



Article Modeling of Low Visibility-Related Rural Single-Vehicle Crashes Considering Unobserved Heterogeneity and Spatial Correlation

Zhenggan Cai¹, Fulu Wei^{1,2,*}, Zhenyu Wang², Yongqing Guo¹, Long Chen¹ and Xin Li¹

- ¹ School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo 255000, China; zhenggancai01@gmail.com (Z.C.); yongqing.guo@sdut.edu.cn (Y.G.); clong11131024@gmail.com (L.C.); lixin.sdut@gmail.com (X.L.)
- ² Center for Urban Transportation Research, University of South Florida, Tampa, FL 33620, USA; zwang9@cutr.usf.edu
- * Correspondence: weifulu@sdut.edu.cn

Abstract: Accident analysis and prevention are helpful to ensure the sustainable development of transportation. The aim of this research was to investigate the factors associated with the severity of low-visibility-related rural single-vehicle crashes. Firstly, a latent class clustering model was implemented to partition the whole-dataset into a relatively homogeneous sub-dataset. Then, a spatial random parameters logit model was established for each dataset to capture unobserved heterogeneity and spatial correlation. Analysis was conducted based on the crash data (2014–2019) from 110 two-lane road segments. The results show that the proposed method is a superior crash severity modeling approach to accommodate the unobserved heterogeneity and spatial correlation. Three variables—seatbelt not used, motorcycle, and collision with fixed object—have a stable positive correlation with crash severity. Motorcycle leads to a 12.8%, 23.8%, and 12.6% increase in the risk of serious crashes in the whole-dataset, cluster 3, and cluster 4, respectively. In the whole-dataset, cluster 2, and cluster 3, the risk of serious crashes caused by seatbelt not used increased by 5.5%, 0.1%, and 30.6%, respectively, and caused by collision with fixed object increased by 33.2%, 1.2%, and 13.2%, respectively. The results can provide valuable information for engineers and policy makers to develop targeted measures.

Keywords: traffic safety; latent class clustering; spatial correlation; single-vehicle crashes; heterogeneity

1. Introduction

The impact of traffic crashes on sustainable development cannot be ignored because accidents will cause significant property damage and personal injury. The analysis of traffic crashes can provide targeted measures to improve traffic safety performance and promote sustainable development. Among various crash types, single-vehicle (SV) crashes account for a high fatality rate; according to the National Highway Traffic Safety Administration, the number of SV crashes and fatalities accounted for 16% and 36.9% of the respective totals for all crashes [1,2]. The impact of SV crashes on road safety is daunting. This phenomenon is particularly obvious in rural areas of China because road infrastructure and medical assistance in these areas are worrisome. The number of fatalities caused by rural SV crashes increased from 2013 to 2017, showing an average growth rate of 4% [3,4]. Therefore, analysis on the severity of rural SV crashes has aroused widespread interest from transportation professionals [5,6]. A primary objective of this research is to clarify the relationship between crash severity and various risk factors [7], and it has been generally recognized that weather has a significant influence on crash severity [8].

Severe weather can affect the traffic environment and is more likely to result in serious crashes [9,10]. In particular, low visibility caused by fog and haze has been considered as a



Citation: Cai, Z.; Wei, F.; Wang, Z.; Guo, Y.; Chen, L.; Li, X. Modeling of Low Visibility-Related Rural Single-Vehicle Crashes Considering Unobserved Heterogeneity and Spatial Correlation. *Sustainability* **2021**, *13*, 7438. https://doi.org/ 10.3390/su13137438

Received: 20 May 2021 Accepted: 30 June 2021 Published: 2 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). crash-concentration environment [11]. Such condition will affect the judgment of traffic obstacles, which has a negative influence on traffic safety [12,13]. In the existing research on rural SV crashes, low visibility has been exhibited as a variable of weather factors; there is a lack of separate research on rural SV crashes under such conditions. For example, Wen et al. [14] investigated the risk factors of SV crash severity in mountainous areas and suggested that compared with sunny weather, the probability of serious and fatal crashes in foggy weather increased by 21.2% and 15.7%, respectively. Visibility conditions were divided into four categories—greater than 200 m, 100 –200 m, 50 –100 m, and less than 50 m—to evaluate the safety performance of rural areas. It was found that the probability of fatal crashes is substantially increased in visibility less than 100 m [15].

Further, research on rural SV crashes under specific inclement weather mainly has focused on rainfall and snowfall. A mixed logit model was established using rain-related rural SV crashes. The regression results show that the influence of male drivers on crash severity exhibited a random effect [5]. Based on the study of snow-related rural SV crashes, it was found that older drivers were 101.7% more likely to be involved in a serious crash compared to middle-aged drivers [6]. Recently, a preliminary analysis of SV crashes associated with low visibility was conducted [16]; however, the research did not separately analyze rural crashes. In urban and rural areas, some substantial differences in crash characteristics have been found [17,18]; thus modeling results cannot be transferred directly [9]. Hence, it is necessary to investigate the effect of contributing factors on the severity of low-visibility-related rural SV (LVR-SV) crashes to provide trustworthy measures.

The rest of this paper is organized as follows. Section 2 presents a comprehensive background review for rural SV crash analysis. Details of the collision dataset are described in the subsequent section. The model framework of this research is exhibited in Section 4. The analysis and discussion of the regression results are given in Section 5. Conclusion of this paper is given in Section 6. The potential limitations, together with future work, are described at the end.

2. Background

2.1. Safety Covariates of Rural Single-Vehicle Crashes

Various risk factors will affect the severity of rural SV crashes, and corresponding research was conducted to clarify the relationship between risk factors and crash severity [8,9]. According to the characteristics of risk factors, three components can be roughly identified: driver characteristics, crash characteristics, and environmenta characteristics.

There is a significant correlation between driver characteristics and rural SV crash severity, which has been widely recognized by transportation professionals. Several driver-related risk factors, such as gender, age, seatbelt usage, drunk driving, and speeding, were investigated in the existing literature. For example, a positive correlation between male driver and serious crashes was revealed [6]. This phenomenon is related to the aggressive driving behavior of male drivers. The study also noted that older and younger drivers were significantly more likely to have severe SV crashes than middle-aged drivers. Older and younger drivers are susceptible to distracted driving [19]. The impact of distracted driving on crash risk was analyzed, and a positive correlation between distracted driving and crash risk was found [20]. Further, the importance of using seatbelts and maintaining a safe speed has been widely emphasized, as they can effectively prevent drivers from being involved in fatal crashes [14,21]. A latent class logit function was established, and three risk variables—seatbelt not used, driving under the influence of alcohol, and fatigue driving—are expected to result in a significant increase in the probability of severe crashes [22].

Crash characteristics mainly include two components: crash type and vehicle type. Specifically, three crash types—collision with fixed object, run-off-road crashes, and collision with animals—have an important impact on the severity of SV crashes [23]. It was found that these variables are negatively correlated with crash severity. However, this finding could not be supported by Haq et al. [24]. A hierarchical Bayesian random intercept approach was established to link collision with fixed object and crash severity. The regression results show that (1) collision with fixed object will cause a significant increase in the possibility of serious crashes and (2) one of the serious consequences of collision with fixed object is rollover, which will increase the risk of serious crashes in some extent [25].

In addition, existing studies have found that the severity of rural SV crashes varies across vehicle types. Three vehicle types—passenger vehicle, motorcycle, and large vehicle—were classified, and the large vehicle was used as a reference variable to establish an ordered logistic regression. The results indicate that compared with large-vehicle crashes, the risk of serious crashes caused by passenger vehicle reduced significantly (odds ratio was 0.553) and caused by motorcycle increased significantly (odds ratio was 3.907) [26]. Considering the substantial differences between different vehicle types, cross-vehicle-types modeling was recommended. For example, a crash severity regression function related to motorcycle and truck crashes was established, respectively, to explore the risk factors [27,28]. Establishing a regression model across vehicle types can obtain targeted findings and improve the model fit; this may be due to the fact that a crash dataset containing a specific vehicle type has more homogeneity compared with a crash dataset containing all vehicle types [29].

Many significant risk variables are included in the environmental characteristics [30], such as road surface, weather, light condition, and intersection, etc. Previous studies have found that there is a negative correlation between adverse road surface conditions and crash severity [23]. It indicates that the reduction in crash severity under adverse conditions is due to drivers being more cautious. A more detailed finding related to adverse road surface was drawn by Yu et al. [10]. Compared with dry road surface, the probability of fatal crashes caused by wet, snow, and ice is reduced by 2.73%, 0.85%, and 2.78%, respectively. A severity prediction function was established by Hou et al. [31] to investigate the influence of environmental factors on crash severity. It was found that the probability of severe injury is expected to be significantly reduced under the conditions of visibility less than 200 m and wet road surface. Meanwhile, both roads with intermediate barriers and dark unlighted conditions will cause a significant increase in the tendency of serious crashes. Further, illumination conditions were divided into five categories to establish a logistic regression. Compared with the daylight condition, the probability of serious collision under dusk, dawn, dark lighted roadway, and dark unlighted roadway conditions increased by 1.57 times, 2.02 times, 1.63 times, and 2.61 times, respectively [32]. This shows that the possibility of serious injury in the early morning is of the most concern, and the provision of street lighting can reduce the occurrence of serious traffic collisions.

2.2. Statistical Techniques for Unobserved Heterogeneity and Spatial Correlation

Unobserved heterogeneity has been recognized as a primary issue in crash severity analysis [33–35] because several risk factors affecting the possibility or severity of the crash cannot be observed [36,37]. Such fact has driven the development of statistical techniques to identify the unobserved heterogeneity. To accommodate the discrete nature of crash severity (no injury, slight injury, serious injury, and fatality), various regression approaches—random parameters logit (RP-logit) model [38,39], random parameters probit model [40], random intercept logit model [41], latent class logit model [10], and finite mixture random parameters model [16,42]—have been widely recommended due to their high flexibility [43–45]. Alternatively, random parameters ordered logit model [46] and random parameters ordered probit model [47] were applied to handle the intuitive ordering of crash severity. For example, Wu et al. [8] established the RP-logit model to analyze the risk factors of single- and multi-vehicle crash severity on rural highways. The results show that the RP-logit model can accommodate the unobserved heterogeneity satisfactorily, and some substantial differences in the risk factors between urban and rural were clarified.

Rural SV crash data in the United States were extracted to calibrate the RP-logit model and the latent class logit model. The regression results indicate that both of these statistical models can effectively identify risk factors, but the latent class logit model has a slightly better fit performance than the RP-logit model [5]. Similar findings were obtained by Cerwick et al. [48]. A hierarchical random intercept function was proposed to accommodate cross-level interaction in crash data. The regression results indicate that the fitting performance of the proposed function is comparable to the RP-logit model, but its generalization is expected to be limited by high complexity [49]. Meanwhile, a random effect tobit model and a random parameters tobit model were established to accommodate the unobserved heterogeneity in crash rate analysis. The importance of capturing the unobserved heterogeneity and the superiority of the random parameters function were confirmed [50].

