Available online www.jsaer.com

***Journal of Scientific and Engineering Research*, 2020, 7(5):191-197**

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Severity Analysis for Single-Vehicle Crashes Based on Mixed Logit Model

Z.G. Cai[1]*, S.S. Hu[1], R.R. Wang[2], Y.Q. Zhou[1]

[1]School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo, Shandong, China
[2]School of Accounting and Finance, Anhui Xinhua University, Hefei, Anhui, China
*Corresponding author: zhenggancai@163.com

Abstract To identify factors that have a significant impact on the severity of single-vehicle crashes, the traffic accident data from 2017 to 2019 in Zibo City was collected to build the Mix logit model. And through the elastic analysis of significant variables to quantify the impact of different factors on the severity of driver injuries. The results show that: (1) The fitting performance of the Mix logit model is better than Multinomial logit model overall, and can identify the heterogeneous influence of different parameters effectively. (2) Driver's gender, age, use of seat belts, drunk driving, fatigue driving, vehicle type, road surface, weather, visibility, time, season have a significant impact on the severity of driver injury in single-vehicle crashes. (3) The gender and age have individual heterogeneity in the impact of the severity for single-vehicle crashes.

Keywords Traffic safety, Traffic accident, Mix logit model; Single-vehicle crashes

Introduction

With the increase of car ownership, the number of traffic accidents has also increased. At least 1.25 million people die from road traffic accidents every year in the world, with an average of at least 3425 deaths per day. Road traffic accidents have become the 9th leading cause of death worldwide [1]. According to statistics, the number of single-vehicle traffic accidents accounts for 6.17% of the total number among different types of traffic accidents. However, the proportion of deaths caused by traffic accidents is as high as 12.06% [2], especially the probability of causing driver death is higher than other types of traffic accidents [3]. In Singapore, for example, the probability of single-vehicle accident that causes a driver's death is 1.7 times than that of a multi-vehicle [4], which has caused huge economic losses and serious impacts on society. Therefore, domestic and foreign scholars have conducted a lot of research to determine the factors that have a significant impact on the occurrence and severity of single-vehicle traffic accidents [5, 6, 7, 8]. For example, female drivers are more likely to be slightly injured and seriously injured in bicycle traffic accidents than male [9], drivers over the age of 65 are more prone to fatal traffic accidents [10]. The probability of a fatal traffic accident for a two-wheeled vehicle is 1.65 times the average level, and the probability of a traffic accident at night is higher than other times [11]. Compared with cities, rural areas are more likely to cause death when bicycle accidents occur [12]. Most of the studies are based on the accident data of foreign countries, due to different driving habits and cultural awareness, there are differences in the traffic operation characteristics between China and other countries, and the research results of accident prevention in other countries are difficult to work in China's traffic safety management directly.

In addition, single-vehicle traffic accidents can be divided into two categories according to the different participants: collision pedestrian accidents and non-collision pedestrian accidents. The single-vehicle accidents of non-collision pedestrians mainly include the damage of vehicles and personnel in vehicles caused by collision with guardrails, piers or other obstacles during the driving process. Among them, traffic accidents involving pedestrians have been extensively studied by scholars because they are easy to cause serious consequences [13, 14, 15]. Some scholars discussed the factors affecting the severity of bicycle accidents from a macro perspective [16], while ignoring the potential impact of pedestrians on traffic accidents. Therefore, in order to clarify the factors that have a significant impact on the severity of traffic accidents of non-collision pedestrians, This study only focuses on the traffic accidents of non-collision pedestrians. In addition, traffic accidents often result in multiple casualties, leading to different standards for evaluating the severity of traffic accidents. In previous studies, there are two methods to classify the severity of accidents: (1) the most severely injured in traffic accidents [17]; (2) the severity of driver injuries [18]. Among them, the former ignores the differences in the probability of injury between drivers and passengers, which may decrease the fitting accuracy. Therefore, this study uses driver injury severity as a measure of accident severity for modeling.

