Available online www.jsaer.com

***Journal of Scientific and Engineering Research*, 2020, 7(6):71-78**

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Analysis of Factors Affecting the Severity of Single-vehicle Accidents in Chinese Cities Based on Multi-layer Order Logit Model

S.S. Hu[1]*, Z.G. Cai[1], Y.Q. Zhou[1], X.H. Guo[2], L.N. Rong[3], Z. Zhang[4]

[1]School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo, Shandong, China
[2]Shouguang Office of Shandong Fishery Mutual Insurance Association, Shouguang, Shandong, China
[3]Faculty of Education, Shandong Normal University, Jinan, Shandong, China
[4]School of Mining, China University of Mining and Technology, Xuzhou, Jiangsu, China
*Corresponding author: hu15689041806@163.com

Abstract To further clarify the factors that have a significant impact on the severity of single-vehicle(SV) crashes, the SV accident data of a city in China was collected, then an ordered Logit model and a multi-layer ordered Logit model was established, and the Bayesian approach was used to estimate the parameters for the model. The goodness of fit for the model was evaluated by Deviance Information Criterion (DIC). The modeling results show that: the multi-layer ordered Logit model has a better fit than the ordered Logit model. And It has a positive correlation with the severity of traffic accidents such as drunk driving, age greater than 59 years old, motorcycle, rollover, collision fixtures, early morning hours, collision pedestrians, holidays. It has a negative correlation with the severity of traffic accidents such as male drivers, bad weather, and visibility less than 50m.

Keywords Traffic safety, Ordered Logit model, Multi-layer ordered Logit model, Bayesian estimation, DIC

Introduction

Traffic safety is closely related to our daily lives. Traffic accidents not only cause a lot of property losses, but also seriously affect the lives of residents. According to statistics, at least 1.25 million people die from road traffic accidents worldwide each year. On average, at least 3425 people die every day. Road traffic accidents have become the ninth leading cause of death worldwide. Among different types of traffic accidents, single-vehicle traffic accidents, the number of accidents only accounts for 6.17% of the total number of accidents, but the proportion of deaths caused by traffic accidents is as high as 12.06% [1-2], especially the probability of causing driver death is higher than other types of traffic accidents. Therefore, it is of great practical significance to analyze the mechanism of SV accidents and the severity of SV accidents.

In recent years, some experts and scholars at home and abroad have conducted a more comprehensive study on SV accidents. Foreign studies on SV accidents are earlier and more systematic. In recent years, there are more studies on the traffic accidents in sections, intersections and highways. Wen Huiying used the nested logit model to study the causes of the severity of motorcycle accidents on road sections [3], and multiple Logit models influenced the factors affecting the severity of SV at road intersections [4]. Liu Xiaoxiao [5] studied the reproduction of single-vehicle collision accidents with PC-Crash through computer technology, which provided a better method for the study of accident occurrence mechanism. Zhong Chunyao [6] conducted a good study on the single-vehicle accident on the highway, analyzed the spatial and temporal distribution of the single-vehicle accident and the pattern of accident morphology. From the ergonomics aspect, he used cluster analysis to study

the human factor and used the correlation between the integral element's geometric linear analysis of the road and the accident establishes a prediction model for single-vehicle accidents on expressways.

Foreign countries have conducted systematic studies on the influencing factors and injuries of SV accidents earlier. In recent years, foreign studies on SV accidents have also become more elaborate, mainly focusing on the impact of driver characteristics and spatiotemporal characteristics on SV accidents. Ruth A. Shults [7] and others studied the characteristics of non-fatal accidents of SV collisions among young drivers in South Carolina. The research shows that unreasonable driving, passenger carrying and speeding of young people increase the driving risk and improve the South Carolina. The safety of young drivers in Rhone offers effective strategies. Mo Zhou [8] *et al* conducted a systematic study on the characteristics of the out-of-control single-vehicle crashes accidents by using two sorted ordered probability models, and revealed the impact of the severity of the driver and passenger injuries on the out-of-control bicycle collision. Zhenning Li [9] and others developed a unified data set through a mixture of Logit and potential Logit models to thoroughly study the factors that affect the severity of driver injuries in rural bicycle accidents on rainy days. Emmanuel Kofi Adanu [10] and others analyzed the severity of SV on weekdays and weekends through latent classes. The model tested the correlation between the factors of bicycle accidents on weekdays and weekends. The results showed that serious bicycle accidents occurred on Sunday is higher than usual. Ali Behnood [11] and others used the random parameter logit model with estimated parameter mean heterogeneity to explore the difference between the severity of driver injury and the number of non-occupants. The study showed that human factors and environmental factors were taken into account. The age and gender of passengers have a significant impact on the severity of driver injuries, indicating a potentially complex interaction between passengers and drivers.