In addition, most crash severity functions were developed using a whole-dataset; however, this modeling approach has some drawbacks. The crash analysis model can identify unobserved heterogeneity, but it cannot reduce them. In addition, extensive unobserved heterogeneity may cause the regression results to deviate from the real situation [34,51]. If the whole-dataset is divided into several sub-datasets, called clustering, and both the homogeneity effect within the sub-dataset and the heterogeneity effect between the sub-datasets are maximized, then the severity prediction functions are established for each sub-dataset. This two-step modeling approach is expected to mitigate the impact of unobserved heterogeneity on model estimation.

Currently, the two-step analysis approach is receiving increasing attention. A latent class clustering (LCC) model and an RP-logit model were combined to investigate motorcycle crash severity. It was found that the LCC model is an efficient crash data clustering technology and can effectively reduce the unobserved heterogeneity in the sub-dataset [52]. Based on crash data from New Mexico, Li et al. [53,54] modeled not only intersection-related crashes but also SV crashes and found that the two-step approach can improve the fitting performance of the severity prediction function. The severities of rain-related rural SV crashes and head-on crashes were investigated respectively, and both of them suggested that the two-step analysis approach can reveal more valuable risk factors [5,55]. A similar conclusion can be found in Yu et al. [10].

Further, the severity prediction functions mentioned above can not only provide valuable risk factors but also effectively identify unobserved heterogeneity in crash data. However, these functions can hardly identify spatial correlation across crashes. The existence of spatial correlation is reasonable. Some characteristics, such as road width and road surface conditions, may be shared in adjacent crashes. This phenomenon has been confirmed by much research [56–58]. Yang et al. [59] considered the spatial correlation of crashes in road safety assessment; the study focused on investigating the interrelationship among crash frequencies. The results show that the model simultaneously considering both spatial correlation and unobserved heterogeneity outperforms the model considering only the unobserved heterogeneity. Risk factors of freeway crashes were explored using a spatial generalized ordered logit model. It was found that the model fit can be improved by accounting for spatial effect and the spatial error term can be effectively estimated by introducing the Gaussian conditional autoregressive (CAR) technique [60]. Klassen et al. [61] obtained the same conclusion by proposing a spatial random intercept logit model and pointed out that ignoring spatial correlation may lead to biased inferences. To examine pedestrian injury severity in bicyclist-pedestrian crashes, a Gaussian CAR spatial Poissonlognormal model was adopted. The regression results highlight the effectiveness of jointly modeling multiple crash severities to improve fit performance [62].

The spatial functions mentioned above are evolved from traditional regression models by considering a spatial error term, including the multinomial logit model, ordered logit model, random intercept logit model, etc. The spatial function evolved from random parameters models is less common. Several studies have proved that the random parameters model outperforms the fixed-parameter model and the random intercept model. Identifying the spatial correlation by extending the random parameters model is expected to exhibit satisfactory fitting performance. Recently, a spatial random parameters Poisson-lognormal model, a spatial random parameters tobit model, and a spatial random parameters Poissonlognormal model with a mixture component were proposed to capture the spatial correlation across crashes [63–65]. These three advanced crash prediction functions are employed to predict crash count, crash rate, and crash frequency, respectively. However, no spatial function evolved from the random parameters model, especially the RP-logit model, for predicting crash severity was found. This is a methodology gap and corresponding studies should be conducted to supplement existing severity prediction functions.

2.3. The Current Research

As discussed above, unobserved heterogeneity and spatial correlation may have a substantial contribution to crash severity. However, these two crucial issues are not clear in the analysis of LVR-SV crash severity. Hence, this research was conducted to comprehensively investigate unobserved heterogeneity and spatial correlation across LVR-SV crashes. A two-step modeling approach combining the LCC model and a novel spatial random parameters logit (SRP-logit) model was proposed to complement the severity prediction function. Specifically, the LCC model was first developed to divide the whole-dataset into several relative homogeneous sub-datasets. Then, the SRP-logit model was established for each sub-dataset to capture the spatial correlation and unobserved heterogeneity. Based on the two-step modeling approach, the following questions will be clarified: (1) Does the SRP-logit model outperform the RP-logit model? (2) Which variables have significant impacts on the severity of LVR-SV crashes? (3) Are there differences in the spatial correlation and the unobserved heterogeneity between the sub-dataset and the whole-dataset? A comprehensive research framework of this paper is exhibited in Figure 1.



Figure 1. Comprehensive research framework of this paper.

3. Data

A rural SV crash dataset of 110 two-lane undivided road segments for a five-year period (2014–2019) was extracted from the Shandong Department of Transportation. The GIS map of the region under study is shown in Figure 2. The crash dataset records the

characteristics of the driver, vehicle, time and location, road conditions, weather, visibility, etc. As visibility and weather are estimated by traffic police, some errors may exist; thus a real-time meteorological dataset was collected from the Shandong Climate Center to calibrate row data.



Figure 2. GIS map of the region under study.

Crash data related to low visibility—less than 100 m—was extracted because the rate of fatal crashes in rural areas greatly increases when visibility is less than 100 m [15]. After excluding abnormal data, 19,187 LVR-SV crashes were sorted. Of note is that rain and snow may also cause low visibility; crashes impacted under these conditions are mainly due to unsatisfactory anti-skid performance and were excluded.

More details of road characteristics were extracted from the Traffic Information System maintained by the Shandong Department of Transportation. All road segments at crash locations were verified as approximately linear horizontally and vertically. By checking the curvature and slope of each site, crash data at sharp turns or ramps were eliminated. In total, 19,014 crashes were used to calibrate the model parameters.

During the modeling process, driver injury severity was regarded as a dependent variable. In China, driver injury severity is divided into four categories—no injury (70.0%), slight injury (20.1%), serious injury (8.6%), and fatality (1.2%). A driver passing away within seven days is regarded as a fatal event. It can be seen that fatality cases are very limited, which may lead to incorrect inference. Considering that the two adjacent injury categories are similar, the combination of fatality and serious injury was named FS injuries. It was not expected to have a substantial influence on model estimation. Hence, the dependent variable contains three categories—no injury, slight injury, and FS injuries. No injury was taken as a reference variable.

Independent variables were processed to optimize the model structure; specifically, continuous variables such as time and driver age were discretized, and the classification of discrete variables was simplified. For example, there were more than 20 crash types in the raw data that were simplified into four categories—collision with non-fixed object, collision with fixed object, collision with pedestrian, and other crashes. Motorcyclists must wear helmets, and passenger car drivers must use seatbelts; because they have similar functions, these data were collectively classified as seatbelts/helmets. Nearly 80 types of motor vehicles were recorded, and the vehicle types, less than 1%, were combined into "other" vehicle type. Five categories—passenger car, motorcycle, pickup, truck, and other vehicle type—were included in vehicle type. Traffic control includes several modes, such as police direction, roadway guide markings, and traffic lights. These control methods were marked as traffic control, otherwise marked as no traffic control. The specific variables classification and proportion statistics are shown in Table 1.

			Driver Injury Severity			
Variables	Description	Number	No Injury	Slight Injury	FS Injuries	
Total number		19,014	70.0%	20.1%	9.8%	
Driver gender	Male	17,209	70.6%	19.3%	10.1%	
0	Female *	1804	64.7%	26.7%	8.5%	
Driver age	<25	2473	67.6%	23.8%	8.4%	
0	[25,50] *	13,210	76.1%	16.0%	7.9%	
	>50	3330	47.9%	32.8%	19.2%	
Seatbelt/helmet	Not used	5778	50.9%	32.4%	16.5%	
	Used *	13,235	78.4%	14.5%	6.9%	
Drunk driving	No *	14,200	72.9%	17.9%	9.2%	
0	Yes	4813	58.7%	26.1%	15.1%	
Career	Company staff *	3080	70.0%	20.5%	9.5%	
	Self-employed	5869	77.0%	15.7%	7.2%	
	Farmer	8305	66.6%	22.1%	11.3%	
	Others	1759	63.0%	23.3%	13.6%	
Vehicle type	Passenger car *	8405	90.3%	6.9%	2.7%	
	Motorcycle	4574	29.0%	47.0%	23.8%	
	Pickup	2356	50.3%	34.7%	14.8%	
	Truck	3011	87.6%	6.5%	5.8%	
	Others	667	86.9%	7.7%	5.2%	
Week	Monday/Friday	5405	70.7%	19.7%	9.6%	
	Tuesday–Thursday *	8361	70.0%	20.2%	9.8%	
	Weekend	5247	69.3%	19.9%	10.8%	
Intersection	No *	12,476	69.5%	19.9%	10.6%	
	Yes	6537	71.0%	20.2%	8.8%	
Time of day	00:00-7:00	2720	64.8%	19.7%	15.4%	
	7:00-10:00 *	2666	77.3%	16.2%	6.4%	
	10:00-17:00	6391	72.2%	19.4%	8.2%	
	17:00-21:00	3898	73.6%	19.4%	6.9%	
	21:00-24:00	3338	60.3%	25.0%	14.6%	
Collision type	Fixed object	14,249	72.2%	21.0%	6.6%	
	Non-fixed object *	1704	22.1%	30.8%	46.9%	
	Pedestrian	3042	89.4%	8.3%	2.2%	
	Others	18	50.1%	44.7%	5.2%	
Traffic control	No control *	10,306	69.2%	20.6%	10.2%	
	Control	8707	71.0%	19.3%	9.6%	

Table 1. Variables classification and statistics.

Note: * indicates that variable is reference variable in model.

As shown in Table 1, the proportion of FS injuries of older drivers (>50) (19.2%) and the time of 00:00–7:00 (15.4%) were higher than other categories of factors. In addition, the proportion of FS injuries caused by collision with a fixed object is the highest (46.9%), a finding that is inconsistent with the general rule that the number of crashes decreases with the increase of injury degree. Such a phenomenon is related to the topic of this research. Generally, the traffic environment of rural areas is complex, and the driver safety awareness is not satisfactory. In low-visibility conditions, drivers may not be able to avoid obstacles in time, which is more likely to result in a serious crash.

4. Methodology

4.1. Latent Class Clustering Model

The LCC model is a probability-based multivariate clustering approach and consists of two types of variables—exogenous and latent class. The exogenous variable is risk factors included in the observation, such as gender and age, and the latent class variable is a kind of dummy variable obtained by estimating clustering probabilities of the exogenous variable. Homogeneous clustering was conducted using latent class variables to reduce unobserved

heterogeneity across observations. A complete LCC approach includes four-step theory construction, parameter estimation, fitness evaluation, and result analysis.

For a given LVR-SV crash dataset, it is assumed that *R* latent class variables are estimated using *J* categorical exogenous variables. Each of the exogenous variable has K_j possible values, for crash i = 1, ..., I. The number of outcomes may vary across the different exogenous variables, hence indexing by *j*. Let Y_{ijk} represent the value of the *J* exogenous variable. $Y_{ijk} = 1$ denotes the response of crash *i* for the *j*th variable is *k*, and $Y_{ijk} = 0$ otherwise, where j = 1, ..., J and $k = 1, ..., K_j$.

