.23	0.02
2.2	--	407.63	417.53	0.02
2.3	--	431.79	441.70	0.01
3	T	398.47	413.33	0.04
3	RT	400.33	417.68	0.04
3	LT	400.34	417.68	0.04
3	LTR	402.18	422.00	0.04
3	L	402.71	417.57	0.03
3	R	402.96	417.82	0.03
3	LR	404.38	421.72	0.03
4.1	T	409.23	421.61	0.02
4.1	R	409.34	421.73	0.02
4.1	L	409.61	422.00	0.01
4.1	RT	410.37	425.23	0.02
4.1	LT	411.19	426.05	0.01
4.1	LR	411.32	426.19	0.01
4.1	LTR	412.29	429.63	0.01
4.2	L	408.08	420.46	0.02
4.2	T	408.12	420.508	0.02
4.2	R	408.30	420.68	0.02
4.2	LT	410.04	424.91	0.02
4.2	LR	410.08	424.94	0.02
4.2	RT	410.10	424.96	0.02
4.2	LTR	412.04	429.38	0.01
4.3	T	402.32	414.71	0.04
4.3	R	402.49	414.87	0.04
4.3	L	402.55	414.94	0.05
4.3	RT	404.30	419.17	0.03
4.3	LT	404.32	419.18	0.03
4.3	LR	404.49	419.35	0.03
4.3	LTR	406.30	423.64	0.03
5	--	400.71	427.96	0.06
6	T	393.99	423.72	0.08
6	RT	395.49	427.70	0.08
6	LT	395.99	428.20	0.08
6	LTR	397.49	432.18	0.08
6	L	402.34	432.06	0.06
6	R	402.71	432.44	0.06
6	LR	404.22	436.43	0.06

independent relationship between variables. Translog transformation and the idea of flexibility have been introduced in our models to overcome the above limitations. Flexibility and elasticity are defined to express the elastic influence of explanatory variables and interaction of explanatory variables on crash rate. The analysis sheds light on crash prediction effect of SHCS of freeways. Thus, totally six types of models, a total of 10 models, were proposed to predict the crash. These 6 types of models include 2 NB models (models 1 and 2), 2GNB models (models 3 and 4), one NB model (model 5), and one

GNB model (model 6) with flexibility and variable elasticity considered. Among the models, model 6 is much better than the other models. All parameter estimates in Table 5 satisfy the models with a confidence of more than 1%. Through the detailed analysis and study, the following conclusions have been drawn:

- (1) Among 3 NB models, with flexibility and elasticity considered in the model, model 5 has relatively small AIC and BIC and big Pseudo R2. It demonstrated that

TABLE 5: Estimated and statistics parameters.

Estimated parameter	Model 1	Model 2			Model 3		Model 4		Model 5	Model 6
		2.1	2.2	2.3		4.1	4.2	4.3		
α_0	1.93	1.37	0.56	1.59	2.96	1.58	0.48	1.42	5.00	3.81
α_1	0.29	0.17	0.30	1.00	0.31	0.15	0.33	1.00	0.78	0.60
α_2	-0.77	-1.10	1.00	-0.23	-1.60	-1.33	1.00	-0.72	-14.05	-11.70
α_3	-0.001	1.00	-0.76	-1.41	-0.001	1.00	-0.65	-0.73	6.62	7.85
α_4	—	—	—	—	—	—	—	—	0.003	0.015
α_5	—	—	—	—	—	—	—	—	9.29	8.49
α_6	—	—	—	—	—	—	—	—	-2.79	-2.41
α_7	—	—	—	—	—	—	—	—	0.38	0.64
α_8	—	—	—	—	—	—	—	—	-0.001	-1.09
α_9	—	—	—	—	—	—	—	—	-0.004	-5.80
λ_0	—	—	—	—	2.06	-0.63	-0.93	-0.82	—	1.17
λ_1	—	—	—	—	-8.22	-0.18	0.12	-0.15	—	-6.35
β	0.36	0.46	0.45	0.811	—	—	—	—	0.23	—
AIC	401.13	407.63	406.32	431.79	398.47	409.22	408.12	431.79	400.71	393.99
BIC	413.51	417.54	416.23	441.70	413.33	421.61	420.50	441.70	427.96	423.72
Pseudo R2	0.03	0.02	0.02	0.01	0.04	0.016	0.023	0.039	0.059	0.080
LR chi2	11.13	6.12	9.38	4.57	15.76	6.52	9.41	15.96	23.54	32.24

TABLE 6: Statistics values of predicted and real crashes.

