Investigating the severity of expressway crash based on the random parameter logit model accounting for unobserved heterogeneity

Fei Ye¹,², Wen Cheng²,³, Changshuai Wang⁴, Haoxue Liu¹ and Jiping Bai²

Abstract
The present study utilized a random parameter logit (RPL) model to explore the nonlinear relationship between explanatory variables and the likelihood of expressway crash severity. The potential unobserved heterogeneity of data brought by China’s road traffic characteristics was fully considered. A total of 1154 crashes happened on Hang-Jin-Qu Expressway from 2013 to 2018 were analyzed. In addition to the conventional impact factors considered in the past, variables related to road geometry were also introduced, which contributed to expressway accidents significantly. The overall stability of the model estimation was examined by likelihood ratio test. Then, the average elastic coefficient of the significant factors at each severity level was also calculated. Several factors that significantly increase the fatal crash probability were highlighted: rainy/snowy/cloudy weather condition, low visibility (100–m), night without light, wet-skid road surface, being female, aged 41+ years, collision with a rigid barrier and some other obstacles, radius and length of horizontal curve, and longitudinal gradient. The parameters of four factors were random and obeyed normal distribution: night without light, being female, driving experience with 10+ years and with large vehicle responsible. These findings provide insights for better understanding of expressway crash severity. Some countermeasures were proposed about driver education, traffic law enforcement, vehicle and road design, environmental improvement, and so on.

Keywords
Traffic safety, expressway crash, injury severity, random parameter logit, elasticity analysis

Date received: 26 July 2021; accepted: 26 November 2021
Handling Editor: Chenhui Liang

Introduction
The high mortality rate caused by road traffic accidents has imposed heavy economic and emotional burdens to the family and society. Tremendous efforts have been made to reduce the frequency and severity of accidents.¹ Some kinds of statistical models were usually built based on crash data to explore the relationship between the accident severity and traffic occupants, vehicles, road infrastructures, environmental conditions. And also, measures have been proposed based on the comprehensive understanding of factors affecting the risks and severity.

The crash rate of the expressway was lower than the other types of roads (such as an urban road).² But the expressway was usually ranked first in all road types

¹School of Transportation Engineering, Chang’an University, Xi’an, China
²School of Rail Transit, Zhejiang Institute of Communications, Hangzhou, China
³School of Economic and Management, Chang’an University, Xi’an, China
⁴School of Transportation, Southeast University, Nanjing, China

Corresponding author:
Fei Ye, School of Transportation Engineering, Chang’an University, 33# Chang’an Middle Road, Xi’an 710064, China.
Email: yefei@qjvtit.edu.cn
for the aspects of mortality. According to statistics of Transportation Administration of China, the expressway crashes account for about 5% of road traffic crashes, but the number of deaths accounts for 10%. Traffic safety issue of expressway has been a hot topic of management agencies and researchers for many decades. Most of them focused on investigating the rate of crashes, but the crash severity has not been paid enough attention to our best knowledge. Therefore, it is valuable to conduct a comprehensive analysis of the factors affecting the expressway crash severity.

However, there are differences in the statistical calibers of traffic accidents due to the various stages of development in different countries. The data integrity needs to be improved, especially in developing countries. Injury severity conditional on crash occurrence can depend on numerous factors all of which are most certainly not observed in crash databases. These unobserved factors can moderate the influence of other observed covariates in the model leading to variation in the parameter effects across different observations. These unobserved variations are referred to as “unobserved heterogeneity,” which is of considerable importance in injury severity analysis. More attention should be attached to the heterogeneity in the modeling process. Judging from the modeling methods cited by scholars, the random parameter logit (RPL) model is still the most widely used.

The current study intents to optimize the traditional discrete choice models with fixed parameters into the RPL model with some random parameters for quantitative analysis of traffic accidents by considering the unobserved heterogeneity among predictor variables. The factors about driver, vehicle, road, and environment are considered as independent variables. The responsible and accuracy of the model are verified, and then the average elastic coefficients of the significant factors at each severity level of the collision are calculated. Some corresponding improvement strategies are proposed based on the findings.

Literature review

Some researchers have studied the crash severity regarding drivers, vehicles, roads, and the environment conditions in recent years. The severe weather and unfavorable segments would increase the crash severity in the merging and diverging locations of freeways. Many significant factors including the average speed of road section, the average daily traffic volume, time periods, weather conditions, the physical characteristics of accident area, and the cause of accident were identified based on predicting the injury severity level and analyzing the influence extent through multiple logistic regression model. The findings showed significant differences in the impact of driving behaviors, environmental conditions, and some other factors on the accident rate under different severity levels. Following from above analysis, we found that many scholars have studied the field of crash severity from various aspects, especially the influencing factors, but there are still some unknown or controversial factors, such as road alignment.

The logistic regression is a promising statistical model that provides significant interpretations in promoting future safety performance based on the results of accident-related data analyzing. Discrete choice models (e.g. Logit and Probit) are more suitable for solving the discrete dependent variable problem of crash severity, and they have also been widely used. The multinomial logit (MNL) model was effective in exploring the risk factors that affect the injury severity of 11,771 traffic accidents in 6 years of Turkey. A joint Poisson regression model was presented with multivariate normal heterogeneities of crash frequency by severity level for freeway sections, considered the presence of common unobserved factors that influence crash frequencies of different severity levels.

And also, the random parameter models, ordered logit models, and latent class cluster models were applied to solve the unobserved heterogeneity problem. The RPL model was developed to explore the driver injury severity in single-vehicle crashes by setting the regression coefficients of age and gender as random parameters. The accidents were divided into serious and non-serious accidents and then compared the fitting effects of the fixed-parameter logit model, RPL model, and random forest model. The results showed that the RPL model and random forest model had better fitness. It was necessary to consider the nonlinear relationship between variables and individual heterogeneity ignored in the traditional logit model. By studying the characteristics of traffic accidents in remote mountainous areas, three random parameter models (RPNB, RPNB-L, RPNB-GE) were constructed to solve the heterogeneity caused by missing variables and verified that the RPNB-L model was effective in terms of prediction ability and superiority in fitness. Numerous studies have shown that the combination of steep slopes, horizontal curves, and sharp bends with insufficient visibility in rainy weather would increase accidents. According to the order characteristics of injury severity, Rezapour et al. used ordered logit models to investigate the impact of various factors on crash injury severity of single and multiple-vehicle downgrade accidents. The results showed that there was a significant difference between the two types of accidents. In order to assess the impact of driver’s age on the severity of accidents accounting for unobserved heterogeneity and age group differences, Osman et al. constructed a
mixed generalized ordered response probit model to analyze the injury severity of commercially licensed drivers involved in single-vehicle crashes considering the discrete ordinal nature of injury severity data.