At present, a large number of discrete selection models have been established to study traffic accident [19]. Including ordered logit model [20] and ordered probit model [21]. One of the disadvantages of the ordered regression model is that there is a clear order relationship between the response variables of the model. When this order relationship is not obvious, the scope of the model will be limited, which is beyond doubt. The Multinomial logit model has been widely used in the past few decades because of its complete theory, simple calculation, and overcoming the shortcomings of dependent variable order, which can flexibly deal with the different effects of various factors on different levels of accidents [22, 23]. Although the Multinomial logit model has certain advantages in terms of simple model structure, with the deepening of research, the disadvantages of the Multinomial logit model are becoming more and more obvious. In other words, the Multinomial logit model assumes that the effect of explanatory variables on the severity of traffic accidents is fixed, while ignoring the individual heterogeneity of different factors. Under the influence of these heterogeneity, the influence of the same factor on different traffic accidents may be different, which leads to the reduction of the fitting accuracy of the model, while the Mix logit model overcomes by introducing the random effect of each variable on the severity of traffic accidents. Therefore, this study uses the Mix logit model to study the influencing factors of single vehicle crashes severity. In order to compare the goodness of fit for the model, a Multinomial logit model was also established, and elastic analysis was used to further quantify the impact of different factors on traffic accidents. The research results can provide a theoretical basis for the traffic management department, clarify the focus of accident prevention, and further improve the level of traffic safety.

Data Description

In order to build a single vehicle traffic accident analysis model, 3815 single vehicle accident data occurred in Zibo City in 2017-2019 were retrieved, and 3491 traffic accidents were selected as the sample database after incomplete and unreasonable data were eliminate. According to the severity of driver injury, traffic accidents can be divided into five levels: no injury, slight injury, serious injury, fatal. Considering that only 47 traffic accidents causing driver death in the sample data, which accounts for a low proportion, it may not be enough to provide sufficient samples for model construction. Therefore, the severe injuries and fatal are combined into one category for analysis and recorded as killed and severe injury (KSI) crashes.

Based on the previous studies and the accident information recorded in the sample database, 13 independent variables are selected from four aspects: driver, motor vehicle, road conditions, driving environment and other factors. Among them, driver factors include age, gender, Alcohol or drug, fatigue driving and whether to use safety belt; motor vehicle factors mainly include vehicle type; road condition factors include road surface condition and road line type; driving environment factors include weather, visibility, region, season; other factors include time and week. Table 1 and table 2 show the classification statistics and coding of independent variables and the distribution of continuous variables (max, min, mean, S.D.). It can be seen from table 1 that the variables involved have been discretized, and all variables are unordered classification variables. In the modeling process, the category coded as 0 is used as the reference class. As the basic variable of the model, 2 classification variables can be directly brought into the model, while multi-classification variables have different effects on traffic accidents due to their own particularity. In order to accurately describe the potential impact of different factors on the accident, virtual variables need to be introduced. If there are n(n > 2) categories in the multi-category variables, n-1 virtual variables need to be set in the model. Take the vehicle type as an example to briefly explain the introduction method of virtual variables. The vehicle type consists of four categories, with the ordinary car as the reference, and the specific assignment method is shown in Table 3.

Table 1: Variable coding and descriptive summary

Factor	Variables	Code	Proportion	Factor	Variables	Code	Proportion
Severity	No injury	0	39.8%	Weather	Clear	0	74.1%
	Slight injury	1	32.4%		Overcast	1	15.7%
	KSI	2	27.8%		Snow	2	4.1%
sex	Female	0	27.6%		Rain	3	6.1%
	Male	1	72.4%	Visibility	⩽50m	0	11.3%
year	＞25	0	32.5%		50m-100m	1	14.7%
	≤25	1	67.5%		100m-200m	2	17.8%
Alcohol or drug	No	0	74.9%		⩾200m	3	56.2%
	Yes	1	25.1%	Region	Urban	0	71.3%
Fatigue driving	No	0	79.4%		Rural	1	28.7%
	Yes	1	20.6%	Season	Spring	0	28.4%
Safety belt	Use	0	65.1%		Summer	1	24.3%
	Non-use	1	34.9%		Autumn	2	24.2%
Vehicle type	Private car	0	24.8%		Winter	3	23.1%
	Light truck	1	15.2%	Time	7:00-9:59	0	28.3%
	Heavy truck	2	20.3%		0:00-6:59	1	13.9%
	Motorcycle	3	39.7%		10:00-15:59	2	19.4%
Road surface	Dry	0	85.1%		16:00-18:59	3	23.5%
	Non drying	1	14.9%		19:00-23:59	4	14.6%
Road Line type	Curve	0	26.7%	Weekend	Yes	0	25.0%
	Straight line	1	73.3%		No	1	75.0%

Table 2: Statistical description for continuous variables

Variables	Max	Min	Mean	S.D.
Year	80	15	40	11.52
Time	23	0	13	7.39
Month	12	1	6	3.38