In summary, the death rate of drivers caused by single-vehicle accidents is relatively high, which has practical research value. In the same type of bicycle accidents, a considerable part of the accidents was caused by the same factor, and the accident occurrence factors are similar. Therefore, there is a certain correlation between the data of the SV accidents, and the ordered Logit model can solve this type of problem well. Due to the complexity of the factors affecting SV accidents, and multiple logit models are mainly used to deal with the case where the classification response variables are three or more types, the essence is the further expansion of the two-category Logit model, which affects the severity of SV accident factor analysis.

Model variable analysis

This article selects a total of 5480 SV data from a city in China from 2010 to 2019. After excluding incomplete or unreasonable data, the remaining 5012 SV data are used to establish a sample database. Based on the previous research results, 18 relevant variables were selected as the variables to be selected from the model. For the convenience of research, the cluster variables are used to study the data. By establishing an ordered Logit model and a multi-layer ordered Logit model, the relationship between these 18 variables and the severity of SV accidents is explored. In the process of constructing the traffic accident analysis model, the factors considered have been discretized, and the specific model coding is shown in Table 1. The coding of independent variables includes two types of binary classification and multi-categorical variables. Binary classification variables are directly included in the model as the basic variables of the model. For multi-classified variables, in order to further characterize the differential impact of different factors on the severity of the accident, dummy variables need to be introduced, and the introduction method of dummy variables will be explained by taking the time of traffic accident as an example. The specific introduction methods are shown in Table 2.

Table 1: Description of model variables

S. No.	Variable name	Code
1	Cycling accident severity (y)	1 = financial damage accident only; 2 = slight injury accident; 3 = severe injury / fatal accident
2	Driver gender (x_1)	0 = Female; 1 = Male
3	Seat belt / helmet (x_2)	0 = used; 1 = not used

4	Drunk driving (x_3)	0 = No; 1 = Yes
5	Driver's age (x_4)	0=25-59； 1=<25； 2=>59
6	Involving motorcycles (x_5)	0 = No; 1 = Yes
7	Involving trucks (x_6)	0 = No; 1 = Yes
8	Motor vehicle rollover (x_7)	0 = No; 1 = Yes
9	Road surface condition (x_8)	0 = dry, 1 = non-dry
10	Location of road section (x_9)	0 = motor lane; 1 = others lane
11	Collision fixture (x_{10})	0 = No; 1 = Yes
12	Road safety attributes (x_{11})	0 = normal road segment; 1 = hidden road segment; 2 = others
13	Traffic control method (x_{12})	0 = no control; 1 = marking line; 2 = others
14	Weather (x_{13})	0 = good weather; 1 = bad weather
15	Visibility (x_{14})	0=>200m； 1=100m-200m； 2=50m-100m； 3=<50m
16	Accident time (x_{15})	0= morning peak(7:00-10:00);1= midday(10:00-17:00) 2= evening peak(17:00-20:00);3= night(20:00-24:00) 4= early morning(0:00-7:00)
17	Date of accident (x_{16})	0 = working day; 1 = holiday
18	Collision with pedestrians (x_{17})	0 = No; 1 = Yes
19	Administrative division (x_{18})	0 = urban / county town; 1 = rural

Table 2: Dummy variables of the time of traffic accident

SV accident time (x_{16})	**Virtual variable**			
	x_{161}	x_{162}	x_{163}	x_{164}
Morning peak (7:00-10:00)	0	0	0	0
Midday (10:00-17:00)	1	0	0	0
Evening peak (17:00-20:00)	0	1	0	0
Night (20:00-24:00)	0	0	1	0
Early morning (0:00-7:00)	0	0	0	1

Model building

Ordered Logit model

The ordered Logit model is mainly used to deal with multi-categorized response variables, and there is an order relationship between the response categories. Its essence is the further expansion of the binary classification Logit model. Suppose that the model response variable y has J categories ($j=1,2...,j$), and the relationship between the values is (y = j) ＜(y = j + 1), when the level is j The ordered Logit model is expressed as (1):

$$\log it\left[P(y\le j)\right]=\log\left[\frac{P(y\le j|X)}{1-P(y\le j|X)}\right]=\beta_{j0}+\sum_{k=1}^{K}x_k\beta_k=\Delta^{\mathrm{T}}\beta_j \quad (1)$$