In the LCC model, the class-conditional probability is represented by π_{jrk} , and its meaning is the *k*th outcome related to *j*th variable is exhibited by a crash in class *r*, where r = 1, ..., R. Further, within each latent class, for each exogenous variable, the sum of the

conditional probabilities for all categories is 1, which can be derived as $\sum_{k=1}^{K_j} \pi_{jrk} = 1$.

Hence, the probability of a specific set of J outcomes was produced by the variable j in crash i for latent class r and can be obtained as:

$$f(Y_i; \pi_r) = \prod_{j=1}^{J} \prod_{k=1}^{K_j} \left(\pi_{jrk} \right)^{Y_{ijk}}$$
(1)

Further, the prior probability of latent class membership is denoted by p_r , which represents the unconditional probability that crash *i* will belong to each latent class before considering the variable response outcome Y_{ijk} provided by the exogenous variables. The probability density function of all latent classes is a form of weighted sum, which can be expressed as (see Linzer et al. [66]):

$$P(Y_i|\pi, p) = \sum_{r=1}^{R} \left[p_r \prod_{j=1}^{J} \prod_{k=1}^{K_j} \left(\pi_{jrk} \right)^{Y_{ijk}} \right]$$
(2)

 p_r and π_{jrk} are unknown model parameters and are determined by maximizing the log-likelihood function. \hat{p}_r and $\hat{\pi}_{jrk}$ are used to represent the estimated results of p_r and π_{jrk} . Based on the outcome of the exogenous variables, the Bayesian inference is used to calculate the posterior probability of each crash belonging to each latent class:

$$\hat{P}(r|Y_i) = \frac{\hat{p}_r f(Y_i; \, \hat{\pi}_r)}{\sum\limits_{q=1}^{R} \hat{p}_q f(Y_i; \, \hat{\pi}_q)}$$
(3)

Model fitness was evaluated by *AIC* (Akaike information criterion), *BIC* (Bayesian information criterion), and *CAIC* (consistent Akaike information criterion). By introducing a penalty term of model complexity, such diagnostic methods provide the criteria to measure complexity and goodness of fit. The smaller the value of *AIC*, *BIC*, and *CAIC*, the better the fitting performance of the model. The calculation formulas of *AIC*, *BIC*, and *CAIC* are as follows:

$$AIC = 2p - 2LL(\beta) \tag{4}$$

$$BIC = p\ln(I) - 2LL(\beta)$$
(5)

$$CAIC = [1 + \ln(I)]p - 2LL(\beta)$$
(6)

where *p* denotes the number of model parameters, $LL(\beta)$ denotes the log-likelihood convergence value, and *I* is the number of observations.

4.2. Spatial Random Parameters Logit Model

The RP-logit model can capture unobserved heterogeneity by allowing model parameters to vary across observations. Such function provides a flexible framework for the RP-logit model. In theory, it can fit any observations with random effects. However, the generality of the model is limited by the inability to capture the spatial correlation across observations [59]. Since adjacent segments have similar geometric characteristics and environments, some factors will be shared across adjacent crashes [60,61]. By considering the spatial correlation, the fitting performance of the RP-logit model is expected to be improved.

In order to better illustrate the modeling approach, the derivation starts from the RP-logit model, which is the original approach of the SRP-logit model. For a given LVR-SV crash at road segment m (m = 1, ..., M), the probability of *ith* crash exhibits a severity level k, marked as P_{ik}^m , which can be given as:

$$P_{ik}^m = P(U_{ik}^m \ge U_{ik}^m) \ \forall k \neq k' \tag{7}$$

 $U_{i,k}^m$ is a propensity function to determine the severity *k* of *ith* crash. *i* = 1,..., *I* is the *ith* crash event where *I* denotes the total number of crashes. *k* = 1,..., *K* is the crash severity. *K* denotes the number of categories of crash severity; *K* = 3 in this research. For such a ternary-dependent variable, $P_{i,k}^m$ is determined by a set of covariates representing specific attributes and a set of unknown parameters. The propensity function is assumed to be linear, which can be expressed as:

$$U_{i,k}^m = \beta_{i,k} X_i^m + \varepsilon_{i,k}^m \tag{8}$$

$$\beta_{i,k}X_i^m = \beta_{i,k,0} + \beta_{i,k,1}x_{i,1}^m + \ldots + \beta_{i,k,L}x_{i,L}^m = \beta_{i,k,0} + \sum_{l=1}^L \beta_{i,k,l}x_{i,l}^m$$
(9)

The term $\beta_{i,k} = (\beta_{i,k,0}, \beta_{i,k,1}, \dots, \beta_{i,k,L})$ represents the vector of model coefficients to be estimated. Among them, $\beta_{i,k,l}$ is the coefficient of the *l*th covariate when the severity of *ith* crash is *k*. $\beta_{i,k,0}$ is the model intercept. $X_i^m = (x_{i,1}^m, \dots, x_{i,L}^m)$ represents a set of covariates related to crash severity. $x_{i,l}^m$ is the *l*th covariate related to *ith* crash. $\varepsilon_{i,k}^m$ denotes the random error term, which obeys Gumbel distribution.

The response variable Y_i^m takes one of three values: $Y_i^m = 0$ denotes no injury and was used as the reference variable, while $Y_i^m = 1$ and $Y_i^m = 2$ represent slight injury and FS injuries, respectively. Let $P_{i,k}^m = \Pr(Y_i^m = k)$ (k = 0, 1, 2) denote the probability of $Y_i^m = k$. Hence, we have:

$$Y_i^m \sim categorical(P_{i,0}^m, P_{i,1}^m, P_{i,2}^m)$$
 (10)

Assuming that the probabilities $P_{i,0}^m$, $P_{i,1}^m$, $P_{i,2}^m$ obey the multinomial logistic distribution, the structure of RP-logit model is shown as follows.

$$\log it(P_{i,k}^m) = \log \frac{P_{i,k}^m}{P_{i,0}^m} = \beta_{i,k,0} + \sum_{l=1}^L \beta_{i,k,l} x_{i,l}^m \ (k = 1, 2)$$
(11)

where all variables in Equations (10) and (11) are the same as those defined earlier.

The regression coefficients to be estimated in RP-logit model may vary across observations. It is assumed that the normal distribution, marked as $N(\cdot)$, is followed by the random coefficients $\beta_{i,k} = (\beta_{i,k,0}, \beta_{i,k,1}, \dots, \beta_{i,k,L})^T$ and its distribution form is as follows.

$$\beta_{i,k} \sim N(\mu_k, \, \phi_k) \tag{12}$$

$$\beta_{i,k} = \begin{pmatrix} \beta_{i,k}^{1} \\ \beta_{i,k}^{2} \\ \cdots \\ \beta_{i,k}^{L} \end{pmatrix}, \ \mu_{k} = \begin{pmatrix} \mu_{k}^{1} \\ \mu_{k}^{2} \\ \cdots \\ \mu_{k}^{L} \end{pmatrix}, \ \phi_{k} = \begin{pmatrix} (\varphi_{1,1}^{k})^{2} & (\varphi_{1,2}^{k})^{2} & \cdots & (\varphi_{1,L}^{k})^{2} \\ (\varphi_{2,1}^{k})^{2} & (\varphi_{2,2}^{k})^{2} & \cdots & (\varphi_{1,L}^{k})^{2} \\ \cdots & \cdots & \cdots & \cdots \\ (\varphi_{L,1}^{k})^{2} & (\varphi_{L,2}^{k})^{2} & \cdots & (\varphi_{L,L}^{k})^{2} \end{pmatrix}$$
(13)

It is necessary to point out that the random parameter $\beta_{i,k}^l$ is determined when its variance $\varphi_{w,l}^k$ is significantly greater than zero. Otherwise, it will be regarded as a fixed parameter across observations. If no random parameter is found, the RP-logit model will degenerate to a multinomial logit model.

To capture the spatial correlation of LVR-SV crashes, a spatial structure is constructed by adding a spatial error term δ_k^m to the RP-logit model, called the spatial random parameters logit (SRP-logit) model. The spatial term is interpreted by Gaussian CAR techniques, which have been widely accepted in spatial statistical analysis. Based on this modification, unobserved heterogeneity across observations can be captured by random parameters and the spatial correlation among adjacent crashes can be captured by the spatial structure. The specific model structure is as follows.

$$\log it(P_{i,k}^m) = \log \frac{P_{i,k}^m}{P_{i,0}^m} = \beta_{i,k,0} + \sum_{l=1}^L \beta_{i,k,l} x_{i,l}^m + \delta_k^m \ (k = 1, 2)$$
(14)

An effective joint density specification often used for the spatial term $\delta_k = (\delta_k^1, \delta_k^2, \dots, \delta_k^M)$ is in terms of pairwise differences in errors and a variance term $\sigma_{\delta_k}^2$ [67,68], thus:

$$P(\delta_k) \propto \exp\left[-0.5\omega^{-1} \sum_{m \neq \overline{m}} \left(\delta_k^m - \delta_k^{\overline{m}}\right)^2\right], \ \omega = \sigma_{\delta_k}^2 \tag{15}$$

A normal distribution is implied in this joint density for δ_k^m conditioning on the effect of $\delta_k^{\overline{m}}$ on the remaining observation road segments. The corresponding conditional form is:

$$\delta_k^m \Big| \delta_k^{\overline{m}} \sim N \left(\overline{\delta_k^m}, \frac{\sigma_{\delta_k}^2}{\sum_{\overline{m} \neq m} w_{m, \overline{m}}} \right)$$
(16)

With

$$\overline{\delta_k^m} = \frac{\sum_{\overline{m} \neq m} \left(w_{m, \overline{m}} \delta_k^m \right)}{\sum_{\overline{m} \neq m} w_{m, \overline{m}}}$$
(17)

where $w_{m,\overline{m}}$ denotes the un-normalized weight between road segment m and \overline{m} , and it is defined as a proximity structure based on distance attenuation. The weight is calculated using exponential decay function of the distance between road segments. In addition, let (t_{m1}, t_{m2}) represent the crash coordinate in segment m (the latitude and longitude). The Euclidean distance between road segment m and \overline{m} is represented by $d_{m,\overline{m}} = \sqrt{(t_{m1} - t_{\overline{m1}})^2 + (t_{m2} - t_{\overline{m2}})^2}$. Hence, we have:

$$w_{m,\,\overline{m}} = e^{-\alpha d_{m,\,\overline{m}}} \tag{18}$$

The term α is a weight coefficient, which controls the decline rate of correlation. At the same time, if the potential spatial correlation of all observed road segments is considered, this will reduce the model fit performance [69]. Inspired by Klassen et al. [61] and Aguero-Valverde et al. [70] and combined with the actual length of road segment, $d_{m,\overline{m}} = 3 \, km$ was used as the threshold to identify the spatial correlation. Specifically, the minimum distance between the observed road segments is greater than this threshold, and a weight with a value of 0 is considered, otherwise it is 1. As pointed out by El-Basyouny and Sayed [71], due to the complexity of traffic interaction around the specific crash site, random variations across road segments may be structured spatially, and the SRP-logit model can be calibrated using the PROC GLIMMIX procedure in SAS version 9.3.