In addition to the statistical and econometric methods, the data-driven methods such as those relating to data mining, artificial intelligence, machine learning, neural networks, support vector machines, and others, were also widely used in the analysis of traffic accident data. Such methods have the potential to handle extremely large amounts of data and provide a high level of prediction accuracy. But, they may not truly understand the effects of specific factors on the resulting injury probabilities. However, the heterogeneity in modeling has been paid attention to in recent years. Some scholars account for potential unobserved heterogeneity by extending traditional statistical and econometric methods models.

Samples were grouped and the latent class cluster analysis were performed to identify homogenous subgroups for a specific crash type-pedestrian crashes. The influencing factors of single-vehicle crashes severity were investigated by developing a latent class logit model as an alternative to the frequently used random parameters models to account for unobserved heterogeneity across observations. The heterogeneity may be attributed to the interaction between different types of factors, including roadway conditions, demographic and behavioral attributes of drivers, environmental conditions, etc. Ignoring the heterogeneity for modeling may lead to inaccurate parameter estimation and bias in analyzing causal factors.

Each of these methods has an implicit trade-off between practical prediction accuracy and their ability to uncover underlying causal relationships. In particular, these methods such as the RPL models, are less performing than machine learning techniques but have the advantage of numerically expressing the impact of certain factors. In the literature, the potential unobserved heterogeneity was identified as an important issue, but the studies in this area were still rare. These effects gave support to the application of the RPL models in injury severity research and verified the individual heterogeneity based on some factors.

**Data preparation**

**Data pre-processing**

The data used in this study were about crashes from the Hang-Jin-Qu Expressway in Zhejiang Province of China, which is designed to speed of 120 km/h and has six lanes in two-ways. Totally 1443 traffic accidents from 2013 to 2018 in the 290 km long sections were recorded in the road maintenance management system of Zhejiang Transportation Group. Excluding the accidents in the ramp entrances of the service area and interchanges, the remaining 1154 accidents with complete information were taken as a sample set. In the original traffic accident data, there were four levels of crash severity according to the definition of the China Ministry of Public Security, including no injury (property damage only), minor injury, serious injury, and fatal. The level of injury sustained by the most severely injured vehicle occupant defined the crash severity. Although there were many ways to classify the crash severity, the most commonly used method was to arrange the severity level into three categories: fatal, injury, and property damage only (PDO). Therefore, this study combines minor injury and severe injury accidents into injury accidents for statistical analysis. The additional information recorded in the system includes driver factors, vehicle factors, road conditions, and environmental conditions. The road design data such as horizontal curve properties and longitudinal profile alignment of the expressway were extracted from the construction drawing design documents. The data of horizontal curve properties include radius and length of horizontal curve. The data of longitudinal profile alignment include longitudinal gradient. The other information about driver and vehicle factors, road, and environmental conditions were generated from the reports of expressway traffic police. Because of the record deviation of the police in collecting accident information or the artificial operation deviation of the modeling personnel, the existence of abnormal values is inevitable. Therefore, a distance-based outlier detection algorithm was applied to preprocess the missing values, outliers, and consistency in the data to improve the model accuracy.

**Data description**

The dependent variable was the crash severity level. The independent variables included in the model were divided into discrete and continuous variables according to their attributes. Table 1 shows the explanatory variables selected in the study and their descriptive statistical characteristics. There were a total of 1154 accidents recorded in three categories including 424 PDO accidents (36.7%), 456 injury accidents (39.5%), and 274 fatal accidents (23.7%). Considering the weather conditions, most accidents (85.1%) usually occur on sunny days (982 times); meantime a high level of accident severity being recognized as a fatal accident always occurs on sunny days (237 times). For visibility, the largest number of accidents occur in high visibility (200 + m) condition (968 times). Furthermore, a greater number of crashes occur on dry road surface (901 times).
| Categorical variable | Description | Description | Injury severity |
|----------------------|-------------|-------------|-----------------|
| Total PDO Injury     | Freq. | % | Freq. | % | Freq. | % | Freq. | % |
| Output Severity      | 1154 | 100 | 424 | 36.7 | 456 | 39.5 | 274 | 23.7 |
| Environmental conditions | 982 | 85.1 | 363 | 85.6 | 382 | 83.8 | 237 | 86.5 |
| Weather conditions   | 968 | 83.9 | 352 | 83.0 | 380 | 83.3 | 236 | 86.1 |
| Rainy/snowy/cloudy  | 126 | 10.9 | 48 | 11.3 | 53 | 11.6 | 25 | 9.1 |
| Visibility           | 60 | 5.2 | 24 | 5.7 | 23 | 5.0 | 13 | 4.7 |
| Daylight*            | 675 | 58.5 | 239 | 56.4 | 286 | 62.7 | 150 | 54.7 |
| Lighting conditions  | 59 | 5.1 | 24 | 5.7 | 17 | 3.7 | 18 | 6.6 |
| Night with light     | 420 | 36.4 | 161 | 38.0 | 153 | 33.6 | 106 | 38.7 |
| Road conditions      | 901 | 78.1 | 340 | 80.2 | 343 | 75.2 | 218 | 79.6 |
| Road surface         | 929 | 80.5 | 398 | 93.9 | 327 | 71.7 | 204 | 74.5 |
| Dry *                | 225 | 19.5 | 26 | 6.1 | 129 | 28.3 | 70 | 25.5 |
| Terrain type         | 1121 | 97.1 | 416 | 98.1 | 432 | 94.7 | 273 | 99.6 |
| Level*               | 1121 | 97.1 | 416 | 98.1 | 432 | 94.7 | 273 | 99.6 |
| Vertical alignment   | 266 | 23.1 | 101 | 23.8 | 100 | 21.9 | 65 | 23.7 |
| Gender               | 94 | 8.2 | 29 | 6.8 | 40 | 8.8 | 25 | 9.1 |
| Male*                | 253 | 21.9 | 84 | 19.8 | 113 | 24.8 | 56 | 20.4 |
| Female               | 175 | 41.3 | 341 | 74.8 | 208 | 75.9 | 121 | 10.5 |
| Age 26–40 years*     | 506 | 43.8 | 196 | 46.9 | 196 | 43.0 | 111 | 40.5 |
| 19–25 years          | 94 | 8.2 | 29 | 6.8 | 40 | 8.8 | 25 | 9.1 |
| 41+ years*           | 554 | 48.0 | 196 | 46.2 | 220 | 48.3 | 138 | 50.4 |
| Driving experience   | 131 | 10.5 | 70 | 15.6 | 32 | 7.0 | 19 | 6.9 |
| 2–3 years*           | 309 | 26.8 | 179 | 42.2 | 83 | 18.2 | 47 | 17.2 |
| 10+ years*           | 869 | 75.3 | 365 | 86.1 | 319 | 70.0 | 185 | 67.5 |
| Collision type       | 169 | 14.6 | 36 | 8.5 | 86 | 18.9 | 47 | 17.2 |
| Collision between vehicles* | 116 | 10.1 | 23 | 5.4 | 51 | 11.2 | 42 | 15.3 |
| Responsible vehicle type | 838 | 72.6 | 234 | 55.2 | 378 | 82.9 | 226 | 82.5 |
| Small vehicle*       | 60 | 5.2 | 29 | 6.8 | 23 | 5.0 | 8 | 2.9 |
| Middle vehicle       | 256 | 22.2 | 161 | 38.0 | 55 | 12.1 | 40 | 14.6 |
| Large vehicle        | 1119 | 97.0 | 399 | 94.1 | 452 | 99.1 | 268 | 97.8 |
| Movement of vehicle  | 35 | 3.0 | 25 | 5.9 | 4 | 0.9 | 6 | 2.2 |
| Going straight*      | 0.170 | 0.462 | 0.131 | 0.372 | 0.199 | 0.524 | 0.183 | 0.475 |
| Length of horizontal curve | 0.688 | 1.239 | 0.739 | 1.319 | 0.609 | 1.118 | 0.742 | 1.295 |
| Longitudinal gradient | 0.124 | 0.920 | 0.151 | 0.855 | 0.104 | 0.942 | 0.116 | 1.008 |

