Available online www.jsaer.com

***Journal of Scientific and Engineering Research*, 2020, 7(5):206-211**

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Analysis of Driver-Injury Severity in Rural Single-Vehicle Crashes Based on Multinomial Logit Model

Z.G. Cai[1]*, S.S. Hu[1], R.R. Wang[2], Y.Q. Zhou[1], Z. Zhang[3], W.Z. Li[4]

[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
[3]School of Mining, China University of Mining and Technology, Xuzhou, Jiangsu, China
[4]School of Science, Nanjing University of Science and Technology, Nanjing, Jiangsu, China
*Corresponding author: zhenggancai@163.com

Abstract To identify the factors that have a significant impact on the severity of single-vehicle(SV) accidents in rural areas. The data of SV crashes during 2017-2019 in rural was collected. The factors such as sex, age and accident type are selected as explanatory variables for the model, and the non-injury, slight injury, killed and serious injury(KSI) are selected as model response variables to establish multinomial logit models, and the maximum likelihood method(MLE) was used to estimate the model parameters. The result shows that: at the 5% significance level, male, age≤25, age≥59, alcohol, motorcycle, colliding fixer, night, roller, et al. are positively correlated with the severity of traffic accidents. And summer, autumn, winter, intersection, use seat belt/helmet, rush hour can reduce the severity of traffic accidents. The research can provide theoretical basis for traffic managers, so as to further reduce the severity of crashes and improve traffic safety.

Keywords Traffic safety, The multinomial logit model, MLE, SV crashes

Introduction

In recent years, with the increase of the number of traffic accidents, especially the number of fatal accidents, the traffic operation efficiency and driving safety have been seriously harmed. In 2017, there were 2064 traffic accidents causing casualties in Zibo City, including 578 SV accidents. Among all SV accidents, the number of accidents in rural areas accounts for 28.7% of the total number, but the proportion of accidents causing death is as high as 35.8% [1]. It can be seen that SV traffic accidents in rural areas have caused huge economic losses and serious impact on society, which can't be ignored in the study of traffic safety. Therefore, it is necessary to analyze the rural SV accident to identify the factors that have a significant impact on the severity of the accident, so as to further clarify the focus of accident prevention and effectively reduce the probability of traffic accidents, especially the probability of fatal traffic accidents.

At present, the SV accident in rural areas is easy to cause casualties, which has been widely concerned by scholars [2, 3], For example, speeding will lead to a significant increase in the probability of serious traffic accidents [6]; rainfall during the crash is positively associated with single-vehicle crashes [7]; Kim, et al. [4] studied the SV crashes in California, and found that the driver characteristics (gender, age, use of seat belt, drunk driving), environmental characteristics (road surface condition, lighting conditions) had a significant impact on the severity of the accident; Wu, et al. [5] used the accident data of New Mexico to study the impact of different factors on the severity of crashes, and found that factors such as collision fixtures, rollovers, and drunk driving will significantly increase the severity of driver injuries. At the same time, due to the differences in the types of data used by different scholars in the research process, the research results are not always comparable. For example, some scholars study the impact of trucks on the severity of crashes [8,11], while others study the impact of motorcycles on traffic accidents [9, 10]. In addition, even if the research theme is the same, different scholars will focus on different points. For example, Wu, et al. [12] studied the impact of heavy fog on traffic accidents; El-Basyouny, et al. [13] studied the impact of time and weather on crash types; Hou, et al. [14] studied the impact of climbing lane on crash frequency. But most scholars use foreign traffic accident data in the research process, and there are differences in traffic operation conditions between different countries, making it difficult for their research results to be transplanted into traffic safety management in China directly. Therefore, it is necessary to use the SV data in rural of China to study the factors that have a significant impact on the severity of traffic accidents. In addition, in the selection of modeling methods, this study uses the multinomial logit models [15, 16] for modeling analysis to study the impact of different factors on the severity of crashes.

DATA

We retrieved the data of SV accidents in rural of Zibo from 2017 to 2019. After excluding the sample data containing abnormal values, a total of 1054 traffic accidents were modeled as samples. When evaluating the severity level of traffic accidents, considering the difference of injury probability between drivers and passengers, in order to unify the standard, we take the severity of driver injury as the severity of traffic accidents. In addition, the database divides the severity of driver into four levels (no injury, slight injury, serious injury, killed), in which the proportion of accidents leading to death of driver is relatively low, which provides enough training samples for the model. In the modeling, serious injury and death are regarded as a kind of factor and recorded as killed and severity (KSI). In the modeling process, the severity of driver injury severity is taken as the response variable, and the gender, age, drunk driving, speeding, vehicle type, road surface condition, collision fixture, weather, visibility, occurrence time, peak hour, et al are taken as the explanatory variables. The specific variable analysis and coding is as follows.