In addition to *AIC* and *BIC*, *McFadden* R^2 is often used to evaluate model fit. The calculation approach is as follows:

$$McFadden R^{2} = 1 - \frac{LL(\beta)}{LL(0)}$$
(19)

where $LL(\beta)$ and LL(0) denote the convergence value and the initial value of log-likelihood, respectively. The range of *McFadden* R^2 is 0 to 1—the larger the value, the better the fitting performance.

4.3. Average Marginal Effect

The average marginal effect of the significant risk variables was calculated to quantify the impact of the independent variables on crash severity. Since the independent variables in this research are binary variables, the calculation paradigm of the average marginal effect, marked as $E_{X_i}^{P_k}$ is as follows:

$$E_{X_{l}}^{P_{k}} = \frac{1}{I} \sum_{i=1}^{I} \left(P_{i,k}(X_{i,l} = 1) - P_{i,k}(X_{i,l} = 0) \right)$$
(20)

where $X_{i,l}$ denotes the value of the *l*th independent variable in *i*th crash. $P_{i,k}(X_{i,l} = 1)$ represents the probability that the severity of crash *i* is *k* under the premise of $X_{i,l} = 1$. For a given risk variable, all the observations have a marginal effect, and the average marginal effect was obtained by calculating the average of all marginal effects. The mathematical meaning of the average marginal effect refers to the probability change of a certain crash severity when a variable changes by one unit while the value of other variables remains stable.

5. Analysis and Discussion

5.1. Analysis of Latent Class Clustering Model

The LCC model was performed based on the exogenous variable shown in Table 1 in Latent GOLD 4.5 software. An exploratory analysis approach was used to determine the optimal number of latent class variables. Starting with the latent class variable of 1, the number of latent class variables was gradually increased and model fitness was checked one by one to select the best model. In total, 12 clustering models were constructed; the specific fitting comparison was shown in Figure 3. As shown in Figure 3, the values of *AIC*, *BIC*, and *CAIC* decrease with the increase of the number of latent class variables. However, the percentage reduction in the values of *AIC*, *BIC*, and *CAIC* is less than 1% starting from a latent class variable of 4. This indicates that dividing the observation into four clusters (sub-datasets) can satisfactorily identify the latent class variable. The distribution statistics of different variables in the different clusters are shown in Appendix A Table A1.



Figure 3. Identification of the number of latent class variables.

Since the proportion of variables in different clusters is different, variables with a high percentage can be identified as characteristic variables; hence, the latent class variables can be described by the characteristic variables. The characteristic variables (nine groups in total) and simple size of different clusters are shown in Table 2.

Variables	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Number of observations	9517	3561	3075	2860
Older driver (>50)	10.8%	8.5%	19.3%	49.7%
Seatbelt/helmet used	83.1%	77.8%	71.4%	13.5%
Seatbelt/helmet not used	16.9%	22.3%	28.6%	86.5%
Drunk driving	23.7%	15.2%	24.5%	45.5%
Passenger car	81.2%	12.7%	17.5%	1.2%
Motorcycle	9.8%	1.8%	68.4%	52.9%
Truck	0.2%	68.2%	7.4%	0.03%
Collision with fixed object	2.5%	3.6%	40.2%	3.3%
Mid-age driver (25–50)	73.1%	88.3%	64.3%	33.6%

Table 2. Distribution of characteristic variables and simple size of each cluster.

As can be seen from Table 2, the percentages of seatbelt/helmet used and passenger car in cluster 1 are 83.1% and 81.2%, respectively, which are significantly higher than those in other clusters. Hence, these two variables are the characteristic variables in cluster 1. In cluster 2, the percentage of truck is 68.2%, which is less than 10% in other clusters. Further, among the different vehicle types in cluster 2, truck has the highest proportion. Hence, truck is one of the characteristic variables of cluster 2. In addition, the proportion of midage driver (25–50) in cluster 2 is significantly higher than that in other clusters. Thus, truck and mid-age driver are the characteristic variables of cluster 2. In cluster 3, the proportion of motorcycles (68.4%) and collision with fixed object (40.2%) is significantly higher than that in other clusters. Therefore, cluster 3 has two characteristic variables are contained in cluster 4, which are older driver, seatbelt/helmet not used, and drunk driving.

In addition, the proportional distribution of crash severity for different clusters is shown in Table 3. Interestingly, the crash severity in cluster 1 contains only two categories, and they are no injury and slight injury. Other clusters include three categories—no injury, slight injury, and FS injuries. This phenomenon can be confirmed by the characteristic variables of cluster 1. As is known, the use of a seatbelt/helmet can provide superior protection for drivers, which will significantly reduce the probability of a serious crash [15]. Moreover, the percentage of FS injuries in cluster 3 is 41.4%, which is significantly higher than that in other clusters. The characteristic variables of cluster 3 is motorcycle and collision with fixed object; both of these variables are risk factors related to high mortality [52]. These findings preliminarily demonstrate the superiority of the LCC model in identifying latent class variables.

Table 3. Distribution of crash severity in each cluster.

Categories	Total Sample Size	No Injury	Slight Injury	FS Injuries
Whole-dataset	19,013	13,309 (70.0%)	3802 (20.1%)	1902 (9.8%)
Cluster 1	9517	9264 (97.3%)	253 (2.6%)	0 (0%)
Cluster 2	3561	3377 (94.8%)	105 (2.9%)	79 (2.2%)
Cluster 3	3075	46 (1.5%)	1757 (57.1%)	1272 (41.4%)
Cluster 4	2860	675 (23.6%)	1655 (57.8%)	530 (18.5%)

5.2. Analysis of Spatial Random Parameters Logit Model

Establishment of the SRP-logit model was a progressive process. The initial variables were added to the model one by one and were retained or deleted depending on the

significance of the parameters (90% confidence level). The modeling was repeated five or more times for each dataset, and the best-fitting model was employed.

Similar to the RP-logit model, the distribution form of random parameters needs to be pre-set. To verify the optimal distribution hypothesis of random parameters, three commonly used probability density distributions—normal distribution, uniform distribution, and log-normal distribution—were tested. Among them, the log-normal distribution provides a limitation, that is, the model parameters need to maintain the same sign (all positive or all negative). However, signs of the parameters varying across the risk factors are often observed [39,55]. Hence, the log-normal distribution is not appropriate in this research. The fitting performance of random parameters obeying normal and uniform distributions was compared. The results are shown in Table 4.

Indicators	Distribution Form	Whole-Dataset	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Number	-	19,013	9517	3561	3075	2860
LL(β)	Normal distribution	-13,762.75	-5661.29	-3285.78	-2632.22	-2168.51
	Uniform distribution	-13,768.13	-5674.34	-3283.78	-2641.53	-2179.42
BIC	Normal distribution	20,490.35	9084.12	4291.55	3971.34	3067.85
	Uniform distribution	20,506.58	9095.87	4291.55	3988.25	3071.32
McFadden R2	Normal distribution	0.342	0.401	0.397	0.371	0.408
	Uniform distribution	0.339	0.400	0.397	0.364	0.406
<i>p</i> -value	Normal distribution		<	0.001		
	Uniform distribution		<	0.001		

 Table 4. Fitting comparisons of random parameters following different distributions.

The optimal distribution hypothesis was determined by comparing the *BIC* and *McFadden* R^2 values. As shown in Table 5, for the whole-dataset, cluster 1, cluster 3, and cluster 4, the random parameters in the SRP-logit model conforming to the normal distribution outperform those conforming to the uniform distribution because the normal distribution in such dataset has a lower *BCI* value and a higher *McFadden* R^2 value. However, no random parameter was captured in cluster 2. Hence, the SRP-logit model established by cluster 2 degenerates to the spatial multinomial logit model.

Table 5. Comparison of fitting performance between RP-logit model and SRP-logit model.

Creach Datasat		BIC	
Crash Dataset	RP-Logit Model	SRP-Logit Model	Difference
Whole-dataset	20,563.21	20,490.35	72.86
Cluster 1	9109.08	9084.12	24.96
Cluster 2	4309.86	4291.55	18.31
Cluster 3	3983.92	3971.34	12.58
Cluster 4	3072.02	3067.85	16.17

According to the optimal distribution hypothesis of random parameters in the SRP-logit model, the $LL(\beta)$ value of the model built by the whole-dataset was -13,762.75, which was less than the sum of the $LL(\beta)$ values of the model built by the different sub-dataset (-13,747.80). The *BIC* value of the whole-dataset model was greater than the sum of the *BIC* values of the sub-dataset model (the difference is about 75.49). Meanwhile, the chi-square test of both the whole-dataset model and the sub-dataset models was significant (p < 0.001). It was shown that the fitting performance of the SRP-logit model can be improved by using the relatively homogeneous sub-dataset to calibrate the parameters.

In addition, the RP-logit model was established for each dataset, and the distribution of random parameters was consistent with the SRP-logit model. Then, the fitting performance was compared between the RP-logit model and the SRP-logit model by the *BIC* criteria. As shown in Table 5, the *BIC* values of the PR-logit model and the SRP-logit model established by cluster 1 were 9109.08 and 9084.12, respectively, which shows a difference of 24.96. Meanwhile, the RP-logit model and the SRP-logit model calibrated by cluster 2 showed a *BCI* difference of 18.31. Similarly, the SRP-logit models calibrated by cluster 3 and cluster 4

were accompanied by more satisfactory *BIC* values, which were 12.58 and 16.17 lower than the RP-logit model, respectively. The same findings can be confirmed for the model built by the whole-dataset.

These statistics showed that the SRP-logit model outperforms the RP-logit model in terms of fitting crash severity, and this finding remained stable across different datasets. Hence, spatial correlation is a recommended modeling approach. In addition, this study also proved that the Gaussian CAR technology is an effective method to identify spatial correlation in modeling crash severity, which further enriches the research on traffic safety. Guo et al. [72] and Klassen et al. [61] reached a similar conclusion and pointed out that a highly significant spatial error component can reveal more crash information compared to a model without spatial structure.

The estimation results of the variance term $\sigma_{\delta_k}^2$ in the SRP-logit model are shown in Table 6, and the independent variable estimation and average marginal effect analysis are shown in Tables 7 and 8, respectively. As shown in Table 6, the spatial correlation was captured in all crash datasets because the spatial variance term was highly significant. Of note is that the significance of variance term varies across the different crash severities and the different datasets. Specifically, in the whole-dataset, a significant spatial correlation was identified for the three severity levels-no injury, slight injury, and FS injuries-under the distance threshold of 3 km. In cluster 1, the spatial parameters of no injury and slight injury were significant. In cluster 2, there is a significant spatial correlation among no-injury crashes. The spatial correlation of slight injury and FS injuries was captured in cluster 3 and cluster 4. These findings indicate that there is a spatial correlation between the severities of LVR-SV crashes. As can be seen in Tables 7 and 8, there are significant difference in the parameter estimation and the average marginal effect between the whole-dataset model and the sub-dataset model. For example, some variables are not significant in the wholedataset model but are significant in the sub-dataset model. By comparing these models, it is possible to reveal some important variable information hidden in the whole-dataset model [67].