	Obs (the sample number)	Average	Std. Dev	Min	Max
The real crashes	88	3.31	2.87	0	15
Prediction result of model 1	88	3.30	1.11	2.38	7.39
Prediction result of model 3	88	3.60	1.44	2.16	10.73
Prediction result of model 6	88	3.38	1.82	2.16	15.13

flexibility and elasticity help to improve the models' overdispersion parameter.

- (2) When T is selected as the overdispersion parameter in GNB models, the AIC and BIC values of models tend to be smaller, and the Pseudo R2 value tends to be larger than the others. The results indicate that the fitting effect of these models is better than those of others.
- (3) With sufficient samples and data, the effective use of the GNB model in analyzing the interactive influence of explanatory variables and predicting freeway basic segments can be demonstrated. The prediction results with relatively good models (model 1, model 3, and model 6) have also been compared to that of real data. The results show model 6 has the best prediction effect.

In summary, the results show suitable length of circle (Li) allows drivers to adapt to the driving environment. Long sharp curve and high traffic volume Ti result in high crash rate. However, when Ti is low (e.g., free flow), the influence of Li on crash rate will decrease. Sharp curve with high traffic volume Ti results in high crash rate. When Ti is low (e.g., free flow), the influence of Ri on crash rate will decrease. Long

sharp curves with really small Ri are difficult for drivers to handle the steering wheel. Drivers are also nervous driving along sharp curves. There must be an interaction between them.

Thus, the findings of this study can help enhance understanding of the relationship among traffic volume, highway horizontal radius, and curve length. Such understanding can help to develop crash prevention strategies for specific conditions. For example, the findings can provide an important guide for designers when applying the horizontal radius and curve length. Moreover, the results could be used as basis to implement a variable traffic speed limit on curves to reduce crash risk while traveling on a hazardous roadway segment.

However, further efforts should be made to demonstrate the differences between the NB and GNB models. The experimental data were limited. Thus, the model fitting effect is slightly far from ideal. What the impact degree of variable interaction on prediction accuracy is, is also our future research topic. With flexibility and elasticity defined to express the elastic influence of explanatory variables and interaction of explanatory variables on crash rate, GNB model 6 is proved to have good fittingness, which offers a certain reference value for crash prediction in general.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study is supported by National Natural Science Foundation of China (51878297). Funding of research is provided by Guangdong Communication Department (2015-02-003, 2015-02-004), Fund of State Key Laboratory of Bridge Engineering Structural Dynamics (Key Laboratory of Bridge Earthquake Resistance Technology, Ministry of Communications, PRC), Open Funding Project of National Engineering Laboratory for Surface Transportation Weather Impacts Prevention (NELJA201701), and Natural Science Foundation of Shaanxi Province (2018JM5117).