Freq.: frequency; SD: standard deviation. The reference category.
Model formulation

The RPL model

The discrete choice models may be suitable for modeling the injury severity level since it is a discrete outcome. The MNL model is the basic model with a simple form and requires a lower sample size in discrete choice models. Some other types of logit models were evolved from the MNL models. The continuous variables can be directly incorporated into model calculations, while the discrete variables need to be preprocessed to adapt to the model. The dichotomous variables in the discrete variables can be coded by 0 or 1 directly. If the number of category is more than 2, we can introduce the dummy variables. When the severity category is I (I ≥ 3), the probability can be derived as follows:

\[ P_n(i) = P(U_{in} \geq U_{jn}), \forall j \neq i, i, j \in I \] (1)

Where, \( P_n(i) \) is the probability that a crash severity category \( i \) occurs in an observed crash \( n \), \( U_{in} \) is a linear function that determines the severity of the crash \( n \). As usual, \( U_{in} \) can be linearly formed by equation (2).

\[ U_{in} = \beta_i X_n + e_{in} \] (2)

Where, \( X_n \) is a vector of explanatory variables (risk factors) that affect the severity level, \( \beta_i \) is a vector of estimable parameters, and \( e_{in} \) is a disturbance term that takes into account the unobserved effects. Assumed that the error term \( e_{in} \) are generalized extreme value distributed independently, a MNL model can be derived to estimate the probability of a crash severity category with the expression in equation (3).

\[ P_n(i) = \frac{\exp(\beta_i X_n)}{\sum_j \exp(\beta_j X_n)}, \forall j \neq i, i, j \in I \] (3)

In this case, the error terms \( (e_{in}) \) are supposed to be independent in each of the severity categories. The regression coefficient \( (\beta_i) \) in equation (3) can be estimated by the standard maximum likelihood estimation method. Before using the MNL models for analysis, it is necessary to set the reference category in advance. However, the MNL models was restricted by the independence of irrelevant alternatives (IIA) problem, that is, its hypothesis that there is no correlation between dependent variables. When one category of the dependent variable is added or removed, it will not affect the occurrence probability of other categories. But in practice, this assumption is usually not satisfied. If some severity categories share unobserved effects (i.e. have correlated disturbances), the most commonly used tests are the Hausman and McFadden (HM) test and the Small and Hsiao (SH) test.

In addition to the IIA assumption, the MNL models also assume that there are no difference between the sample individuals, that is, the effects of the variables in the models on the dependent variables are fixed. In order to solve MNL models’ failure to consider the limitations of individual differences and IIA assumptions, the RPL models are widely applied in various fields, including in researches about traffic accident injury severity. By allowing the parameters of all or part of the independent variables to vary across the observations, the RPL model can explain the unobserved heterogeneity in the data, and will not be limited by the “IIA” assumption. According to random utility theory, the utility function that determines the injury severity of accident can be expressed as equation (2). In order to capture the effects of unobserved heterogeneity due to randomness associated with some of the factors necessary to understand injury severity, the RPL model is generated as equation (4) by extending the MNL models.

\[ P_n(i|\varphi) = \frac{\exp(\beta_i X_n)}{\sum \exp(\beta_j X_n)} \int f(\beta_i|\varphi) d\beta_i \] (4)

Where \( P_n(i|\varphi) \) is the probability of injury severity level \( i \) conditional on \( f(\beta_i|\varphi) \), \( f(\beta_i|\varphi) \) is the density function of \( \beta_i \) and \( \varphi \) is the parameter vector with known density function. The probability is calculated by weighted average for different values of \( \beta_i \) across observations. Typically, some elements of \( \beta_i \) are fixed and some are randomly distributed with specific statistical distribution. If the variance of \( \varphi \) is statistically significant, the modeled injury severity levels vary with \( X \) across observations. The probability is a weighted average of different coefficient values \( \beta_i \) in an accident, which may be fixed or random. If all parameters are randomly distributed, the weighted value is determined by the probability density function \( f(\beta_i|\varphi) \). If all parameters are fixed distribution, the RPL model is equivalent to the standard MNL model. The regression coefficient in equation (4) can be estimated by simulating maximum likelihood estimation. Generally, the probability density functions of normal distribution, lognormal distribution, triangular distribution, and uniform distribution are considered. Previous studies have proved that the normal distribution is the most suitable one for analyzing accident injury severity data. In this study, maximum likelihood estimation is performed through a simulation-based approach to overcome the computational complexity of estimating the parameters \( \beta_i \) of the RPL model.

Model test and evaluation

Usually, it is necessary to test the validity of the model in a regression analysis. The most common methods are
the likelihood ratio test or the Chi-square test, which can be calculated by equation (5).

\[ \chi^2 = -2(LL(0) - LL(\beta)) \sim \chi^2(K) \tag{5} \]

Where, \( LL(0) \) is the initial value of the log-likelihood at zero, namely, the value of the log-likelihood when the model only includes a constant term that without independent variables; \( LL(\beta) \) is the convergence value of the log-likelihood function, namely, the value of the log-likelihood function when all significant independent variables and constant terms are included in the model; \( K \) is the number of parameters in the model, \( \chi^2(K) \) is the chi-squared value (critical value) for a given significance level. If the \( \chi^2 \) is greater than \( \chi^2(K) \), it indicates that the model with added parameters is better than the model with constant terms only, which means that the model is effective. The Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and McFadden’s Pseudo R² are usually used to evaluate the fitness of a theoretical model. The calculation methods of each evaluation index are shown in the following formulas:

\[ AIC = -2LL(\beta) + 2K \tag{6} \]
\[ BIC = -2LL(\beta) + \ln(N) \times K \tag{7} \]
\[ \text{Pseudo}R^2 = 1 - \frac{LL(\beta)}{LL(0)} \tag{8} \]

Where \( N \) is the number of crashes. The smaller the value of the AIC and BIC indicators, the higher the model fitness. The McFadden’s Pseudo R² value ranges from 0 to 1, which indicates that the model get better fitness when the value is closer to 1.