Severity	Whole-Dataset	Cluster 1	Cluster 2	Cluster 3	Cluster 4
No injury	1.067 *	0.731 *	2.549 **		
Slight injury	0.814 **	0.415 ***		0.612 *	1.013 *
FS injuries	1.261 *			2.164 **	0.321 ***

Table 6. Coefficient estimation of spatial error term in SRP-logit model.

Note: * indicates 90% confidence level; ** indicates 95% confidence level; *** indicates a confidence level above 95%.

5.3. Discussion

Based on the regression results in Tables 7 and 8, a total of 18 factors that have significant impact on crash severity were identified. Among them, three factors (seatbelt not used, motorcycle, and collision with fixed object) lead to a significant increase in the probability of serious crashes, and these findings are consistent across different datasets. It is suggested that these three risk factors should be considered emphatically by the policy makers to reduce the severity of rural crashes. In most cases, the other three factors—drunk driving, old drivers, and the time of 21:00–7:00 (21:00–24:00 and 0:00–7:00)—also have a positive correlation with crash severity. In addition, truck and collision with pedestrian have a negative correlation with crash severity. These findings are consistent with the official statistics [3] and previous studies [4,6,10,54]. A detailed discussion of the risk factors is shown below.

Variables	Whole-Dataset		Cluster 1	Cluster 2		Cluster 3		Cluster 4	
vallables	S	F	S	S	F	S	F	S	F
Male driver	-0.63 **	-0.28 ***	-0.29 *				-1.51 ***	0.31 *	1.40 **
Std. dev.		0.50							0.42
Young (<25)		-0.39 **		-0.59 **	-0.60 *			-0.56 **	0.82 **
Std. dev.		1.06							
Old (>50)	0.62 **	1.06 ***	0.81 **	-1.58 *	-1.34 **		1.58 *		0.88 ***
Seatbelt not used	0.71 *	1.24 **	0.99 **		0.68 **	1.05 **	1.79 **	-1.17 **	
Drunk driving	0.33 **	0.31 **	0.35 *		-2.64 ***	1.04 *	2.68 *	-0.22 *	0.35 **
Std. dev.	1.25		1.21						
Self-employed	-0.41 **	-0.52 **	-0.13 *						
Farmer	-0.17 *							-0.32 *	-0.43 *
Motorcycle	3.07 **	3.32 ***	0.41 *			0.70 *	3.42 **		2.45 *
Pickup	-1.97 **	-1.12 **	-2.21 **	-0.59 *	-1.22 **				2.407 *
Std. dev.		2.24						0.52	
Truck		-0.73 **	-0.43 *	-0.46 *	-1.30 *				
Others	0.49 *								
10:00-17:00	0.22 *								
Std. dev.	0.53								
17:00-21:00	0.27 *							0.38 *	
Std. dev.	1.28							2.45	
21:00-24:00	0.93 **	1.28 **	-0.89 **	1.29 ***			0.34 **	0.35 *	0.60 **
0:00-7:00	0.68 **	1.26 ***	-1.29 **	1.08 **	2.63 ***		0.26 **		
Control	0.20 *							0.38 **	
Fixed object	1.89 **	3.64 **		1.71 **	1.64 ***	1.70 *	2.52 *		
Std. dev.					1.51				
Pedestrian	-1.77 **	-2.08 *	-0.56 **	-1.41 **	-0.60 **				
Others		3.29 **			3.34 **				
Intercept	-2.56 ***	-4.91 *	-3.03 **	-3.24 **	-5.22 **			2.41 **	

 Table 7. Coefficient estimation results for SRP-logit model.

Note: S = slight injury; F = FS injuries. * indicates 90% confidence level; ** indicates 95% confidence level; *** indicates a confidence level above 95%.

 Table 8. Average marginal effect of SRP-logit model.

Variables	Whole-	Dataset	Cluster 1	Clus	ster 2	Clu	ster 3	Clus	ter 4
variables	S	F	S	S	F	S	F	S	F
Male	-0.065	0.005	-0.061				-0.114	0.034	0.123
Young (<25)		-0.022		-0.003	-0.031			-0.076	0.038
Old (>50)	0.035	0.053		-0.007	-0.021		0.218		0.147
Seatbelt not used	0.038	0.055	0.041		0.001	0.132	0.306	0.254	
Drunk driving	0.028	0.007	0.010		-0.043	0.039	0.109	-0.015	0.028
Self-employed	-0.023	-0.017	-0.006						
Farmer	-0.016							-0.030	-0.029
Motorcycle	0.174	0.128	0.030			0.187	0.238		0.126
Pickup	0.211	0.074	-0.016	-0.049	-0.076				0.078
Truck		-0.030	-0.019	-0.012	-0.059				
Others		0.019							
10:00-17:00	0.018								
17:00-21:00	0.026							0.022	
21:00-24:00	0.064	0.052	-0.002	0.018			0.105	0.019	0.051
0:00-7:00	0.031	0.062	-0.003	0.016	0.003		0.097		
Control		0.011							0.035
Fixed object	0.045	0.332		0.014	0.012	0.073	0.132		
Pedestrian	-0.134	-0.049	-0.027	-0.031	-0.001				
Others		0.411			0.014				

5.3.1. Driver Characteristics

A significant correlation was found between male driver and the crash severity. The parameter of this variable (FS injuries) follows a normal distribution, with a mean of -0.28 and a standard deviation of 0.84 in the whole-dataset model, showing that 37.0% of male drivers are more likely to cause FS injuries. Further, the regression coefficient of male driver varies across different sub-datasets. For FS injuries, the parameters of male driver in cluster 3 and cluster 4 were negative (fixed parameter) and positive (random parameter), respectively. This variable in clusters 1 and 2 was not significant. It is shown that the unobserved heterogeneity related to male driver (FS injuries) presenting in the whole-dataset was divided into cluster 3 and cluster 4 by the LCC model. Among these, the heterogeneity of male driver in cluster 4. Hence, a conclusion can be drawn (consistent with Liu and Fan [55]), under the specific risk factor, that the LCC model can effectively eliminate the heterogeneity in some of the sub-datasets, but this function cannot be supported by all the sub-datasets.

Young driver (under age 25) was significantly related to crash severity, which is consistent with Feng et al. [73]. However, the coefficient of such variable varies across different datasets. In the whole-dataset model, the parameter of young driver (FS injuries) follows a normal distribution with a mean of -0.39 and a standard deviation of 1.06. However, no random effects were found in sub-dataset models. The coefficient of young driver (FS injuries) in cluster 2 and cluster 4 was negative and positive, respectively; this variable was not significant in cluster 1 and cluster 2. This finding shows that the unobserved heterogeneity of young driver in the whole-dataset can be effectively captured by the LCC model and divided into cluster 2 and cluster 4. Further, these coefficients in the sub-dataset can be explained by the characteristic variables. In cluster 2, one characteristic variable was truck, which can provide satisfactory protection measures for a driver [48]. In cluster 4, both seatbelt/helmet not used and drunk driving were the characteristic variables, and there is a significant positive correlation between these two variables and serious crashes [52]. These findings reflect the infinite potential of the LCC model to mine data heterogeneity.

Older driver (above age 50) shows a significant influence in most datasets. In the whole-dataset, cluster 3, and cluster 4, the likelihood of FS injuries caused by older drivers increased by 5.3%, 21.8%, and 14.7%, respectively. Wu et al. [9] reached a similar conclusion. Therefore, older drivers should drive carefully and maintain safe driving habits to reduce crash severity. However, the coefficient of this variable in cluster 2 suggested that older driver was expected to significantly reduce the probability of FS injuries (the average marginal effect was -0.021). A characteristic variable in cluster 2 was truck. Under low-visibility driving conditions, more driving experience may be a dominant factor in determining crash severity. This conclusion reflects the advantages of the LCC model to mine the potential variable information; it can maximize the heterogeneity across different sub-datasets and reveal the important variable information hidden in the whole-dataset.

The influence of seatbelt/helmet on crash severity cannot be ignored. The probability of FS injuries caused by seatbelt/helmet not used increases by 5.5%, 0.1%, and 30.6% in the whole-dataset, cluster 1, and cluster 3, respectively, which is consistent with previous studies [54]. The impact of seatbelt/helmet not used on cluster 3 is significantly higher than that in other datasets. Cluster 3 has two characteristic variables—motorcycle and collision with fixed object. Rollovers often occur in motorcycle crashes, which can cause serious injury, especially head injury, to drivers; the available literature confirms that a significant cause of motorcyclists being killed is head injury [52]. Not wearing a helmet leaves a rider's head unprotected, which leads to a significant increase in the probability of a serious injury. However, 30.4% of drivers still do not use seatbelts/helmets in the whole-dataset. It is necessary to take some mandatory management measures to strengthen the safety awareness of drivers.

In the whole-dataset, cluster 3, and cluster 4, drunk driving causes a significant increase in the probability of serious crashes (marginal effects 0.007, 0.109, and 0.028, respectively) [55]. These results demonstrate the pernicious effects of drunk driving on traffic safety to policy makers and the general public. Driving behavior can be affected by alcohol, which can lead to serious collisions. In cluster 2, however, there was a significant negative correlation between drunk driving and severe crashes (marginal effect-0.043). This may be due to the high proportion of trucks in cluster 2 (68.2%), and some risk compensation operations will be carried out by experienced truck drivers to reduce the influence of alcohol [74].

5.3.2. Vehicle Characteristics

According to the model calibration results, motorcycles, pickups, and trucks have significant impacts on the severity of LVR-SV crashes. As there are some differences between various vehicle types, which cause crash severity to vary across different vehicle types, this finding was proved by Rezapour et al. [26]. Specifically, there was a significant positive correlation between motorcycle and serious crashes. In the whole-dataset, cluster 3, and cluster 4, the probability of FS injuries caused by a motorcycle increased by 12.8%, 23.8%, and 12.6%, respectively. Wei and Cai [15] obtained a similar finding and pointed out that motorcycles cannot provide superior protection for riders, and high mortality is inevitable. Meanwhile, judgment related to the driving environment is influenced by low-visibility conditions, which increase the probability of serious crashes [75].

In the whole-dataset, the parameter of pickup (FS injuries) obeys a normal distribution, with a mean of -1.12 (0.11) and a standard deviation of 2.24 (0.13). The heterogeneity exhibited in the whole-dataset was identified and divided into different sub-datasets by the LCC model. In cluster 2 and cluster 4, the coefficient of pickup (FS injuries) was fixed (-1.22) and random (mean 1.42, standard deviation 1.35), respectively; it was not significant in clusters 1 and 3. This result shows that the LCC model can effectively eliminate the heterogeneity of pickup in cluster 2, but this finding cannot be supported by cluster 4. The residual heterogeneities need to be identified by the SRP-logit model. This finding indicates that the LCC model and the SRP-logit model can compensate for the shortcomings of each other, and combining the two models can provide a superior modeling approach.