Table 3: virtual variables of vehicle types

Vehicle types	Virtual variables		
	Light truck	Heavy truck	Motorcycle
Private car	0	0	0
Light truck	1	0	0
Heavy truck	0	1	0
Motorcycle	0	0	1

Methodology

Mix logit model

Mixed logit model is developed on the basis of multinomial logit model, also known as random effect model. Using the random effect method, assuming that the severity of the accident *i* is *j*, the random utility can be written as:

$$U_{ij} = x'_{ij}\beta_j + \varepsilon_{ij} \quad (i=1, \cdots, n; \ j=1, \cdots, J) \tag{1}$$

Where U_{ij} is the effect when the severity of the *i*th accident is *j*; x_{ij} represents a set of independent variables; ε_{ij} is the observed disturbance; β_j is the regression coefficient of the independent variable; *n* represents the total number of sample accidents; *J* denote the total number of categories indicating the severity of single-vehicle crashes.

Obviously, the effect of accident *i* on *j* is higher than that of other schemes, If ε_{ij} is assumed to follow a type 1 extreme-value distribution, Then the probability that the severity of accident *i* is *j* can be written as:

$$P(y_i = j \mid x_{ij}) = P(U_{ij} \geq U_{ik}, \forall k \neq j) = \frac{\exp(x_i'\beta_j)}{\sum_{j=1}^{J}\exp(x_i'\beta_j)} \tag{2}$$

By considering the random distribution parameters among various factors, the Mix logit model can be further expressed as:

$$P(y_i = j \mid x_{ij}) = \int P(y_i = j \mid x_{ij}) f\left(\beta_j \mid \varphi\right) d\beta \tag{3}$$

where $f\left(\beta_j \mid \varphi\right)$ is the probability density function of the random vector β_j and φ denotes a vector of parameters describing the probability density function.

Elasticity Analysis

Because the estimated coefficient of the mixed Logit model cannot measure the quantitative relationship between the independent variables and the severity of traffic accidents, for discrete variables, the calculation formula of elasticity is shown in 4:

$$E_{x_{im}}^{P(y_i=j)} = \frac{P(y_i = j \mid x_{im} = 1) - P(y_i = j \mid x_{im} = 0)}{P(y_i = j \mid x_{im} = 0)} \tag{4}$$

Where $E_{x_{im}}^{P(y_i=j)}$ is the elastic coefficient of the *m*-th significant independent variable related to the *j*-th accident level. x_{im} represents the *m*-th significant independent variable related to the *j*-th accident level. It can measure the probability change of accident severity *j* caused by x_m changing from 0 to 1.

Results

Parameter estimation

The parameter estimation of the Mix logit model is implemented using STATA16.0, with a significance level of 0.05.

Table 4: Model parameter estimation

Variables	Coef.	t-Ratio	Elasticity	Variables	Coef.	t-Ratio	Elasticity
Random coefficient				Heavy truck [K]	-0.326	2.45	-0.64
Mean:Year≤25[K]	-1.350	1.87	0.01	Motorcycle [K]	2.544	6.04	2.86
S.D.:Year≤25[K]	2.962	2.45	-	Non-dry road[K]	-0.447	2.38	-8.11
Mean: Male[S]	-0.957	3.74	-14.33	Snow[K]	-1.034	3.65	-8.14
S.D.: Male[S]	0.862	2.52		Rain[S]	0.152	2.15	2.98
Fixed coefficient				⩾200m[S]	-0.235	3.46	-1.27
≤25[S]	0.235	2.07	2.56	50m-100m[S]	-0.186	3.10	0.89
≤25[K]	-0.438	5.49	-13.7	0:00-6:59[K]	0.464	2.98	8.34
Male[N]	-0.506	10.35	-25.81	19:00-23:59[S]	0.441	2.88	9.50
Male[S]	-0.631	11.33	-34.65	Summer[K]	0.107	2.18	3.54
Alcohol/drug[S]	0.541	3.54	2.65	Autumn[S]	0.218	3.07	3.29
Alcohol/drug[K]	1.656	5.90	6.90	Working day[N]	0.102	2.17	0.37
Use seat belts[K]	-1.203	17.54	-35.81	Constant[K]	-5.274	13.54	-
Rural[K]	1.042	5.49	1.85	Constant[N]	-1.873	10.37	-
FD[K]	0.276	3.13	1.04	-			
McFadden R2			0.43	-			
Log likelihood initial value			-14251.65	-			
Log likelihood convergence value			-13384.74	-			