**Interpretation of parameters**

The regression parameters obtained by using the MNL logit model can only qualitatively but not quantitatively reflect the impact of significant independent variables on crash severity. It means that the model parameters cannot be interpreted directly. In order to paraphrase the impact of each independent variable on the injury severity, the normally derivatives of the injury severity probabilities are calculated. When giving a change in an observed factor, \( x_{nm} \), the change of the probability that crash \( n \) is severity \( i \) can be calculated as equation (9).

\[ \frac{\partial P_n(i)}{\partial x_{nm}} = \frac{\partial S_m}{\partial x_{nm}} P_n(i)(1-P_n(i)) \tag{9} \]

If the explanatory variable is dichotomous, we must introduce pseudo-elasticity to assess the effect of individual parameter estimates on crash severity probabilities. The direct pseudo-elasticities can be computed as equation (10).

\[ E_{X_{nm}}^{P(i)} = \left[ \exp(\beta_m) \frac{\sum_i \exp(ASC_i + \beta_i X_n)}{\sum_i \exp(\Delta(ASC_i + \beta_i X_n))} - 1 \right] \times 100 \tag{10} \]

Where, \( E_{X_{nm}}^{P(i)} \) is the direct pseudo-elasticity of the \( nth \) variable from the vector \( X_n \), \( i \) is the number of possible severity categories, \( ASC_i + \beta_i X_n \) is the value of the function determining the severity category when \( x_{nm} \) equals zero, and \( \Delta(ASC_i + \beta_i X_n) \) is the value after \( x_{nm} \) has been changed from zero to one.

**Results**

**Model parameter estimation**

The insignificant variables were excluded successively with the significance level of 0.1 in each utility function. According to the testing results for multicolinearity with Variance Inflation Factors (VIF), the remaining variables were all treated as independent variables. The parameters were estimated by taking the PDO crash as a baseline item. At the same time, a reference item was specified for each multi-classification explanatory variables. The specific reference item was marked and listed in Table 1. In order to obtain more accurate calibration results, the utility functions for different levels of accident severity were estimated. To test the fitness of the RPL model, a fixed parameter MNL model was constructed for comparison. The likelihood ratio statistics were constructed to test whether it was significant or not.

The \( \chi^2 \) value of the MNL model was 1113.256, which was greater than the critical value of 73.683 (taking the significance level of 0.01). The RPL model was constructed and the significance of the model was examined. The utility function was consistent with the MNL model, but some of the fixed parameters were changed into random parameters. The \( \chi^2 \) value of the RPL model was 1903.308, which was greater than the critical value of 78.616 (taking the significance level of 0.01). A total of four variables with random parameters were identified through the RPL model, that was why the freedom of the RPL model was 52 but 48 for the MNL model. The log-likelihood with constant only of the two models were the same, but the log-likelihood convergence value increased from -1438.453 of the MNL model to -1043.427 of the RPL model. The McFadden R² value constructed by log-likelihood convergence value and log-likelihood value with only constant term was 0.477, which was higher than 0.279 of MNL model. It indicated that the fitness of RPL model was better than MNL model, which was consistent with previous studies on traffic safety analysis. The results of parameter estimation and model comparison were summarized in Table 2, all the parameters included in the
model were statistically significant at a 0.10 significance level or higher. The tabulated values outside the parentheses were the posterior means of parameters, and those inside the parentheses were their posterior standard deviations. Although the effective parameters of the traditional MNL model were less than the RPL model, the RPL model still showed a low AIC value (2190.854) (AIC of the MNL model was 2972.906). The difference of AIC was considered to be substantial, which indicated that the RPL model was more desirable than the MNL model.

As shown in the last three rows of Table 2, the superiority of the RPL model in the fitness was further confirmed by its higher classification accuracy for each severity level and the entire dataset. We identified some significant variables that contributed to injury and fatal crashes in the MNL model. Further comparison between the two models revealed that they were still significant in the RPL model, and their coefficients were relatively similar in the two models. It gave support to the consistency between the two models to some extent.

Parameters producing statistically significant standard errors for their assumed distribution were found to be random, which were: the indicator variables for driver gender (female), responsible vehicle type (large vehicle) for injury crashes, and the indicator variables for driving experience (10 + years), lighting conditions (night without light) for fatal crashes. When the estimated standard errors were not statistically different from zero, the parameters were fixed across the observations. Normal distribution appeared to provide the best statistical fitness for these random parameters. The mean and standard deviations were shown in Table 2. The average elastic coefficients of the significant factors in the RPL model were list in Table 3.

**Interpretation of the results**

The significant variables for each severity outcome were discussed next. They were categorized into five parts.

**Effects of driver factors.** Regarding the effects of the driver’s gender, the parameter of the indicator variable was found to be significant and random across injury crashes but fixed for fatal crashes. The female drivers were more likely to result in a fatal crash outcome, but for injury crashes, the parameters of driver’s gender were normally distributed with a mean of 1.82 and a standard deviation of 1.23. This means that for the majority (93.1%) of the expressway crashes, a female driver would increase the likelihood of an injury outcome, while for 6.9% of the crashes, a female driver would decrease the likelihood of an injury outcome. The probability of injury accidents increased by 14.4%, and the probability of fatal accidents increased by 15.7%. This random effect is likely because female drivers are weaker than male drivers for their inherent personality traits and psychological characteristics. Female drivers’ psychological endurance, emergency handling ability in the face of crisis, and strain ability are lower than those of male drivers, and their judgment of distance and speed is also less accurate, which results in higher casualty rates of accidents than male drivers.

The impact of driver’s age on the severity of the accident showed a trend of polarization. Compared with 26-40 years, the age of 19–25 years was negatively correlated with accident severity, and the age of 41 + years was positively correlated with accident severity. On the other hand, the accidents with drivers’ age of 19–25 years were less likely to result in injury or fatal outcomes, which was 1.1% for injury and 0.3% for fatal crashes. The drivers aged 41 + years would increase the probability by 1.0% for injury and 1.3% for fatal outcomes. That might be due to the risk compensation effects. The unfamiliar with the vehicle, the road situation and the lack of driving experience of the younger drivers would lead to their more cautious driving behaviors, which reduced the crash severity. Regarding the impact of driver age on the severity of accidents, relevant studies have not drawn unanimously recognized conclusions. With the increase of age, most drivers' physical function and health would not be as good as younger drivers, and they were more likely to sustain a serious physical injury in the accident. In addition, the drivers’ reaction time also increases with age. In the high-speed driving environment, traffic accidents are more likely to cause serious casualties.