Table 1: Variables considered for the model

Number	Variables	Code(Number)
1	SV severity	1=No injury(384); 2=Slight injury(379); 3=KSI(290)
2	Sex	0=Female(282); 1=Male(772)
3	Year	0=25-59(872); 1≤25(118); 2≥59(62)
4	Alcohol	0=No(907); 1=Yes(147)
5	Speeding	0=No(847); 1=Yes(206)
6	Vehicle type	0=Private car(240); 1=Truck(354); 2=Motorcycle(371); 3=Others(87)
7	Road surface	0=Dry(903); 1=Non-dry(150)
8	Colliding fixer	0=No(822); 1=Yes(231)
9	Weather	0=clear(880); 1=Non-clear(173)
10	Visibility	0=$>100m$(511); 1=$\leq 100m$(542)
11	Time	0=Day(405); 1=Night(648)
12	Weekend	0=No(741); 1=Yes(312)
13	Use seat belt/helmet	0=No(504); 1=Yes(549)
14	Season	0= Spring(274); 1=Summer(265); 3=Autumn(277); 4=Winter(237)
15	Intersection	0=No(835); 1=Yes(218)
16	Place	0=Highway(876); 1=Path(177)
17	Rollover	0=No(885); 1=Yes(168)
18	Road type	0=Curve(201); 1=Others(852)
19	Rush hour	0=No(741); 1=Yes(312)

Multinomial Logit model

Using the random effect method, suppose that the severity of the accident i is j, and the random effect can be written as:

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

Where U_{ij} is the effect when the severity of the *i*-th accident is *j*; x_i 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 SV crashes.

Assuming that ε_{ij} is independently identically distribution(IID) and obey the type I extreme value distribution, then it can be proved that the probability of the severity of *i*-th accident is j can be written as:

$$P(y_i = j|x_i) = \frac{\exp(x_i'\beta_j)}{\sum_{j=1}^{J}\exp(x_i'\beta_j)} \tag{2}$$

Obviously, the probability sum of the accident severity is 1, that is $\sum_{j=1}^{J} P(y_i = j|x_i) = 1$. because the model can't recognize all the coefficients β_j, (j=1,…,J) at the same time. Therefore, a scheme (J=1 in this paper) is usually taken as the base category, and then the corresponding coefficient $\beta_1 = 0$. Therefore, when the severity of the *i*-th accident is *j*, the probability can be further expressed as:

$$P(y_i = j|x_i) = \begin{cases} \dfrac{1}{1+\sum_{j=2}^{J}\exp(x_i'\beta_j)} (j=1) \\ \dfrac{\exp(x_i'\beta_j)}{1+\sum_{j=2}^{J}\exp(x_i'\beta_j)} (j=2,\ldots,J) \end{cases} \tag{3}$$

In the model, maximum likelihood estimation (MLE) was used to estimate the parameters. The likelihood function of the *i*-th accident is as follows:

$$L_i(\beta_1,\ldots,\beta_J) = \prod_{j=1}^{J}[P(y_i = j|x_i)]^{1(y_i=j)} \tag{4}$$

Then the log-likelihood function is:

$$\ln L_i(\beta_1,\ldots,\beta_J) = \sum_{j=1}^{J} 1(y_i = j)\cdot \ln P(y_i = j|x_i) \tag{5}$$

In the formula, $1(\cdot)$ is the indicator function, and the value is 1 if the expression in parentheses holds, otherwise, the value is 0. We can get the log-likelihood function of the whole sample through sum up the log-likelihood functions of all individuals, and maximize it to get the coefficient estimate $\hat{\beta}_1,\ldots,\hat{\beta}_J$.

Result Analysis

Parameter Estimation

The model parameter estimation is implemented by STATA16.0 software, and the (MLE) was used to estimate the model parameters, with a significance level of 5% to test the significance of the explanatory variables on the response variables. The final model contains only factors that have a significant impact on the severity for the SV crashes, while other factors are not included in the model. The specific model parameter estimation results are shown in Table 2. The model shows that factors such as sex, year, Alcohol, vehicle type, colliding fixer, time, unused seat belt, season, intersection, rollover, peak hours, etc. have a significant impact on the severity for SV traffic accidents.