Note: 1. No-injury: N; 2. Slight injury: S; 3. Killed and severity injury(KSI): K; 4. Fatigue driving: FD

In the final model, only significant variables are retained, while other non-significant variables are removed. In the parameter estimation process, it is assumed that all the parameters to be estimated are random parameters, and each parameter is assumed to follow a normal distribution, a uniform distribution and other distribution forms in turn. It was found that only the age variable corresponding to the KSI accident and the gender variable corresponding to minor injuries were random parameters, both of which were in a normal distribution. In terms of the goodness of fit for the model, the McFadden R^2 value is 0.43, indicating that the overall goodness of fit of the model is good. The specific model parameter estimation is shown in Table 4.

In order to compare the goodness of fit, the multinomial logit model was also established. But the Akaike Information Criterion (AIC) shows that the goodness of fit for the two model is not much different. However, the mixed logit model can better identify the heterogeneity of parameters, which is more explanatory than the multinomial logit model, so only the estimated results of the mixed logit model are given in this research.

Result Analysis

Driver Factor

Compared with drivers of other ages, drivers under the age of 25 have a 2.56% increase in the probability of slight injury accidents and a 13.7% decrease in the probability of KSI. In addition, the parameters of young drivers who have KSI accidents obey the normal distribution with the mean value of -1.350 and the standard deviation of 2.962. The cumulative frequency of normal distribution shows that 30.4% of young drivers are more likely to have KSI accidents than those aged over 25.

Compared with female drivers, male drivers have a 25.81% and 34.65% lower probability of no injury and slight injury in single-vehicle accidents. The parameters corresponding to the traffic accidents causing slight injuries to drivers obey the normal distribution with the mean value of -0.957 and the standard deviation of 0.862. The calculation results of the cumulative frequency of the normal distribution show that 19.6% of male drivers are more likely to have minor injuries.

The probability of slight injury and KSI accident caused by drunk driving increased by 2.65% and 6.90% respectively. The use of seat belts can reduce the probability of KSI accidents significantly (35.81%). Fatigue driving increases the probability of a minor injury accident by 1.04%.

Motor Vehicle Factors

Compared with the Private car, the probability of non-injury accident of heavy truck is reduced by 0.64%. The probability of KSI accident of motorcycle is 2.86% higher than that of Private car.

Road and Driving Environment Factors

Compared with the dry road surface, the probability of KSI accident of non-dry road surface is reduced by 8.11%.

Compared with sunny days, the probability of KSI accidents is reduced by 8.14% in snowy days and increased by 2.98% in rainy days.

When the visibility is $\geq$ 200m, the probability of slight injury accidents is reduced by 1.27 compared with that of visibility less than 50m, and the probability of slight injury accidents is increased by 0.89% when the visibility is 50m-100m.

Other Factors

Compared with the weekend, the probability of no-injury traffic accidents increased by 0.37%. The probability of KSI accident in 0:00-6:59 was 8.34% higher than that in 7:00-9:59. The probability of minor injury accidents in the evening (19:00-23:59) was 9.5% higher than that in the morning (7:00-9:59).

The probability of KSI accident in summer is 3.54% higher than that in spring, and the probability of slight injury accident in autumn is 3.29% higher than that in spring.

The probability of KSI accidents in rural areas is increased by 1.85% compared with that in urban areas

Conclusions

Considering the influence of different factors on the driver injuries severity of single-vehicle crashes. Based on the data of single-vehicle accidents in Zibo from 2017 to 2019, and the mix logit model was established. Elastic analysis was used to study different factors quantitatively. The model shows that the Mix logit model can identify the different effects of various factors on accidents effectively.

The model shows that factors such as the gender, age, seat belt use, drunk driving, fatigue driving, vehicle type, road surface, weather, visibility, time, season and other factors have a significant impact on the severity of driver injuries in single-vehicle accidents. The parameters of age and gender follow normal distribution, indicating that male drivers and drivers younger than 25 years have heterogeneous effects on the severity of single-vehicle accidents.

Research can further clarify the impact of different factors on the severity of traffic accidents, provide a theoretical basis for the traffic management department, and further improve the level of traffic safety.

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]. Kim, J. K., Ulfarsson, G. F., Kim, S., & Shankar, V. N. (2013). Driver-injury severity in single-vehicle crashes in California: A mixed logit analysis of heterogeneity due to age and gender. *Accident Analysis & Prevention*, 50: 1073-1081.