Regarding the years of driving experience, the probability of injury and fatal outcomes were both reduced for driving experience of 2–5 years and driving experience of 10 + years. For injury level, the reductions were 7.7% and 13.0%. And for fatal level, the reductions were 9.9% and 17.7%. The parameters of driving experience 10 + years in the utility function of fatal level were random. For 89.5% of accidents with drivers’ driving experience 10 + years, the probability of fatal level was lower than that with drivers’ driving experience of 10–25 years. The potential reasons are as follows: (a) Drivers with driving experience of 2–5 years tend to strictly follow the speed limit on expressway, so that their driving speeds are relatively lower. In addition, they concentrate highly on driving task, which gives rise to lower severity of accidents. (b) For drivers with much more experience, the driving skills and risk perception are higher than those with less experience, so the accident severity is also relatively lower. In short, the drivers with medium driving age (3–10 years) will generally cause more serious outcomes than other
Table 2. Coefficients (and standard errors) of the logit regression models for injury severity outcomes.

| Variable                        | Description                  | MNL Injury (standard error) | MNL Fatal (standard error) | RPL Injury (standard error) | RPL Fatal (standard error) |
|---------------------------------|------------------------------|-----------------------------|----------------------------|----------------------------|----------------------------|
| **Environmental conditions**    |                              |                             |                            |                            |                            |
| Weather conditions              | Rainy/snowy/cloudy           | 0.332 (0.092)**             | 0.699 (0.221)**            | 0.214 (0.149)**            | 0.772 (0.151)**            |
| Visibility                      | Medium visibility (100–200 m)| 0.119 (0.097)**             | 0.301 (0.039)**            | 0.137 (0.091)**            | 0.336 (0.042)**            |
| Low visibility (100–m)          | Low visibility               | -0.253 (0.062)**            | -0.501 (0.041)**           | -0.316 (0.066)**           | -0.858 (0.302)**           |
| Night with light                | Night without light          | 0.217 (0.174)**             | 0.797 (0.213)**            | 0.339 (0.315)**            | 1.103 (0.117)**            |
| SD for random parameter         |                              |                             |                            |                            |                            |
| **Environmental conditions**    | Wet-skid                     | 0.323 (0.173)**             |                             | 0.637 (0.217)**            |                             |
| **Terrain type**                |                              |                             | 2.413 (1.172)**            | 2.005 (1.510)**            |                             |
| Vertical alignment              | Upgrade/downgrade            | 0.237 (0.033)**             |                             | 0.328 (0.104)**            |                             |
| Road geometry                   | Radius of horizontal curve   | 0.254 (0.091)**             | 0.463 (0.131)**            | 0.286 (0.039)**            | 0.313 (0.062)**            |
|                                 | Length of horizontal curve   | -0.133 (0.092)**            | -0.173 (0.011)**           | -0.132 (0.018)**           | -0.192 (0.014)**           |
|                                 | Longitudinal gradient        | 0.227 (0.081)**             |                             |                             | 0.136 (0.026)**            |
| **Driver factors**              |                              |                             |                            |                            |                            |
| Gender                          | Female                       | 1.424 (0.293)**             |                             | 1.824 (0.195)**            |                             |
|                                 | Age                          |                             |                             |                             |                             |
|                                 | 19–25 years                  | -0.317 (0.177)**            |                             | -0.322 (0.221)**           | -0.199 (0.115)**           |
|                                 | 41+ years                    | 1.282 (0.491)**             | 2.119 (0.453)**            | 1.535 (0.335)**            | 2.370 (0.457)**            |
|                                 | Driving experience           |                             |                             |                             |                             |
|                                 | 2–years                      | 1.445 (0.271)**             |                             | -1.445 (0.278)**           |                             |
|                                 | 10+ years                    | -0.747 (0.312)**            |                             | -1.317 (0.414)**           |                             |
|                                 | SD for random parameter      |                             |                             |                             |                             |
| **Vehicle factors**             |                              |                             |                            |                            |                            |
| Collision type                  | Collision with a rigid barrier| 0.811 (0.312)**             |                             | 0.973 (0.356)**            |                             |
|                                 | Collision with other facilities| 1.222 (0.472)**             | 1.675 (0.521)**            | 1.218 (0.458)**            | 1.819 (0.478)**            |
| Responsible vehicle type        | Middle vehicle               | -0.239 (0.073)**            |                             |                             |                             |
|                                 | Large vehicle                | 1.177 (0.241)**             |                             | 1.283 (0.285)**            |                             |
|                                 | SD for random parameter      |                             |                             |                             |                             |
| Movement of vehicle             | Not going straight           | -2.565 (0.617)**            |                             | -2.435 (0.732)**           |                             |
| Constant                        |                             | 0.364 (0.193)**             |                             |                             |                             |
| Log-likelihood                  |                             | -1.438 (0.453)**            |                             |                             |                             |
| Log-likelihood with constant only|                             | -1995.081                   |                             |                             |                             |
| Freedom                         |                              |                             |                             |                             |                             |
| AIC                             | 48                           |                             |                             |                             |                             |
| BIC                             | 2972.906                     |                             |                             |                             |                             |
| McFadden's R²                   | 0.279                        |                             |                             |                             | 0.477                      |

SD: standard deviation.
* ** *** denote statistically significant at 90%, 95%, 99% respectively.
× A variable with random parameters.
drivers because of their overconfidence in driving skills and risk perception.

**Effects of vehicle factors.** The indicators of the collision with a rigid barrier or other facilities tend to increase the probability of more severe injuries since the parameters were positive. The probability of injury accidents in the collision with a rigid barrier and other facilities increased by 5.1% and 3.0%, respectively. The probability of fatal crashes increased by 6.8% and 5.4%, respectively. Compared with vehicles, the stiffness of the guardrails or other rigid structures was larger, the energy absorbed by the collision was less, so the vehicles sustain more damage by the impact. When the drivers take some emergency actions, the vehicles may turn sharply to increase the possibility of continuous collisions, and the probability of casualties tend to be higher.

The model results also suggested that the type of responsible vehicle was statistically significantly associated with injury severities. Compared with small vehicles, middle vehicles are negatively correlated with the accident injury severity, while large vehicles are positively correlated. The probability of injury and fatal of middle vehicles decreased by 0.5% and 2.5% respectively, and that of large vehicles increased by 14.8% and 16.8% respectively. The coefficients of responsible vehicle type were specified with a random parameter that obey normal distribution, the mean value was 1.44, and the standard deviation was 1.05. From the probability density diagram of large vehicle parameters, it was found that 91.5% of large vehicles have higher probability of injury than small vehicles. It was consistent with past researches.38,39 The possible reason is that large vehicles have large self-weight, high center of gravity, more blind areas of sight, and they have internal wheel difference and long braking distance when turning, which lead to high probability of heavy truck accidents and more serious outcomes. These large vehicles have stronger collision aggression, which can bring greater harm to other vehicles that involved in the collisions.