Where: $P(y \le j|X)$ is Cumulative probability, and $\sum_{j=1}^{J} P(y=j|X)=1$; X represents a vector of independent variables; β_{j0} is the regression intercept of the jth grade; x_k is the kth independent variable; β_k represents the regression coefficient of the independent variable x_k ; Δ、β_j are the model variable set and regression coefficient set, $\Delta=(1,\ x_1,\ x_2,\ldots,\ x_K)^{\mathrm{T}}$, $\beta_j=(\beta_{j0},\ \beta_1,\ \beta_2,\ldots,\ \beta_K)^{\mathrm{T}}$.
Then the ordered Logit probability model can be expressed as:

$$P(y=1|X)=\frac{\exp(\Delta^{\mathrm{T}}\beta_1)}{1+\exp(\Delta^{\mathrm{T}}\beta_1)} \tag{2}$$

$$P(y=2|X)=P(y\le 2|X)-P(y\le 1|X) \tag{3}$$

…

$$P(y=J|X)=1-P(y\le J-1|X) \tag{4}$$

Multi-layer ordered Logit model

One of the limitations of the ordered Logit model is the assumption of independence, that is, there is no correlation between different levels of data. It is difficult to meet this strict constraint in actual analysis, which will lead to deviations in parameter calibration. Therefore, scholars have proposed a multi-layer ordered Logit model, which overcomes the criticism that ordered logit models require independent sample data. The multi-layer ordered Logit model can be expressed as formula (5):

$$\log\left[\frac{P(y\le j|X)}{1-P(y\le j|X)}\right]=\beta_{j0}+X\beta+ZU \tag{5}$$

Where: β_{j0} is the regression intercept of the *j*-th level; X is the independent variable design matrix with fixed slope vector β represents the level 1 fixed regression coefficient vector; Z is the independent variable design matrix with random slope; U represents the random regression coefficient vector.

Model estimation and testing

Bayesian estimation of the model

The Bayesian inference method is used to calibrate the model parameters to be estimated. Bayesian inference defines the unknown parameters of the model as random variables. Using the population, samples and a priori information, the posterior distribution of the parameters is estimated by the Bayesian formula, so that the posterior distribution is the result obtained by excluding information irrelevant to the model parameters, so it is reasonable to infer the model parameters based on the posterior distribution. On the other hand, the Bayesian estimation method can construct a flexible framework for the model, further improve the model's fitting performance to the sample data, and ensure that the model performs a reasonable regression on the data. In addition, Bayesian analysis needs to set the a priori information of the model parameters, because the historical data that can be directly cited by the research content is less, so this paper uses the a priori probability distribution without information. And take the parameter vector β_j to follow the normal distribution, namely:

$$\beta_j \sim N(0_k, 10^6 \mathrm{I}_k) \tag{6}$$

Where: 0_k means $K\times 1$ order zero vector matrix; I_k represents the identity matrix of order $k\times k$.
According to the Bayesian principle, the posterior distribution of model parameters can be expressed as formula (7):

$$f\left(\beta_j\middle|Y_i\right)\propto f\left(Y_i\middle|\beta_j\right)\pi\left(\beta_j\right)=f\left(Y\middle|\Delta\right)N\left(\beta_j\middle|0_k,10^6\mathrm{I}_k\right) \tag{7}$$

Where: $f\left(\beta_j\middle|Y_i\right)$ represents the posterior probability distribution of the parameter β_j under the given sample Y_i; $f\left(Y_i\middle|\beta_j\right)$ is the model likelihood function; $\pi\left(\beta_j\right)$ is the prior distribution of the model parameter β_j before the given sample data Y_i; $N(\bullet)$ represents the density function of the normal distribution.

Secondly, because the posterior distribution of the model parameter β_j under the normal distribution is more complicated, Bayesian reasoning overcomes the high computational complexity when solving the parameter posterior distribution through Markov Chain Monte Carlo (MCMC). Therefore, when using Bayesian method to calibrate model parameters, it is often done by MCMC algorithm [12].