There is a significant negative correlation between truck and the severity of LVR-SV crashes, which is consistent with previous research [76]. In the whole-dataset and cluster 2, the probability of FS injuries caused by truck was reduced by 3.0% and 5.9%, respectively, compared with passenger car. In addition, truck (slight injury) was not significant in the whole-dataset model, but a significant influence was shown in cluster 1 and cluster 2. This finding suggests that the significance of truck (slight injury) was concealed by the whole-dataset; this potential impact can be demonstrated by using the sub-dataset to calibrate the SRP-logit model.

5.3.3. Other Characteristics

In the whole-dataset and cluster 4, the probability of slight injury related to the time of 17:00–21:00 increased by 2.6% and 2.2%, respectively, compared with 7:00–10:00. Chang et al. [52] reached a similar conclusion and pointed out that this phenomenon was caused by aggressive driving caused by traffic chaos during the rush hour [77] because the sample in this research was rural crashes and there is no evening peak phenomenon in rural China. Therefore, a more suitable explanation is needed for the findings of this research. Under low-visibility conditions, it is difficult to maintain a satisfactory driving environment, and the infrastructure of rural roads is lacking. Furthermore, drivers are fatigued after a day's work and may not be able to avoid a crash under such conditions. This phenomenon is particularly significant for drivers who are not familiar with the local traffic environment [14].

In addition, there is a significant positive correlation between FS injuries and the time of 21:00–7:00 (21:00–24:00 and 0:00–7:00), which is consistent with previous research [15].

In the whole-dataset, cluster 3, and cluster 4, the probability of FS injuries related to the 21:00–24:00 increased by 5.2%, 10.5%, and 5.1%, respectively; this variable corresponding to FS injuries was not significant in cluster 1 and cluster 2. Further, 0:00–7:00 (FS injuries) was significant in the whole-dataset, cluster 2, and cluster 3 (average marginal effect 0.062, 0.003, and 0.097, respectively). These findings suggested that serious crashes are more likely to occur at night and satisfactory driving conditions should be maintained to reduce crash severity.

In the whole-dataset, cluster 2, and cluster 3, the probability of FS injuries caused by collision with fixed object increased by 33.2%, 1.2%, and 13.2%, respectively. Zhu et al. [78] reached a similar conclusion and pointed out that this finding may be related to the high probability of rollover caused by collision with a fixed object. In addition, the parameter of collision with fixed object corresponding to FS injuries was fixed across observations in the whole-dataset, but it was random across observations in cluster 2 (mean 1.64, standard deviation 1.51). This finding suggests that the model calibrated using the sub-dataset, as obtained by the LCC model, can reveal the variable information hidden by the whole-dataset.

There was a significant negative correlation between collision with pedestrian and crash severity. In the whole-dataset and cluster 2, the probability of serious injury caused by collision with pedestrians was reduced by 4.9% and 0.1%, respectively. Pedestrians are vulnerable compared with motor vehicles, and a collision with a pedestrian is not expected to cause serious injury to motor vehicle occupants but could seriously threaten the lives of pedestrians.

5.4. Recommendations

The method demonstrated in this study can be used to identify risk factors associated with crash severity under low-visibility conditions. For areas where low-visibility conditions occur frequently, such as Chongqing and Zibo, China, the proposed method can be used to extract significant influencing factors. Then, effective solutions can be proposed according to the factor characteristics to improve traffic safety performance and promote sustainable development.

Based on the findings in this article and previous engineering experience, the greatest benefit received by policy makers is the clarification of risk factors of the LVR-SV crash severity. Some measures can be adopted to improve traffic safety.

(1) Certain risky driving behaviors (such as seatbelt not used, drunk driving, overspeed, and fatigue driving) have significant influence on crash severity. Legal measures are effective means to overcome risky driving behavior. Several electronic policies can be installed on the roadside in rural areas to monitor risky driving behavior and impose penalties. Meanwhile, an in-vehicle crash warning system that identifies such risky behaviors in real time may contribute to improving traffic safety performance. The primary purpose of these measures is to remind drivers to maintain safe driving habits.

In fact, these dangerous behaviors are closely related to the lack of adequate safety awareness among drivers. Improving the safety awareness of drivers in rural areas is fundamental and important for improving road safety. Appropriate publicity and education are necessary. In addition, an on-board navigation system can report the low-visibility conditions ahead and advise drivers to slow down appropriately.

(2) A conclusion was drawn by this study that collision with fixed object will lead to a significant increase in the risk of fatal collisions. Hence, the traffic management department should focus on improving hard guardrails or other pavement fixtures on both sides of the road (such as concrete wall pillars), replacing hard guardrails with soft guardrails that can absorb collision energy, or installing a buffer energy-absorbing device on the surface of a fixed object to reduce crash severity.

(3) The impact of motorcycles on traffic safety must be seriously improved. The supervision, management, and education of motorcycle riders should be strengthened. Dangerous driving and abnormal driving license status should be given severe punishment

to reduce high-risk travel behavior. The government should encourage residents to choose alternative modes of transportation instead of motorcycles.

(4) In rural China, road infrastructure is commonly inadequate. The Department of Transportation should install sufficient roadway ancillary facilities, including road condition signs, speed limit signs, pavement marking, and streetlights. Among them, installation of streetlights is most essential because nights without street lighting will lead to a dramatic increase in the risk of fatal collisions [48]. Meanwhile, to ameliorate the effects of low-visibility conditions on crashes, we recommend installing blinking lights on the roadside to alert drivers.

Among the safety measures mentioned above, the countermeasures corresponding to seatbelt not used, motorcycle, and collision with fixed object should be implemented first because the positive correlation between these variables and crash severity remains stable across different datasets. Then, other measures could be implemented one after another.

It is necessary to point out that we have received strong support from the Zibo Department of Transportation and several corresponding measures have been implemented in some of the studied road segments. In the future, traffic crash data before and after the implementation of safety measures will be collected and compared to evaluate the effectiveness of the measures. This part is the focus of our future research.

6. Conclusions

The primary objective of this research was to investigate the risk factors of LVR-SV crash severity. The injury severity of drivers was used as the dependent variable—no injury, slight injury, and FS injuries. A two-step modeling approach was proposed. First, the LCC model was used to divide the whole-dataset into several sub-datasets. By comparing the fitting indicators of different numbers of sub-datasets, the whole-dataset was divided into four sub-datasets. Then, the SRP-logit model was established for the whole-dataset and the sub-datasets to identify the risk factors of crash severity and the spatial correlation among adjacent crashes. In total, 18 significant factors were identified—male driver, young age (<25, >50), seatbelt/helmet not used, drunk driving, career (self-employed and farmer), vehicle type (motorcycle, pickup, truck), time of day (10:00–17:00, 17:00–21:00, 21:00–24:00, 0:00–7:00), collision type (collision with fixed object, collision with pedestrian), and traffic control.

The statistical comparison showed that the SRP-logit model outperforms the RP-logit model and is a superior modeling technique. The significance of the spatial error term in the SRP-logit model indicates that there is a spatial correlation between the severity of LVR-SV crashes. Meanwhile, the significance of the spatial error term varies across different datasets. In the whole-dataset, the spatial correlation was captured among all the three severity levels; however, the spatial correlation in the sub-dataset was less than the whole-dataset. For example, the spatial parameter of no injury in cluster 2 was significant, but slight injury and the FS injuries were not significant. This finding suggests that the LCC model can reduce the spatial correlation in the sub-dataset, but it cannot be completely eliminated.

In addition, dividing the whole-dataset into sub-datasets through the LCC approach and calibrating the model by the sub-dataset can further improve the fitting performance of the SRP-logit model. Hence, a two-step analysis approach is advocated to model crash severity.

Some interesting findings were obtained. First, the LCC model can effectively eliminate the unobserved heterogeneity across observations in some sub-datasets, but it cannot eliminate them in all sub-datasets. For example, male drivers showed random effects in the whole-dataset and cluster 4, but no random effects were found in cluster 3. This finding suggests that the LCC model can effectively eliminate the heterogeneity of male drivers in cluster 3, but there was residual heterogeneity in cluster 4. Second, the sub-dataset model can reveal some important variable information that was hidden by the whole-dataset model. Truck (slight injury) was not significant in the whole-dataset model, but a significant influence was captured in cluster 1 and cluster 2.

7. Limitations of the Study

This research has some limitations.

(1) The risk factors considered in this study are not rich enough. If variables such as traffic volume and vehicle speed can be considered, it is expected to further improve the fitting performance of the SRP-logit model, although it is difficult to obtain such information. Therefore, in the future, a millimeter-wave radar could be installed on several rural road segments to collect valuable data.

(2) The crash extracted in this research occurred in China; the crash characteristics may vary with different cultural backgrounds and traffic conditions. Thus, the transferability of the model proposed in this research across countries should be studied to further enrich the existing literature.

(3) This study used Gaussian CAR technology to account for the spatial correlation error of crashes. We will continue to study the decrease of the spatial correlation error by using other alternative methods.

Author Contributions: Study conception and design, Z.C., F.W., Y.G. and Z.W.; data collection, F.W., X.L. and L.C.; data analysis and interpretation of results, Z.W. and Z.C.; draft manuscript preparation, F.W., Y.G. and Z.C.; manuscript revision, Y.G., F.W. and Z.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was jointly supported by (1) the National Natural Science Foundation of China (Grant No. 71901134 and 71871057); (2) the National Science Foundation for Distinguished Young Scholars (Grant No. 51925801); (3) the Natural Science Foundation of Shandong (Grant No. ZR2018BF024); and (4) the Postdoctoral Research Assistance Program of Jiangsu (Grant No. 2018K118C).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are thankful for the support of the Shandong Department of Transportation (SDOT) and the Zibo Department of Transportation (ZDOT).

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

Table A1. Proportional distribution of independent variables in LCC model.