Compared with the “going straight” state, the indicator of “not going straight” state of vehicles were negatively correlated with the severity of crashes, and the decrease rates of injury and fatal were 4.4% and 3.8%. The possible underlying reason is that the driver in the “not going straight” state is usually in a high alert state. This can be attributed to the risk compensation effects mentioned before.

**Effects of road conditions.** The road surface conditions and terrain type were found to affect severity outcomes. Compared with the dry road surface, the wet-skid pavement was more prone to injury accidents, and the average probability of injury accidents increased by 4.9%. It was consistent with some previous studies where normal road surface condition was found to provoke more severe accidents.39,40 The potential reason is that under the condition of wet-skid pavement, although the

### Table 3: Elasticity analysis of the RPL model for injury severity outcomes.

| Variable                   | Description                          | PDO  | Injury | Fatal  |
|----------------------------|--------------------------------------|------|--------|--------|
| Environmental conditions   | Weather conditions                   | 0.0192 | -0.0177 | 0.0319 |
|                            | Visibility                           | 0.0051 | -0.0066 | -0.0229 |
|                            | Lighting conditions                   | 0.0612 | 0.0871  | 0.0478  |
| Road conditions            | Road surface                         | -0.0376 | 0.0490 | 0.0298 |
|                            | Terrain type                         | -0.0040 | 0.0052  | -0.0510 |
|                            | Vertical alignment                    | -0.0017 | 0.0023  | 0.0044  |
| Driver factors             | Gender                               | -0.1106 | 0.1442  | 0.1567  |
|                            | Age                                  | -0.0082 | -0.0107 | -0.0033 |
|                            | Driving experience                   | 0.0073  | 0.0095  | 0.0129  |
|                            |                                      | 0.0589  | -0.0767 | -0.0987 |
|                            |                                      | 0.0998 | -0.1300 | -0.1771 |
| Vehicle factors            | Collision type                       | -0.0393 | 0.0512  | 0.0681  |
|                            | Responsible vehicle type             | -0.0128 | 0.0297  | 0.0539  |
|                            | Movement of vehicle                  | 0.1134  | -0.1480 | -0.1676 |
|                            |                                      | 0.0338  | -0.0441 | -0.0377 |
| Numeric variable           | Radius of horizontal curve           | -0.0142 | 0.0185  | 0.0292  |
|                            | Length of horizontal curve           | -0.0093 | 0.0121  | 0.0542  |
|                            | Longitudinal gradient                | -0.0036 | 0.0471  | 0.0523  |
Vehicle speed generally decreases, the vehicle braking distance increases significantly, thereby increasing the probability of injury accidents. Interestingly, compared with the flat road, the model results showed that the rolling and/or mountainous sections make the probability of fatal outcomes decreased by 5.1%. This result is reasonable, and the driver is likely to be more cautious in the mountainous sections. From the recorded accident data, the mountain accident accounts for only 2.9% of the total number of accidents. For each 1% increase in radius of horizontal curve, the probability of injury and fatal was expected to increase by 1.9% and 2.9%, respectively. For each 1% increase in the length of horizontal curve, the probability of injury and fatal was expected to increase by 1.2% and 5.4%, respectively. The results also showed that the probability of injury and fatal was expected to increase by 4.7% and 5.2%, respectively, for each 1% increase in vertical slope compared with the flat sections. This is also consistent with previous research results.16,41 As the above cited works pointed out, steeper slope will shorten the line of sight distance, thereby urging the driver to take appropriate action in response to the upcoming traffic accident time.

**Effects of environmental conditions.** When the crash happened at night without light, the probability of injury accidents increased by 4.0% and the probability of fatal accidents increased by 15.1%. The parameters of “night without light” factor in fatal outcomes were random, and the parameter distribution showed that in the case of 82.2%, the probability of fatal outcomes caused by night without light factors was higher than that in the daylight. It can be seen that the indicator of “night without light” significantly increased the probability of fatal outcomes. The potential reason is that the driver only relies on vehicle lights and reflective signs to drive in night without light conditions, which is difficult to meet the requirements of safe driving on expressways. The probability of rear vehicle rear-end collision or side collision increases without timely detection of the presence of front vehicles. In dark conditions, the probability of fatigue driving and inattention increases significantly, and at night it is more likely to appear faster, which aggravates the severity of the accident. In 17.8% of the accidents, the probability of fatal outcomes at night without light was lower than that in daylight, which might be due to the increased driving vigilance of these drivers in the absence of lighting at night and the lower speed of vehicles, which reduces the severity of the accident.

When the visibility was 100–m, the accident rate decreased and casualties decreased. The probability of injury accidents decreased by 8.7%, and the probability of fatal accidents decreased by 4.8%. The possible reason is that when visibility is very low, the vehicle speed decreases significantly. Due to the decrease of vehicle speed, it is not easy to cause serious accidents. The driver is more cautious and the probability of accidents is also reduced. When the visibility was between 100 and 200 m, the probability of injury and fatal accidents decreased by 6.6% and 9.3%, respectively. At the same time, the probability of PDO accidents increased by 5.1%. The potential reason is that when the visibility on the expressway decreases, the driver will generally improve the vigilance and then decelerate, thereby reducing the accident severity. However, on account of the obstruction of environment, the overall accident risk still increases.

**Interaction among influencing factors.** The coefficient and variance-covariance matrix between random parameters were shown in Table 4. The data outside the parenthesis was the random parameter correlation coefficient, and the data inside the parenthesis was the random parameter distribution variance-covariance.

| Variable                                  | Driver gender (Female) | Driving experience (10 + years) | Responsible vehicle type (Large vehicle) | Lighting conditions (Night without light) |
|-------------------------------------------|------------------------|---------------------------------|------------------------------------------|------------------------------------------|
| Driver gender (female)                    | 1.000 (0.175***)       |                                 | 1.000 (0.098***)                         | 1.000 (0.081***)                         |
| Driving experience (10 + years)          | -0.173 (-0.102)        |                                 | 0.318 (0.198*)                          | 1.000 (1.251***)                         |
| Responsible vehicle type (large vehicle) | 0.712 (0.259)          | 0.318 (0.198*)                  | 1.000 (1.251***)                        |                                           |
| Lighting conditions (night without light) | 0.773 (0.097***)       | -0.172 (-0.072***)              | 0.821 (0.534***)                        | 1.000 (0.081***)                        |

* ***denote statistically significant at 90%, 99% respectively.