Model goodness-of-fit test

The goodness-of-fit test of the Bayesian model for the data can be achieved by the variance information criterion (DIC)[13]. DIC is the Bayesian extension of the Akaike information criterion (AIC). The complexity and goodness of fit of the Bayesian model are comprehensively quantified to evaluate the fitting performance of the model, and the fitting of different models can be compared. The calculation method is shown in formula (8):

$$DIC=\overline{D}+pD=2\overline{D}-D(\overline{\theta}) \tag{8}$$

Where: $pD=\overline{D}-D(\overline{\theta})=\sum_j p_j$; $\overline{D}$ is the Bayesian variance of the model parameters, Calculate the deviation of each iteration by MCMC sampling; $D(\overline{\theta})$ represents the deviation of the posterior expected parameters; pD represents the effective number of model parameters.

pD can measure the complexity of the Bayesian model, its essence is the punishment after the model complexity increases which can describe the intra-group correlation of each group through the Bayesian variance of the model parameters and the parameter posterior. The difference between the expected deviations is calculated. The difference between DIC and other fitting evaluations is that the critical value of DIC cannot be determined, but the smaller the value of DIC, the better the model fit. When the DIC difference between the two models is greater than 10, it means that the model with a smaller DIC is significantly better than the one with a larger DIC, and the model with a larger DIC needs to be eliminated; It is believed that the fitting performance of the two models is quite different; and when the difference of DIC is less than 5, it indicates that the fitting performance difference of the two models is small. If the conclusion of the model is significantly different, the model should be further tested It cannot be assumed that the model with a smaller DIC has better fitting performance.

Model parameter calibration

The Bayesian algorithm is used to calibrate the severity of single-vehicle accidents in ordered and multi-layer ordered Logit models. The calibration results are shown in Table 3 and Table 4. And through the Markov chain Monte Carlo (MCMC), the model parameters were simulated 30,000 times, and the model was checked by monitoring the Markov chain Monte Carlo (MCMC) dynamic graph of the model iteration (Figure 1). Of convergence. Once all the parameter values are in the region without strong periodicity, the conclusion of the model convergence can be drawn. It can be seen from Figure 1 that the model parameters have no strong periodicity, and it is considered to be convergent and the model is more reasonable and effective.

Table 3: Parameter calibration of ordered Logit model

Variable	Symbol	Parameter estimation (standard error)	OR
Driver gender	1	-1.373(0.357)	0.266
Drunk driving	1	1.247(0.571)	5.264
Driver's age	2	0.508(0.508)	1.724
Motorcycle	1	1.364(0.404)	4.441
Roll over	1	0.761(0.348)	2.498
Collision fixture	1	0.972(0.320)	2.589
Weather	1	-0.067(0.154)	.948
Time of occurrence	4	0.137(0.076)	1.144
Collision with pedestrians	1	1.696 .806(0.167)	2.247
Intercept 1		-0.768(0.404)	
Intercept 2		1.146(0.411)	
DIC		11474.723	

Table 4: Parameter calibration of multi-layer ordered Logit model

Variable	Symbol	Parameter estimation (standard error)	OR
Driver gender	1	-1.182（0.293）	0.314
Drunk driving	1	2.513（0.642）	4.857
Driver's age	2	0.675（0.413）	2.314
Motorcycle	1	2.849（0.506）	4.387
Roll over	1	0.697（0.416）	1.982
Collision fixture	1	0.845（0.261）	1.978
Weather	1	-.0.103（0.201）	0.765
Visibility	3	-0.151（0.218）	0.845
Time of occurrence	4	0.245（0.103）	1.251
Collision with pedestrians	1	1.861（0.261）	2.375
Date of accident	1	1.564（0.532）	
Intercept 1		1.235（0.314）	
Intercept 2		-1.547（0.512）	
DIC		11281.548	

It can be seen from Table 3 and Table 4 that the DIC value of the multilayer ordered logit model is 11281.548, which is smaller than the ordered logit model DIC value of 11474.723. In addition, there are 9 factors in the ordered logit model that have a significant impact on the SV accident, and The multi-layer ordered Logit model contains 11 factors that have a significant impact on the SV accident, indicating that the multi-layer ordered Logit model has a significantly better fit than the ordered Logit model, and can identify more significant SV accident severity. The influencing factors show that the multi-layer ordered Logit model has good rationality and effectiveness, and the multi-layer ordered Logit model has a good fit for the factors that affect the severity of single-vehicle accidents. Therefore, this paper uses multi-layer ordered Logit parameter calibration results to analyze the factors that have significant impact.

Compared with female drivers and male drivers, the probability of serious traffic accidents is 0.314 times than that of females; the probability of serious traffic accidents caused by drunk driving is 4.857 times than that of non-drink driving. Compared with drivers aged 25-59 years, drivers older than 59 years of age have a 2.314-fold increase in the probability of serious traffic accidents. Traffic accidents involving motorcycles have a 4.387-fold increase in the probability of serious traffic accidents compared to other models. Compared to other types traffic accidents, the probability of traffic accidents involving pedestrians causing serious consequences increased by

2.375. Rollover, collision fixtures, early morning hours, holidays and traffic accidents have a positive correlation; bad weather, visibility less than 50m and traffic accidents have a negative correlation.