Variables	Description	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Number		9517	3561	3075	2860
Injury severity	No injury	9210	3377	46	675
	Slight injury	253	105	1757	1655
	FS injuries	54	79	1272	530
Driver gender	Female	12.3%	0.1%	2.6%	21.7%
0	Male	87.7%	99.9%	97.4%	78.3%
Driver age	<25	73.1%	88.3%	64.4%	36.6%
Ũ	[25,50]	16.1%	3.2%	16.3%	13.7%
	>50	10.8%	8.5%	19.3%	49.7%
Seatbelt/helmet	Used	83.1%	77.8%	71.4%	13.5%
	Not used	16.9%	22.2%	28.6%	86.5%
Drunk driving	No	76.2%	84.8%	75.5%	54.5%
Ū.	Yes	23.8%	15.2%	24.5%	45.5%
Career	Company staff	16.6%	16.2%	14.8%	16.5%
	Self-employed	38.4%	26.3%	25.5%	19.5%
	Farmer	36.9%	49.4%	50.9%	49.4%
	Others	8.2%	8.1%	8.9%	14.7%

Variables	Description	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Vehicle type	Passenger car	81.2%	12.7%	17.5%	1.2%
	Motorcycle	9.9%	1.8%	68.4%	52.9%
	Pickup	6.2%	9.4%	5.9%	43.3%
	Truck	0.1%	68.2%	7.4%	0.1%
	Others	2.7%	7.9%	0.8%	2.7%
Week	Monday/Friday	43.4%	45.2%	44.2%	43.7%
	Tuesday–Thursday	28.8%	28.2%	26.7%	29.5%
	Weekend	27.8%	26.6%	29.1%	26.9%
Intersection	No	64.9%	64.8%	76.1%	57.4%
	Yes	35.0%	35.1%	23.9%	42.6%
Time of day	00:00-7:00	15.0%	16.1%	6.2%	16.4%
-	7:00-10:00	35.1%	33.6%	26.3%	36.9%
	10:00-17:00	23.7%	15.9%	17.6%	20.2%
	17:00-21:00	17.1%	11.0%	30.3%	14.7%
	21:00-24:00	9.2%	23.3%	19.7%	11.8%
Collision type	Collision with fixed object	75.3%	82.6%	55.9%	83.5%
	Collision with non-fixed object	2.5%	3.6%	40.2%	3.3%
	Collision with pedestrian	22.2%	13.5%	3.8%	13.2%
	Others	0.1%	0.3%	0.9%	0.8%
Traffic control	No control	50.0%	59.5%	54.5%	59.5%
	Control	49.9%	40.5%	45.5%	40.5%

Table A1. Cont.

References

- 1. National Highway Traffic Safety Administration (NHTSA). Traffic Safety Facts Annual Report Tables. 2018. Available online: https://cdan.nhtsa.gov/tsftables/tsfar.htm# (accessed on 15 June 2021).
- Yan, X.; He, J.; Zhang, C.; Liu, Z.; Qiao, B.; Zhang, H. Single-vehicle crash severity outcome prediction and determinant extraction using tree-based and other non-parametric models. *Accid. Anal. Prev.* 2021, 153, 106034. [CrossRef]
- 3. Traffic Management Bureau of Ministry of Public Security of China. *Statistics Annals of Road Traffic Accident of People's Republic of China (2017)*; Traffic Management Science Institute of Ministry of Public Security: Wuxi, China, 2018.
- 4. Chen, Y.; Luo, R.; Yang, H.; King, M.; Shi, Q. Applying latent class analysis to investigate rural highway single-vehicle fatal crashes in China. *Accid. Anal. Prev.* **2020**, *148*, 105840. [CrossRef]
- Li, Z.; Ci, Y.; Chen, C.; Zhang, G.; Wu, Q.; Qian, Z.; Prevedouros, P.D.; Ma, D.T. Investigation of driver injury severities in rural single-vehicle crashes under rain conditions using mixed logit and latent class models. *Accid. Anal. Prev.* 2019, 124, 219–229. [CrossRef]
- 6. Yu, H.; Yuan, R.; Li, Z.; Zhang, G.; Ma, D.T. Identifying heterogeneous factors for driver injury severity variations in snow-related rural single-vehicle crashes. *Accid. Anal. Prev.* **2020**, *144*, 105587. [CrossRef]
- Xu, C.; Tarko, A.P.; Wang, W.; Liu, P. Predicting crash likelihood and severity on freeways with real-time loop detector data. *Accid. Anal. Prev.* 2013, 57, 30–39. [CrossRef] [PubMed]
- 8. Wu, Q.; Chen, F.; Zhang, G.; Liu, X.C.; Wang, H.; Bogus, S.M. Mixed logit model-based driver injury severity investigations in single-and multi-vehicle crashes on rural two-lane highways. *Accid. Anal. Prev.* **2014**, *72*, 105–115. [CrossRef] [PubMed]
- 9. Wu, Q.; Zhang, G.; Zhu, X.; Liu, X.C.; Tarefder, R. Analysis of driver injury severity in single-vehicle crashes on rural and urban roadways. *Accid. Anal. Prev.* 2016, *94*, 35–45. [CrossRef] [PubMed]
- 10. Yu, H.; Li, Z.; Zhang, G.; Liu, P. A latent class approach for driver injury severity analysis in highway single vehicle crash considering unobserved heterogeneity and temporal influence. *Anal. Methods Accid. Res.* **2019**, *24*, 100110. [CrossRef]
- Xu, C.; Wang, W.; Liu, P. A genetic programming model for real-time crash prediction on freeways. *IEEE T Intell. Transp.* 2012, 14, 574–586. [CrossRef]
- 12. Shangguan, Q.; Fu, T.; Liu, S. Investigating rear-end collision avoidance behavior under varied foggy weather conditions: A study using advanced driving simulator and survival analysis. *Accid. Anal. Prev.* **2020**, *139*, 105499. [CrossRef]
- 13. Ma, D.; Xiao, J.; Ma, X. A decentralized model predictive traffic signal control method with fixed phase sequence for urban networks. *J. Intell. Transport. S.* **2020.** [CrossRef]
- 14. Wen, H.; Xue, G. Injury severity analysis of familiar drivers and unfamiliar drivers in single-vehicle crashes on the mountainous highways. *Accid. Anal. Prev.* 2020, 144, 105667. [CrossRef] [PubMed]

- 15. Wei, F.; Cai, Z.; Guo, Y.; Liu, P.; Wang, Z.; Li, Z. Analysis of roadside accident severity on rural and urban roadways. *Intell. Autom. Soft Comput.* **2021**, *28*, 753–767. [CrossRef]
- Li, Z.; Chen, C.; Wu, Q.; Zhang, G.; Liu, C.; Prevedouros, P.D.; Ma, D.T. Exploring driver injury severity patterns and causes in low visibility related single-vehicle crashes using a finite mixture random parameters model. *Anal. Methods Accid. Res.* 2018, 20, 1–14. [CrossRef]
- 17. Ma, D.; Song, X.; Li, P. Daily traffic flow forecasting through a contextual convolutional recurrent neural network modeling inter-and intra-day traffic patterns. *IEEE T Intell. Transp.* **2021**, *22*, 2627–2636. [CrossRef]
- 18. Guo, Y.; Sayed, T.; Essa, M. Real-time conflict-based Bayesian Tobit models for safety evaluation of signalized intersections. *Accid. Anal. Prev.* **2020**, *144*, 105660. [CrossRef]
- 19. Qin, L.; Li, Z.R.; Chen, Z.; Bill, A.; Noyce, D.A. Understanding driver distractions in fatal crashes: An exploratory empirical analysis. *J. Saf. Res.* 2019, *69*, 23–31. [CrossRef] [PubMed]
- Choudhary, P.; Pawar, N.M.; Velaga, N.R.; Pawar, D.S. Overall performance impairment and crash risk due to distracted driving: A comprehensive analysis using structural equation modelling. *Transp. Res. Part F Traffic Psychol. Behav.* 2020, 74, 120–138. [CrossRef]
- 21. Xie, Y.; Zhao, K.; Huynh, N. Analysis of driver injury severity in rural single–vehicle crashes. *Accid. Anal. Prev.* **2012**, *47*, 36–44. [CrossRef]
- 22. Adanu, E.K.; Hainen, A.; Jones, S. Latent class analysis of factors that influence weekday and weekend single-vehicle crash severities. *Accid. Anal. Prev.* 2018, 113, 187–192. [CrossRef]
- 23. Rahimi, E.; Shamshiripour, A.; Samimi, A.; Mohammadian, A. Investigating the injury severity of single-vehicle truck crashes in a developing country. *Accid. Anal. Prev.* 2020, 137, 105444. [CrossRef]
- 24. Haq, M.T.; Zlatkovic, M.; Ksaibati, K. Investigating occupant injury severity of truck-involved crashes based on vehicle types on a mountainous freeway: A hierarchical Bayesian random intercept approach. *Accid. Anal. Prev.* **2020**, *144*, 105654. [CrossRef]
- 25. Anarkooli, A.J.; Hosseinpour, M.; Kardar, A. Investigation of factors affecting the injury severity of single-vehicle rollover crashes: A random-effects generalized ordered probit model. *Accid. Anal. Prev.* **2017**, *106*, 399–410. [CrossRef]
- 26. Rezapour, M.; Moomen, M.; Ksaibati, K. Ordered logistic models of influencing factors on crash injury severity of single and multiple-vehicle downgrade crashes: A case study in Wyoming. *J. Saf. Res.* **2019**, *68*, 107–118. [CrossRef] [PubMed]
- Vajari, M.A.; Aghabayk, K.; Sadeghian, M.; Shiwakoti, N. A multinomial logit model of motorcycle crash severity at Australian intersections. J. Saf. Res. 2020, 73, 17–24. [CrossRef] [PubMed]
- 28. Islam, S.; Jones, S.L.; Dye, D. Comprehensive analysis of single- and multi-vehicle large truck at-fault crashes on rural and urban roadways in Alabama. *Accid. Anal. Prev.* 2014, *67*, 148–158. [CrossRef]
- Lee, J.; Yasmin, S.; Eluru, N.; Abdel-Aty, M.; Cai, Q. Analysis of crash proportion by vehicle type at traffic analysis zone level: A mixed fractional split multinomial logit modeling approach with spatial effects. *Accid. Anal. Prev.* 2018, 111, 12–22. [CrossRef] [PubMed]
- 30. Rusli, R.; Haque, M.M.; King, M.; Voon, W.S. Single-vehicle crashes along rural mountainous highways in Malaysia: An application of random parameters negative binomial model. *Accid. Anal. Prev.* **2017**, *102*, 153–164. [CrossRef]
- Hou, Q.; Huo, X.; Leng, J.; Cheng, Y. Examination of driver injury severity in freeway single-vehicle crashes using a mixed logit model with heterogeneity-in-means. *Physica A* 2017, 531, 121760. [CrossRef]
- Duddu, V.R.; Pulugurtha, S.S.; Kukkapalli, V.M. Variable categories influencing single-vehicle run-off-road crashes and their severity. *Transp. Eng.* 2020, 2, 100038. [CrossRef]
- 33. Yu, H.; Li, Z.; Zhang, G.; Liu, P.; Ma, T. Fusion convolutional neural network-based interpretation of unobserved heterogeneous factors in driver injury severity outcomes in single-vehicle crashes. *Anal. Methods Accid. Res.* **2021**, *30*, 100157. [CrossRef]
- 34. Mannering, F.L.; Shankar, V.; Bhat, C.R. Unobserved heterogeneity and the statistical analysis of highway accident data. *Anal. Methods Accid. Res.* **2016**, *11*, 1–16. [CrossRef]
- Li, Z.; Wang, W.; Liu, P.; Bigham, J.M.; Ragland, D.R. Using geographically weighted Poisson regression for county-level crash modeling in California. *Saf. Sci.* 2013, *58*, 89–98. [CrossRef]
- 36. Xu, C.; Liu, P.; Wang, W.; Zhang, Y. Real-time identification of traffic conditions prone to injury and non-injury crashes on freeways using genetic programming. *J. Adv. Transport.* **2016**, *50*, 701–716. [CrossRef]
- Xu, C.; Liu, P.; Wang, W.; Li, Z. Evaluation of the impacts of traffic states on crash risks on freeways. *Accid. Anal. Prev.* 2012, 47, 162–171. [CrossRef] [PubMed]
- Roque, C.; Jalayer, M.; Hasan, A.S. Investigation of injury severities in single-vehicle crashes in North Carolina using mixed logit models. J. Saf. Res. 2021, 77, 161–169. [CrossRef]
- 39. Guo, Y.; Wu, Y.; Lu, J.; Zhou, J. Modeling the unobserved heterogeneity in e-bike collision severity using full bayesian random parameters multinomial logit regression. *Sustainability* **2019**, *11*, 2071. [CrossRef]
- 40. Tay, R. A random parameters probit model of urban and rural intersection crashes. Accid. Anal. Prev. 2015, 84, 38–40. [CrossRef]
- 41. Xu, C.; Wang, W.; Liu, P. Identifying crash-prone traffic conditions under different weather on freeways. J. Saf. Res. 2013, 46, 135–144. [CrossRef]
- 42. Buddhavarapu, P.; Scott, J.G.; Prozzi, J.A. Modeling unobserved heterogeneity using finite mixture random parameters for spatially correlated discrete count data. *Transport. Res. Part B Methodol.* **2016**, *91*, 492–510. [CrossRef]