Table 2: Model parameter estimation

Variable	Slight injury				KSI			
	Coef.	**S.E.**	**z**	**RR**	**Coef.**	**S.E.**	**z**	**RR**
Male	-	-	-	-	0.305	0.395	1.05	1.36
Age≤25	0.533	0.303	1.76	1.70	-	-	-	-
Age≥59	-	-	-	-	1.229	0.454	0.50	3.15
Alcohol	0.863	0.364	2.37	2.37	0.999	0.363	2.75	2.72
Use seat belt/helmet	-	-	-	-	-1.305	0.395	-1.05	0.36
Motorcycle	2.917	0.310	9.40	18.49	3.645	0.314	11.58	38.28
Colliding fixer	0.706	0.241	2.93	2.03	1.246	0.239	5.20	3.48
Night	-	-	-	-	0.557	0.229	2.43	1.74
Summer	-	-	-	-	-0.575	0.258	-2.23	0.56
Autumn	-	-	-	-	-0.615	0.254	-2.42	0.54
Winter	-	-	-	-	-0.531	0.258	-2.06	0.59
Intersection	-0.568	0.233	-2.43	0.57	-	-	-	-
Rollover	-	-	-	-	1.364	0.253	5.40	3.91
Rush hour	-0.536	0.215	-2.49	0.59	-	-	-	-
Intercept[KSI]	-	-	-	-	-1.614	0.575	-2.81	0.20

Driver factor

A. The probability of KSI caused by a male driver in SV accident is 1.36 times than that of female driver; B. Young drivers (≤25) who are more likely to have slight injuries are 1.7 times than that of middle-aged drivers (25-59); older drivers (≥59) who are more likely to have KSI traffic accidents are 3.15 times than that of middle-aged drivers (25 -59); C. The probability of lightly injured in SV traffic accident for a driver under the influence of alcohol is 2.37 times than that of a normal driver, and the probability of KSI traffic accident is 2.72 times than that normal driver. Obviously, drunk driving has a serious impact on traffic safety. We should strengthen the supervision of drunk driving and put an end to drunk driving. D. The use of seat belts or helmets can reduce the probability of KSI accidents significantly

Types of motor vehicles and accidents

A. The probability of a light injury and KSI for a driver in a motorcycle traffic accident is higher than that in private car. The probability of a KSI traffic accident on a motorcycle is 38.28 times than that of private car; the probability of motorcycle's slight injury traffic accident is 18.49 times higher than that of private car. This is because the motorcycle is limited to its own disadvantages, so that the drivers are exposed and are unable to provide the necessary protection and safety measures [17, 18]. At the same time, motorcycle accidents are prone to rollover, fall and other phenomena, which will cause serious injury to the driver. B. The probability of colliding fixer causing the driver to be slightly injured and KSI are 2.03 and 3.48 times than that of the collision of non-fixed objects respectively. C. In the event of an accident, the probability of a Rollover phenomenon causing the driver 's KSI is 3.91 times than that of a non-rollover.

Other influencing factors

A. The probability of KSI accident at night is 1.74 times than that of the day. This may be due to the poor visibility at night, the driver is easy to ignore the roadside obstacles. In addition, the traffic control is relatively weak, and there are many violations of traffic rules, resulting in serious traffic accidents. B. Season has a significant impact on the severity of SV accidents. The probability of KSI crashes in summer is 0.56 times than that of spring, the probability of a KSI accident in autumn is 0.54 times than that of spring, and the probability of KSI accident in winter is 0.59 times than that of spring. C. The probability of slight injury due to SV accident at Intersection is lower than that at a non-intersection. D. The probability of causing slight injuries to the driver during a SV accident during peak hours is lower than that during off-peak hours.

Goodness of fit test

After the model parameter estimation is completed, the goodness of fit for the model needs to be tested. The specific model test results are shown in Table 3. The value of Pseudo R^2 is 0.24, LR statistics is 487.93, corresponding P-value is less than 0.001, so the joint significance of all coefficients for the model meets the requirements.

Table 3: Goodness of fit test of the model

Number of observations	1053	P-value	＜0.001
LR statistics	487.93	Pseudo R^2	0.24
Degrees of freedom	32	Log likelihood	-908.15

The test for Independence from Irrelevant Alternatives(IIA)

As the precondition for the establishment for the multinomial logit model is to meet the IIA hypothesis, it is necessary to test the IIA hypothesis. The specific model verification results are shown in Table 4. The results of the Hausman test shows that the IIA hypothesis will not be rejected if any of the slight injury and KSI are omitted, and the IIA hypothesis will not be rejected if the minor injury accident is used as the reference category for the non-injury accident too. Therefore, the IIA hypothesis test is passed.