There was a significant interaction between the indicators of “large vehicle” and driving experience for “10 + years,” driving experience for “10 + years” and “night without light,” “female” drivers and “night without light,” “night without light” and “large vehicle.” That is, the two factors jointly affect the severity outcomes. The correlation coefficient between the indicators of “night without light” and “large vehicle” was 0.821, that is, the impact of large vehicle on accidents was higher than that of night without light or large vehicle alone. Night without light conditions will lead to high incidence of accidents, and the characteristics
of large vehicle are more likely to lead to serious accidents. At the same time, the correlation between the indicators of “night without light” and “large vehicle” was the strongest in the four groups. The correlation coefficient between female drivers and night without light was also reaching 0.773, and there was a significant interaction between the two factors. The risk of accidents of female drivers in night without light conditions was higher than that only one of the two factors in the same conditions. Due to the weak psychological quality of female drivers at night, it may be more difficult for them to make timely and accurate decisions. The correlation coefficient between large vehicle and driving experience for 10+ years was 0.318. From the results of elasticity analysis, it can be concluded that driving experience for 10+ years would reduce the severity of accidents, while large vehicle would increase the severity of accidents. From the correlation coefficient of the two factors, when the two factors appear at the same time, the increased risk of large vehicle offsets the risk of driving experience reducing, which was not conducive to traffic safety improvement. However, the correlation coefficient between night without light and driving experience 10+ years was −0.172, indicating that the combined benefit completely offsets the risk caused by night without light, and the driving environment at this time was conducive to traffic safety.

Safety countermeasures

From the parameter estimation and elastic analysis results of the RPL model, it can be seen that several factors have great impact on injury and fatal accidents, including the indicators of driving experience of 2–10 years, vehicle collided with a rigid barrier, large vehicle as responsible, visibility of 100–m and so on. Therefore, such risk factors should be paid greater/more attention when formulating road traffic safety improvement measures.

Drivers with driving experience of 2–10 years should be treated as novice drivers. They just lack more experience for emergency response in high speed driving process, and are easy to be panic in the face of dangerous conditions, which increases the driving risk. But they are more careful than experienced drivers when driving on expressway. Therefore, it is necessary to conform the safety education and training of drivers, especially for female drivers with 3–10 years’ experience. For older drivers, it is needed to strengthen the examination of their physical health status. And also, enough attention should be paid to the training of psychological quality, the training of emergency response measures, and the improvement of driving standardization.

When the vehicle collides with a rigid barrier, the damage caused by the strong rigidity of the guardrail and other structures to the vehicle is serious, especially at night and when the lighting conditions are not good, which can make the collision frequency increased. Therefore, when there are rigid obstacles such as piers in the middle zone or the lateral clearance range of the expressway, safety protection, or energy absorption facilities should be added to reduce the severity of the accident after the vehicle collision. When the type of the guardrail in the middle zone is selected, the semi-rigid guardrail should be selected as far as possible to meet the requirements of the protection level. The passive safety protection facilities can be added to the accident-prone sections where the concrete guardrail has been set and installation of reflective film to improve visibility, and reasonable arrangement of isolation facilities; lighting conditions should be improved for accident-prone sections without lighting.

Due to the high center of gravity, heavyweight, and less sensitive braking performance of large vehicles, accidents may occur with a high probability of fatal outcomes. Therefore, the appropriate value and mode of speed limit should be set according to the operation time and characteristics of large trucks and striking speed limit warning signs. Auxiliary stability devices should be installed on large vehicles when conditions are available to strengthen the stability of vehicle operation, strengthen the detection of vehicle airbags and the detection of driver fatigue driving, and improve drivers’ safety awareness such as roadside signs, vehicle broadcasting, and increase the wearing rate of drivers’ seatbelt through various transmission channels. The seatbelt can also reduce the severity of accidents to some extent when accidents occur.

When the visibility of the expressway decreases or the road surface is wet and slippery, the road conditions and vehicle spacing ahead will be misjudged. Therefore, it is necessary to strengthen the reflective level of road signs at low visibility, enhance the reflective level of vehicle rear reflective signs, and install dynamic variable speed limit signs and warnings to further limit vehicle speed, improve drivers’ vigilance and reduce the severity of accidents. With regard to vehicle design, efforts can be made to reduce the collision aggression of trucks and other heavy vehicles. The infrastructure for the expressway should be designed to minimize or reduce the use of steep slopes.

Conclusion and limitation

The present study established a RPL model to analyze the influencing factors of expressway traffic accident severity in China. The dependent variable of the model is crash severity that was categorized into three levels: PDO, Injury, Fatal. The results showed that compared with the traditional MNL model, the RPL model can
reveal the effect of various factors on crash severity more reasonably. In addition, since the RPL model can analyze the unobserved heterogeneity and the interaction of various factors on the severity of the accident, it has better fitness and wider application prospect.

The factors affecting the severity of the accident are systematically analyzed from the aspects of environment and road conditions, drivers, and vehicle conditions. The results showed that: (a) When the collision barrier and pier, female drivers, and drivers aged 40 + years are more likely to cause injuries and deaths; (b) When visibility is 200 – m, driver’s driving experience of 2—years or 10 + years, large vehicle responsible and the vehicle “not going straight,” the probability of PDO accidents increases, while the probability of injury and fatal accidents decreases; (c) The probability of injury accidents increases under the condition of wet-skid road surface, while the concrete guardrail and “night without light” conditions are more likely to lead to fatal outcomes; (d) When large vehicles run on the road section without lighting facilities at night, if there is a traffic accident, the severity of the accident is generally more serious.

There are some limitations in this study. We only analyzed the factors that can be collected, but there were still many factors not well collected or that have a significant impact on the accident but were unknown, such as instantaneous speed, real-time weather, road design elements and whether the seatbelts were used. Different sample sizes will affect the parameter estimation of the model. This study does not group the samples in depth, and statistics in the sample sizes applicable to different models may also lead to a certain deviation in the calibration results. In the follow-up study, the influence of different sample sizes on parameter estimation can be explored in depth, and sample grouping can be carried out to obtain better fitting results. In addition to the RPL model constructed in this paper, there are also other models to analyze the accident severity, such as the nested logit model and the ordered logit model. Comprehensive comparative analysis of the practicability of each model is also the direction of subsequent research.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was jointly supported by the National Key R&D Program of China (2019YFE0108000), the National Social Science Foundation of China (19CKGL004), and the National Natural Science Foundation of China (51178054).