- 43. Mannering, F.L.; Bhat, C.R. Analytic methods in accident research: Methodological frontier and future directions. *Anal. Methods Accid. Res.* 2014, 1, 1–22. [CrossRef]
- 44. Guo, Y.; Liu, P.; Wu, Y.; Chen, J. Evaluating how right-turn treatments affect right-turn-on-red conflicts at signalized intersections. *J. Transp. Saf. Secur.* **2020**, *12*, 419–440. [CrossRef]
- 45. Guo, Y.; Sayed, T.; Zheng, L. A hierarchical Bayesian peak over threshold approach for conflict-based before-after safety evaluation of leading pedestrian intervals. *Accid. Anal. Prev.* 2020, 147, 105772. [CrossRef]
- 46. Azimi, G.; Rahimi, A.; Asgari, H.; Jin, X. Severity analysis for large truck rollover crashes using a random parameter ordered logit model. *Accid. Anal. Prev.* 2020, *135*, 105355. [CrossRef] [PubMed]
- 47. Yu, M.; Zheng, C.; Ma, C.; Shen, J. The temporal stability of factors affecting driver injury severity in run-off-road crashes: A random parameters ordered probit model with heterogeneity in the means approach. *Accid. Anal. Prev.* **2020**, *144*, 105677. [CrossRef] [PubMed]
- 48. Cerwick, D.M.; Gkritza, K.; Shaheed, M.S.; Hans, Z. A comparison of the mixed logit and latent class methods for crash severity analysis. *Anal. Methods Accid. Res.* **2014**, *3*–4, 11–27. [CrossRef]
- 49. Chen, C.; Zhang, G.; Tian, Z.; Bogus, S.M.; Yang, Y. Hierarchical Bayesian random intercept model-based cross-level interaction decomposition for truck driver injury severity investigations. *Accid. Anal. Prev.* **2015**, *85*, 186–198. [CrossRef] [PubMed]
- 50. Guo, Y.; Li, Z.; Liu, P.; Wu, Y. Modeling correlation and heterogeneity in crash rates by collision types using full bayesian random parameters multivariate Tobit model. *Accid. Anal. Prev.* **2019**, *128*, 164–174. [CrossRef]
- 51. Ma, D.; Xiao, J.; Song, X.; Ma, X.; Jin, S. A back-pressure-based model with fixed phase sequences for traffic signal optimization under oversaturated networks. *IEEE Trans. Intell. Transp.* 2020, *99*, 1–12. [CrossRef]
- 52. Chang, F.; Xu, P.; Zhou, H.; Chan, A.H.S.; Huang, H. Investigating injury severities of motorcycle riders: A two-step method integrating latent class cluster analysis and random parameters logit model. *Accid. Anal. Prev.* **2019**, *131*, 316–326. [CrossRef]
- 53. Li, Z.; Chen, C.; Ci, Y.; Zhang, G.; Wu, Q.; Liu, C.; Qian, Z. Examining driver injury severity in intersection-related crashes using cluster analysis and hierarchical Bayesian models. *Accid. Anal. Prev.* **2018**, *120*, 139–151. [CrossRef] [PubMed]
- 54. Li, Z.; Wu, Q.; Ci, Y.; Chen, C.; Chen, X.; Zhang, G. Using latent class analysis and mixed logit model to explore risk factors on driver injury severity in single-vehicle crashes. *Accid. Anal. Prev.* **2019**, *129*, 230–240. [CrossRef] [PubMed]
- 55. Liu, P.; Fan, W.D. Exploring injury severity in head-on crashes using latent class clustering analysis and mixed logit model: A case study of North Carolina. *Accid. Anal. Prev.* **2020**, *135*, 105388. [CrossRef]
- 56. Zeng, Q.; Wen, H.; Huang, H.; Wang, J.; Lee, J. Analysis of crash frequency using a Bayesian underreporting count model with spatial correlation. *Physica A* **2020**, *545*, 123754. [CrossRef]
- 57. Guo, Y.; Sayed, T.; Zheng, L.; Essa, M. An extreme value theory based approach for calibration of microsimulation models for safety analysis. *Simul. Model. Pract. Theory* **2021**, *106*, 102172. [CrossRef]
- 58. Bi, H.; Ye, Z. Exploring ride sourcing trip patterns by fusing multi-source data: A big data approach. *Sustain. Cities Soc.* **2021**, *64*, 102499. [CrossRef]
- Yang, D.; Xie, K.; Ozbay, K.; Yang, H. Fusing crash data and surrogate safety measures for safety assessment: Development of a structural equation model with conditional autoregressive spatial effect and random parameters. *Accid. Anal. Prev.* 2021, 152, 105971. [CrossRef] [PubMed]
- 60. Zeng, Q.; Gu, W.; Zhang, X.; Wen, H.; Lee, J.; Hao, W. Analyzing freeway crash severity using a Bayesian spatial generalized ordered logit model with conditional autoregressive priors. *Accid. Anal. Prev.* **2019**, 127, 87–95. [CrossRef]
- 61. Klassen, J.; El-Basyouny, K.; Islam, M.T. Analyzing the severity of bicycle-motor vehicle collision using spatial mixed logit models: A city of Edmonton case study. *Saf. Sci.* **2014**, *62*, 295–304. [CrossRef]
- 62. Munira, S.; Sener, I.N.; Dai, B. A Bayesian spatial Poisson-lognormal model to examine pedestrian crash severity at signalized intersections. *Accid. Anal. Prev.* 2020, 144, 105679. [CrossRef]
- 63. Cai, Q.; Abdel-Aty, M.; Lee, J.; Wang, L.; Wang, X. Developing a grouped random parameters multivariate spatial model to explore zonal effects for segment and intersection crash modeling. *Anal. Methods Accid. Res.* **2018**, *19*, 1–15. [CrossRef]
- 64. Zeng, Q.; Wen, H.; Huang, H.; Abdel-Aty, M. A Bayesian spatial random parameters Tobit model for analyzing crash rates on roadway segments. *Accid. Anal. Prev.* 2017, 100, 37–43. [CrossRef]
- 65. Huang, H.; Chang, F.; Zhou, H.; Lee, J. Modeling unobserved heterogeneity for zonal crash frequencies: A Bayesian multivariate random-parameters model with mixture components for spatially correlated data. *Anal. Methods Accid. Res.* **2019**, *24*, 100105. [CrossRef]
- 66. Linzer, D.A.; Lewis, J. poLCA: An R Package for polytomous variable latent class analysis. J. Stat. Softw. 2011, 42, 1–29. [CrossRef]
- 67. Xu, X.; Xie, S.; Wong, S.C.; Xu, P.; Huang, H.; Pei, X. Severity of pedestrian injuries due to traffic crashes at signalized intersections in Hong Kong: A Bayesian spatial logit model. *J. Adv. Transport.* **2017**, *50*, 2015–2028. [CrossRef]
- 68. Banerjee, S.; Carlin, B.P.; Gelfand, A.E. *Hierarchical Modeling and Analysis for Spatial Data*, 2rd ed.; Chapman & Hall; CRC Press: Boca Raton, FL, USA, 2014.
- 69. Dong, N.; Huang, H.; Xu, P.; Ding, Z.; Wang, D. Evaluating spatial-proximity structures in crash prediction models at the level of traffic analysis zones. *Transp. Res. Rec.* 2014, 2432, 46–52. [CrossRef]
- Aguero-Valverde, J.; Jovanis, P. Spatial correlation in multilevel crash frequency models: Effects of different neighboring structures. *Transport. Res. Rec.* 2010, 2165, 21–32. [CrossRef]

- El-Basyouny, K.; Sayed, T. Urban arterial accident prediction models with spatial effects. *Transport. Res. Rec.* 2009, 2102, 27–33. [CrossRef]
- 72. Guo, Y.; Osama, A.; Sayed, T. A cross-comparison of different techniques for modeling macro-level cyclist crashes. *Accid. Anal. Prev.* 2018, 113, 38–46. [CrossRef]
- 73. Feng, S.; Li, Z.; Ci, Y.; Zhang, G. Risk factors affecting fatal bus accident severity: Their impact on different types of bus drivers. *Accid. Anal. Prev.* **2016**, *86*, 29–39. [CrossRef]
- 74. Naik, B.; Tung, L.W.; Zhao, S.; Khattak, A.J. Weather impacts on single-vehicle truck crash injury severity. *J. Saf. Res.* **2016**, *58*, 57–65. [CrossRef] [PubMed]
- 75. Abdel-Aty, M.; Ekram, A.; Huang, H.; Choi, K. A study on crashes related to visibility obstruction due to fog and smoke. *Accid. Anal. Prev.* 2011, 43, 1730–1737. [CrossRef] [PubMed]
- 76. Bédard, M.; Guyatt, G.H.; Stones, M.J.; Hirdes, J.P. The independent contribution of driver, crash, and vehicle characteristics to driver fatalities. *Accid. Anal. Prev.* 2002, *34*, 717–727. [CrossRef]
- 77. Zhang, T.; Chan, A.H.; Zhang, W. Dimensions of driving anger and their relationships with aberrant driving. *Accid. Anal. Prev.* **2015**, *81*, 124–133. [CrossRef] [PubMed]
- Zhu, X.; Srinivasan, S. A comprehensive analysis of factors influencing the injury severity of large-truck crashes. *Accid. Anal. Prev.* 2011, 43, 49–57. [CrossRef] [PubMed]