Table 4: Hausman tests of IIA assumption

Injury severity	χ^2-Value	*p*-Value	Null hypothesis	IIA property
No-injury	-0.16	1.00	Fail to reject	Holds
Slight-injury	0.84	1.00	Fail to reject	Holds
KSI	2.17	1.00	Fail to reject	Holds

Conclusion

In this paper, the multinomial logit model was constructed based on the research object of SV traffic accidents in rural areas. And the effects of gender, age, drunk driving, speeding, vehicle type, road surface condition, collision fixture, weather, visibility, time, peak hour, etc. on the severity for SV accident were analyzed. The modeling result shows that: Male, age≤25, age≥59, alcohol, use seat belt/helmet, motorcycle, colliding fixer, night, rollover and other factors are positively correlated with the severity of traffic accidents. And summer, autumn, winter, intersection, use seat belt/helmet, rush hour can reduce the severity of traffic accidents.

We only built the multinomial Logit model in this paper, but how other types of crash analysis models (such as mix logit model, nested logit model) fit the accident data needs further study.

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

[2]. Russo, F., Pace, R. D., Dell'Acqua, G., & De-Luca, S. (2017). Estimating an injury crash rate prediction model based on severity levels evaluation: the case study of single-vehicle run-off-road crashes on rural context. *Transportation Research Procedia*, 27, 1088-1096.

[3]. Li, Z., Ci, Y., Chen, C., Zhang, G., Wu, Q., & Qian, Z. (2019). Investigation of driver injury severities in rural single-vehicle crashes under rain conditions using mixed logit and latent class models. *Accident Analysis & Prevention*, 124, 219-229.

[4]. 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.

[5]. 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.

[6]. Xie, Y., Zhao, K., Huynh, N. (2012). Analysis of driver injury severity in rural single-vehicle crashes. *Accident Analysis & Prevention*, 47, 36-44.
[7]. Rusli, R., Haque, M. M., King, M., & Voon, W. S. (2017). Single-vehicle crashes along rural mountainous highways in Malaysia: an application of random parameters negative binomial model. *Accident Analysis & Prevention*, 102, 153-164.
[8]. Azimi, G., Rahimi, A., Asgari, H., Jin, X. (2020). Severity analysis for large truck rollover crashes using a random parameter ordered logit model. *Accident Analysis & Prevention*, 2020, 135, 105355.
[9]. Murphy, P., Morris, A. (2020). Quantifying accident risk and severity due to speed from the reaction point to the critical conflict in fatal motorcycle accidents. *Accident Analysis & Prevention*, 141, 105548.
[10]. Xin, C., Wang, Z., Lee, C., Lin, P. S., Chen, T., & Guo, R. (2019). Development of crash modification factors of horizontal curve design features for single-motorcycle crashes on rural two-lane highways: a matched case-control study. *Accident Analysis & Prevention*, 123, 51-59.
[11]. Rahimi, E., Shamshiripour, A., Samimi, A., Mohammadian, A. K. (2020) Investigating the injury severity of single-vehicle truck crashes in a developing country. *Accident; analysis & prevention*, 137, 105444.
[12]. Wu, Y., Abdel-Aty, M., & Lee, J. (2018). Crash risk analysis during fog conditions using real-time traffic data. *Accident Analysis & Prevention*, 114, 4-11.
[13]. El-Basyouny, K., Barua, S., Islam, M. T. (2014). Investigation of time and weather effects on crash types using full Bayesian multivariate Poisson lognormal models. *Accident analysis & prevention*, 73, 91-99.
[14]. Hou, Q., Meng, X., Huo, X., Cheng, Y., & Leng, J. (2019). Effects of freeway climbing lane on crash frequency: application of propensity scores and potential outcomes. *Physica A: Statistical Mechanics and its Applications*, 517, 246-256.
[15]. 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.
[16]. Chen, Z., & Fan, W. (2018). A multinomial logit model of pedestrian-vehicle crash severity in North Carolina. *International Journal of Transportation Science & Technology*, 8, 43-52.
[17]. Rifaat, S. M., Tay, R., de-Barros, A. (2012). Severity of motorcycle crashes in Calgary. *Accident analysis & prevention*, 49(1), 44–49.
[18]. Bella, F., Calvi, A., D'Amico, F. (2012). Impact of pavement defects on motorcycles' road safety. *Procedia-Social and Behavioral Sciences*, 53, 943-952.