ORCID iD
Fei Ye https://orcid.org/0000-0003-1226-2289

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

References
1. Mannering FL and Bhat CR. Analytic methods in accident research: methodological frontier and future directions. Anal Methods Accid Res 2014; 1: 1–22.
2. Zeng Q, Wen H, Huang H, et al. Incorporating temporal correlation into a multivariate random parameters tobit model for modeling crash rate by injury severity. Transportmetrica 2018; 14: 177–191.
3. Zhao J and Deng W. Traffic accidents on expressways: new threat to China. Traffic Inj Prev 2012; 13: 230–238.
4. Mergia WY, Eustace D, Chimba D, et al. Exploring factors contributing to injury severity at freeway merging and diverging locations in Ohio. Accid Anal Prev 2013; 55: 202–210.
5. Ratanaavaraha V and Suangka S. Impacts of accident severity factors and loss values of crashes on expressways in Thailand. IATSS Res 2014; 37: 130–136.
6. Chen Z and Fan W. A multinomial logit model of pedestrian-vehicle crash severity in North Carolina. Int J Transp Sci Technol 2019; 8: 43–52.
7. Al-Ghamdi AS. Using logistic regression to estimate the influence of accident factors on accident severity. Accid Anal Prev 2002; 34: 729–741.
8. Savolainen PT, Mannering FL, Lord D, et al. The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. Accid Anal Prev 2011; 43: 1666–1676.
9. Çelik AK and Oktay E. A multinomial logit analysis of risk factors influencing road traffic injury severities in the Erzurum and Kars provinces of Turkey. Accid Anal Prev 2014; 72: 66–77.
10. Ye X, Pendyala RM, Shankar V, et al. A simultaneous equations model of crash frequency by severity level for freeway sections. Accid Anal Prev 2013; 57: 140–149.
11. Chen F, Chen S and Ma X. Analysis of hourly crash likelihood using unbalanced panel data mixed logit model and real-time driving environmental big data. J Saf Res 2018; 65: 153–159.
12. Ma C, Hao W, Xiang W, et al. The impact of aggressive driving behavior on driver-injury severity at highway-rail grade crossings accidents. J Adv Transport 2018; 2018: 1–10.
13. Milton JC, Shankar VN and Mannering FL. Highway accident severities and the mixed logit model: an exploratory empirical analysis. Accid Anal Prev 2008; 40: 260–266.
14. Malysheksy NV and Mannering FL. Markov switching multinomial logit model: an application to accident severity. Accid Anal Prev 2009; 41: 829–838.
15. Kim J-K, Ulfarsson GF, Kim S, et al. Driver-injury severity in single-vehicle crashes in California: a mixed logit analysis of heterogeneity due to age and gender. Accid Anal Prev 2013; 50: 1073–1081.
16. Yu R and Abdel-Aty M. Using hierarchical Bayesian binary probit models to analyze crash injury severity on high speed facilities with real-time traffic data. Accid Anal Prev 2014; 62: 161–167.
17. Rra D, Mmha B, Apa C, et al. Applying a random para-meter negative binomial lindley model to examine multi-vehicle crashes along rural mountainous highways in Malaysia. Accid Anal Prev 2018; 119: 80–90.
18. Rezapoour M, Moomen M and 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.
19. Osman M, Mishra S and Paleti R. Injury severity analysis of commercially-licensed drivers in single-vehicle crashes: accounting for unobserved heterogeneity and age group differences. Accid Anal Prev 2018; 118: 289–300.
20. Sasidharan L, Wu KF and Menendez M. Exploring the application of latent class cluster analysis for investigating pedestrian crash injury severities in Switzerland. Accid Anal Prev 2015; 85: 219–228.
21. Adanu EK, Hainen A and Jones S. Latent class analysis of factors that influence weekday and weekend single-vehicle crash severities. Accid Anal Prev 2018; 113: 187–192.
22. Mannering FL, Shankar V and Bhat CR. Unobserved heterogeneity and the statistical analysis of highway accident data. Anal Methods Accid Res 2016; 11: 1–16.
23. Flask T, Schneider WH and Lord D. A segment level analysis of multi-vehicle motorcycle crashes in Ohio using Bayesian multi-level mixed effects models. Saf Sci 2014; 66: 47–53.
24. Haleem K and Gan A. Effect of driver’s age and side of impact on crash severity along urban freeways: a mixed logit approach. J Saf Res 2013; 46: 67–76.
25. Chang LY and Mannering F. Analysis of vehicle occupancy and the severity of truck- and non-truck-involved accidents. Accid Anal Prev 1999; 31: 579–592.
26. Lee J and Mannering F. Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accid Anal Prev 2002; 34: 149–161.
27. Angiulli F and Fassetti F. Distance-based outlier queries in data streams: the novel task and algorithms. Data Min Knowl Discov 2010; 20: 290–324.
28. Potoglou D, Carlucci F, Cirà A, et al. Factors associated with urban non-fatal road-accident severity. Int J Inj Contr Saf Promot 2018; 25: 303–310.
29. Zeng Q, Gu W, Zhang X, et al. Analyzing freeway crash severity using a Bayesian spatial generalized ordered logit model with conditional autoregressive priors. Accid Anal Prev 2019; 127: 87–95.
30. Wu Q, Chen F, Zhang G, et al. 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.
31. Morgan A and Mannering FL. The effects of road-surface conditions, age, and gender on driver-injury severities. Accid Anal Prev 2011; 43: 1852–1863.
32. Khattak AJ. Injury severity in multivehicle rear-end crashes. Transp Res Rec 2001; 1746: 59–68.
33. Agebie BRDK. The effect of gender on two-passenger vehicle highway crash-injury severity: a mixed logit empirical analysis. J Transp Saf Secur 2016; 8: 280–291.
34. Koornstra MJ. Risk-adaptation theory. Transp Res Part F Traffic Psychol Behav 2009; 12: 77–90.
35. Hatfield J and Fernandes R. The role of risk-propensity in the risky driving of younger drivers. Accid Anal Prev 2009; 41: 25–35.
36. Moták L, Gabaude C, Bougeant J-C, et al. Comparison of driving avoidance and self-regulatory patterns in younger and older drivers. Transp Res Part F Traffic Psychol Behav 2014; 26: 18–27.
37. Doroudgar S, Chuang HM, Perry PJ, et al. Driving performance comparing older versus younger drivers. Traffic Inj Prev 2017; 18: 41–46.
38. Shao X, Ma X, Chen F, et al. A random parameters ordered probit analysis of injury severity in truck involved rear-end collisions. Int J Environ Res Public Health 2020; 17: 395.
39. Yamamoto T and Shankar VN. Bivariate ordered-response probit model of driver’s and passenger’s injury severities in collisions with fixed objects. Accid Anal Prev 2004; 36: 869–876.
40. Quddus MA, Noland RB and Chin HC. An analysis of motorcycle injury and vehicle damage severity using ordered probit models. J Saf Res 2002; 33: 445–462.
41. Christoforou Z, Cohen S and Karlaftis MG. Vehicle occupant injury severity on highways: an empirical investigation. Accid Anal Prev 2010; 42: 1606–1620.