[4]. Rifaat, S. M., & Chin, H. C. (2007). Accident severity analysis using ordered probit models. *Journal of Advanced Transportation*, 41 (1): 91–114.

[5]. Moradi, A., Nazari, S. S. H., & Rahmani, K. (2019). Sleepiness and the risk of road traffic accidents: A systematic review and meta-analysis of previous studies. *Transportation Research Part F: Traffic Psychology & Behaviour*, 65: 620-629.

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

[7]. Wu, Q., Zhang, G., Zhu, X., Liu, X. C., & Tarefder, R. (2016). Analysis of driver injury severity in single-vehicle crashes on rural and urban roadways. *Accident Analysis & Prevention*, 94: 35-45.

[8]. Geedipally, S. R., & Lord, D. (2010). Investigating the effect of modeling single-vehicle and multi-vehicle crashes separately on confidence intervals of Poisson-gamma models. *Accident Analysis and Prevention*, 42(4): 1273-1282.

[9]. Hou, Q., Huo, X., Leng, J., & Cheng, Y. (2019). Examination of driver injury severity in freeway single-vehicle crashes using a mixed logit model with heterogeneity-in-means. *Physica A: Statistical Mechanics and its Applications*, 531: 121760.

[10]. Kim, J. K., Ulfarsson, G. F., Kim, S., & Shankar, V. N. (2013). Driver-injury severity in single-vehicle crashes in California: A mixed logit analysis of heterogeneity due to age and gender. *Accident Analysis & Prevention*, 50: 1073-1081.

[11]. Zhou, M. & Chin, H. C. (2019). Factors affecting the injury severity of out-of-control single-vehicle crashes in Singapore. *Accident analysis & prevention*, 124: 104-112.

[12]. Xie, Y., Zhao, K., & Huynh, N. (2012). Analysis of driver injury severity in rural single-vehicle crashes. *Accident Analysis & Prevention*, 47: 36–44.

[13]. Liu, J., Hainen, A., Li, X., Nie Q., & Nambisan S. (2019). Pedestrian injury severity in motor vehicle crashes: An integrated spatio-temporal modeling approach. *Accident analysis & prevention*, 132: 105272.

[14]. Zhai, X., Huang, H., & Sze, N. N. (2019). Diagnostic analysis of the effects of weather condition on pedestrian crash severity. *Accident analysis and prevention*, 122: 318-324.

[15]. Li, Y., & Fan, W. D. (2019). Modelling severity of pedestrian-injury in pedestrian-vehicle crashes with latent class clustering and partial proportional odds model: A case study of North Carolina. *Accident Analysis & Prevention*, 131: 284-296.

[16]. Li, Z., Wu, Q., Ci, Y., Chen, C., Chen, X., & Zhang, G. (2019). Using latent class analysis and mixed logit model to explore risk factors on driver injury severity in single-vehicle crashes. *Accident Analysis & Prevention*, 129: 230-240.

[17]. Ahmed, M. M., Franke, R., Ksaibati, K., & Shinstine, D. S. (2018). Effects of truck traffic on crash injury severity on rural highways in Wyoming using Bayesian binary logit models. *Accident Analysis & Prevention*, 117: 106-113.

[18]. Li, Z., Ci, Y., Chen, C., Zhang, G., Wu, Q., Qian, Z., Prevedouros, P. D., & Ma, D. T. (2019). Investigation of driver injury severities in rural single-vehicle crashes under rain conditions using mixed logit and latent class models. *Accident analysis & prevention*, 62: 219–229.

[19]. Savolainen, P. T., Mannering, F. L., Lord, D., & Quddus, M. A. (2011). The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. *Accident Analysis & Prevention*, 43: 1666–1676

[20]. Kim, M., Kho, S. Y., & Kim, D. K. (2017). Hierarchical ordered model for injury severity of pedestrian crashes in South Korea. *Journal of Safety Research*, 61: 33–40.

[21]. Xu, X., Huang, D., & Guo, F. (2019). Addressing spatial heterogeneity of injury severity using Bayesian multilevel ordered probit model. *Research in Transportation Economics*, 24: 100748.

[22]. Chen, Z., & Fan, W. (2019). A multinomial logit model of pedestrian-vehicle crash severity in North Carolina. *International Journal of Transportation Science & Technology*, 8: 43-52.

[23]. Vajari, M. A., Aghabayk, K., Sadeghian, M., & Shiwakoti, N. (2020). A multinomial logit model of motorcycle crash severity at Australian intersections. *Journal of Safety Research*, 73: 17-24.