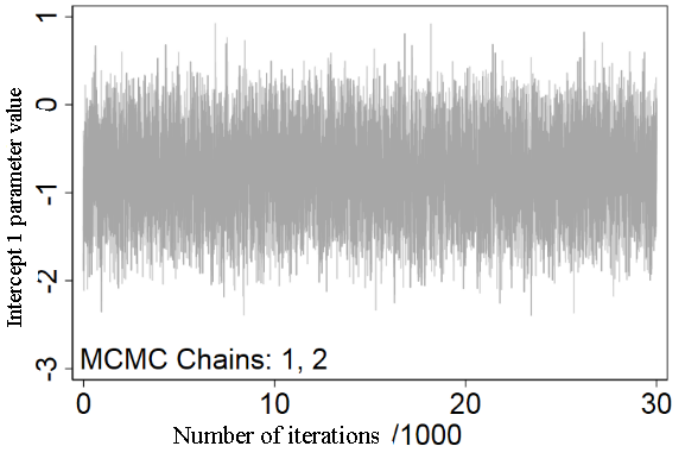

Figure 1: MCMC test of model parameters

Conclusions

Taking SV accident in a city in China as the research object, eighty independent variables such as the driver's gender, age, time of occurrence, weather and so on were selected to analyze the factors affecting the severity of the accident. An orderly Logit model was constructed, and Bayesian algorithm was used to estimate. The study found that the fitting goodness of the multi-layer ordered Logit model is significantly better than the ordered Logit model. Factors such as the driver's gender, age, drunk driving, weather, time of occurrence affect the severity of bicycle accident has a significant impact. To provide a certain degree of theoretical support for the development of reasonable policies to reduce the severity of bicycle accidents in Chinese cities, and it has a certain degree of guiding significance for the study of bicycle accidents in the world.

Acknowledgements

Thanks to the traffic accident data provided by Zibo Traffic Police Detachment, the teachers and classmates who provided strong support during the writing of the paper.

References

[1]. World Health Organization, World health statistics 2016: monitoring health for the SDGs sustainable development goals. 2016.

[2]. Traffic Management Bureau of the Ministry of Public Security of China (TMBMPSC), Annual Statistical Report of China Road Traffic Collisions (2017). Beijing, China, 2018.

[3]. Wen, H.Y., Tang, Z.G. (2019). Cause of Crash Injury Severity in Single-Motorcycle Crash on Roadway Segments. *Journal of Chongqing Jiaotong University(natural science)*, 38(02):117-125. (China)

[4]. Wen, H.Y., Tang, Z.G. (2019). Analysis of Crash Injury Severity in Single-vehicle Crashes Occurring at Intersections. *Highway Engineering*, 44(02):55-61+102. (China)

[5]. Liu, X.X. (2019). Research on Reconstruction of Single Vehicle Collision Accident Based on PC-Crash. *Automotive Practical Technology*, 01:71-72. (China)

[6]. Zhong, C.Y. (2008). The Statistics Analysis on the Single Vehicle Accidents in Expressways. *Chang'an University*. (China)

[7]. Ruth A. S., Gwen B., Tracy J. S., et al (2019). Characteristics of Single Vehicle Crashes with a Teen Driver in South Carolina, 2005–2008 . *Accident Analysis and Prevention*, 122:325–331.

[8]. Zhou M., Hoong C. C. (2019). Factors affecting the injury severity of out-of-control single-vehicle crashesin Singapore [J]. Accident Analysis and Prevention, 124 :104–112.

[9]. Li Z.N., Wu Q., Ci Y.S., et al. (2019). Using latent class analysis and mixed logit model to explore risk factors on driver injury severity in single-vehicle crashes. *Accident Analysis and Prevention*, 129 :230–240.

[10]. Emmanuel K.A., Alexander H., Steven J. (2018). Latent class analysis of factors that influence weekday and weekend singlevehicle crash severities. *Accident Analysis and Prevention*, 113:187–192.

[11]. Ali B., Fred M. (2017). The effect of passengers on driver-injury severities in single-vehicle crashes: A random parameters heterogeneity-in-means approach. *Analytic Methods in Accident Research*,14: 41–53.

[12]. Shaheed M. S., Gkritza K., Carriquiry A. L., et al. (2016). Analysis of occupant injury severity in winter weather crashes: A fully Bayesian multivariate approach. *Analytic Methods in Accident Research*,11:33-47.

[13]. Spiegelhalter D. J., Best N. G., Carlin B. P., et al. (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 2002.4.