References

- [1] V. Prasad, U. Schwerdtfeger, F. El-Awa, D. Bettcher, and V. da Costa e Silva, "Closing the door on illicit tobacco trade, opens the way to better tobacco control," *Eastern Mediterranean Health Journal*, vol. 21, no. 6, pp. 379-380, 2015.
- [2] X. Wang, S. Yang, Y. Yan, X. Li, J. Du, and J. Tang, "Analysis on distribution of freeway accidents under various conditions in China," *Advances in Mechanical Engineering*, vol. 8, no. 8, pp. 1-12, 2016.
- [3] S. S. Durduran, "A decision making system to automatic recognize of traffic accidents on the basis of a GIS platform," *Expert Systems with Applications*, vol. 37, no. 12, pp. 7729-7736, 2010.
- [4] J. White, C. Thompson, H. Turner, B. Dougherty, and D. C. Schmidt, "WreckWatch: automatic traffic accident detection and notification with smartphones," *Mobile Networks and Applications*, vol. 16, no. 3, pp. 285-303, 2011.
- [5] D. Lord and F. Mannering, "The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives," *Transportation Research Part A: Policy and Practice*, vol. 44, no. 5, pp. 291-305, 2010.
- [6] Z. Liande, S. Xiaoduan, C. Yongsheng et al., "Accident prediction model of freeway," *Journal of Beijing University of Technology*, vol. 35, no. 7, pp. 966-971, 2009.
- [7] M. Zhuanglin, "Temporal-spatial analysis model of traffic accident and its prevention method on expressway. Beijing," *Journal of Traffic and Transportation Engineering*, vol. 12, no. 2, 2012.
- [8] T. Chengcheng, *Study on Key Technologies of Accident Prediction and Prevention for Two-Lane Highways*, Tongji University, Shanghai, China, 2009.
- [9] Q. Hou, A. Tarko P, and X. Meng, "Investigating factors of crash frequency with random effects and random parameters models: New insights from Chinese freeway study[J]. Accident Analysis Prevention," in Meng X. *Investigating factors of crash frequency with random effects and random parameters models: New insights from Chinese freeway study[J]. Accident Analysis Prevention*, p. 120, 1-12, 120, 2018.
- [10] X. Wang, X. Li, X. Fu, L. Zhao, and X. Liu, "Construction of crash prediction model of freeway basic segment based on interactive influence of explanatory variables," *Journal of Southeast University(English Edition)*, vol. 31, no. 2, pp. 276-281, 2015.
- [11] X. Li, *Highway Accident Prediction Model and Its Application [D]*, Tongji University, 2013.
- [12] F. L. Mannering and C. R. Bhat, "Analytic methods in accident research: methodological frontier and future directions," *Analytic Methods in Accident Research*, vol. 1, pp. 1-22, 2014.
- [13] 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.
- [14] M. Mohammadi, G. Shafabakhsh, and A. Naderan, "Effects of modal shares on crash frequencies at aggregate level[J]. Accident Analysis Prevention," in Naderan A. *Effects of modal shares on crash frequencies at aggregate level[J]. Accident Analysis Prevention*, pp. 120-295, 295-303, 120, 2018.
- [15] S.-P. Miaou, "Some limitations of the models in the highway safety manual to predict run-off-road crashes," *Transportation Research Record*, no. 2377, pp. 38-48, 2013.
- [16] S. P. Washington, M. G. Karlaftis, and F. L. Mannering, *Statistical and Econometric Methods for Transportation Data Analysis*, CRC Press, Boca Raton, Fla, USA, 2nd edition, 2011.
- [17] G. Yannis, J. Golias, and E. Papadimitriou, "Driver age and vehicle engine size effects on fault and severity in young motorcyclists accidents," *Accident Analysis & Prevention*, vol. 37, no. 2, pp. 327-333, 2005.
- [18] J. D. Hill and L. N. Boyle, "Assessing the relative risk of severe injury in automotive crashes for older female occupants," *Accident Analysis and Prevention*, vol. 38, no. 1, pp. 148-154, 2006.
- [19] B.-S. Liu, "Association of intersection approach speed with driver characteristics, vehicle type and traffic conditions comparing urban and suburban areas," *Accident Analysis & Prevention*, vol. 39, no. 2, pp. 216-223, 2007.
- [20] S.-P. Miaou and H. Lum, "Modeling vehicle accidents and highway geometric design relationships," *Accident Analysis & Prevention*, vol. 25, no. 6, pp. 689-709, 1993.
- [21] L. Fridstrøm, J. Ifver, S. Ingebrigtsen, R. Kulmala, and L. K. Thomsen, "Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts," *Accident Analysis & Prevention*, vol. 27, no. 1, pp. 1-20, 1995.
- [22] R. Li, X. Liu, and J. Liu, "Research on traffic accident frequency prediction based on translog production function," *Hunan Daxue Xuebao/Journal of Hunan University Natural Sciences*, vol. 40, no. 4, pp. 49-54, 2013.
- [23] J. Zeng, "Research on influencing factors of socio-economic losses of road traffic accident based on generalized linear models," *Wuhan Ligong Daxue Xuebao/Journal of Wuhan University of Technology*, vol. 32, no. 6, pp. 155-158, 2010.
- [24] Liu. Xiang, Study on road safety based on pedestrian and environment,.
- [25] A. Couto and S. Ferreira, "A note on modeling road accident frequency: A flexible elasticity model," *Accident Analysis & Prevention*, vol. 43, no. 6, pp. 2104-2111, 2011.
- [26] L. Wu, F. Liang, and Z. Shi, "Transcendental logarithmic production function model for traffic accident losses in China[J]," in *Mathematics in Practice and Theory*, vol. 22, pp. 56-61, 56-61, 22, 2010.

- [27] H. Bozdogan, "Model selection and Akaike's Information Criterion (AIC): the general theory and its analytical extensions," *Psychometrika*, vol. 52, no. 3, pp. 345–370, 1987.
- [28] L. R. Christensen, D. W. Jorgenson, and L. J. Lau, "Transcendental logarithmic production frontiers," *Review of Economics Statistics*, vol. 55, no. 1, pp. 28–45, 1973.



Hindawi

Submit your manuscripts at
www.hindawi.com

