Accident Analysis and Prevention 124 (2019) 219–229

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Investigation of driver injury severities in rural single-vehicle crashes under rain conditions using mixed logit and latent class models

Zhenning Li[a], Yusheng Ci[b], Cong Chen[c], Guohui Zhang[a,*], Qiong Wu[a], Zhen (Sean) Qian[d], Panos D. Prevedouros[a], David T. Ma[a]

[a] Department of Civil and Environmental Engineering, University of Hawaii at Manoa, 2500 Campus Road, Honolulu, HI, 96822, United States
[b] Department of Transportation Science and Engineering, Harbin Institute of Technology, 73 Huanghe Road, Harbin, Heilongjiang, 150090, China
[c] Center for Urban Transportation Research, University of South Florida, 4202 East Fowler Avenue, CUT100, Tampa, FL, 33620, United States
[d] Civil and Environmental Engineering, Carnegie Mellon University Pittsburgh, PA, 15213-3890, United States

ARTICLE INFO

Keywords:
Driver injury severity
Mixed logit model
Latent class model
Unobserved heterogeneity
Temporal instability

ABSTRACT

Due to limited visibility and low skid resistance on road surface, single-vehicle crashes under rain conditions, especially those occurred in rural areas, are more likely to result in driver incapacitating injuries and fatalities. A three-year crash dataset including all rural single-vehicle crashes under rain conditions from 2012 to 2014 in four South Central states, i.e., Texas, Arkansas, Oklahoma, and Louisiana, are selected in this paper to analyze the impact factors on driver injury severity. The mixed logit model (MLM) and the latent class model (LCM) are developed on the same dataset. Several parsimony indices, e.g., AIC and BIC, and as well as McFadden pseudo r-squared, are calculated for all the models to evaluate their respective performance. Results show that choosing the uniform distribution as the prior for random parameters could better improve the goodness-of-fit of the MLM than using normal and lognormal distributions. In addition, the two-class LCM also shows superiority when compared to three- and four-class LCMs. Finally, a careful comparison between these two models is conducted, and the results indicate that the LCM has a slightly better performance in analyzing the aforementioned dataset in this study. Model estimation results show that *curve*, *on grade*, *signal control*, *multiple lanes*, *pickup*, *straight*, *drug/alcohol impaired*, and *seat belt not used* can significantly increase the probability of incapacitating injuries and fatalities for drivers in the two models. On the other hand, *wet*, *male*, *semi-trailer*, and *young* can significantly decrease the probability of incapacitating injuries and fatalities for drivers. This study provides an insightful understanding of the effects of these attributes on rural single-vehicle crashes under rain conditions and beneficial references for developing effective countermeasures for severe injury prevention.

1. Introduction

Comparing with urban areas, rural areas are more concentrated with fatal traffic crashes and fatalities. NHTSA revealed that rural areas held only 19% of the total U.S. population but induced 48% of fatal traffic crashes and 49% of traffic fatalities in 2015 (NHTSA, 2017). Therefore, rural traffic crashes have drawn worldwide research interest (Islam and Brown, 2017; Rusli et al., 2017; Wang et al., 2017). Similarly, single-vehicle crashes are also found to be more fatality-concentrate than multi-vehicle crashes, as was evidenced by the fact that single-vehicle crashes accounted for 28.9% of all crashes, but 58.1% of all fatal crashes in the U.S. in 2015 (NHTSA, 2017). Given these two fatality-concentrate features of rural crashes and single-vehicle crashes, it is of particular interest to investigate the injury severity patterns in rural single-vehicle crashes, a sub crash type that belongs to both these two fatality-concentrate crash types. Injury severity in traffic crashes is an indispensable research area in crash data analysis, in addition to crash frequency analysis, and driver injury severity has been widely used as a representative indicator at the individual level (Behnood and Mannering, 2017a; 2017b; Chen et al., 2016b, 2015a; Seraneeprakarn et al., 2017). Numerous studies have already been conducted to investigate the contributing factors and propose effective countermeasures to mitigate driver injury severity in traffic crashes (Chang and Chien, 2013; Chen et al., 2016c, 2015a; Kim et al., 2013; Wu et al., 2014).

Previous studies showed that driving in the rain may be associated

* Corresponding author.
E-mail addresses: li2016@hawaii.edu (Z. Li), ciyusheng1999@126.com (Y. Ci), congchen1@cutr.usf.edu (C. Chen), guohui@hawaii.edu (G. Zhang), wuqiong@hawaii.edu (Q. Wu), seanqian@cmu.edu (Z.S. Qian), pdp@hawaii.edu (P.D. Prevedouros), tianwei@hawaii.edu (D.T. Ma).

https://doi.org/10.1016/j.aap.2018.12.020
Received 25 August 2018; Received in revised form 7 December 2018; Accepted 22 December 2018
Available online 23 January 2019
0001-4575/ © 2019 Elsevier Ltd. All rights reserved.

with higher crash risk than that in clear weather (Jung et al., 2010). A sizable portion of severe traffic crashes is brought about by these issues and induces significant fatalities and serious injuries. According to the Texas Department of Transportation (TxDOT, 2016), 16,818 rural crashes (159 fatal crashes) occurred under rain conditions in 2015, which is four times as many as those related to all other inclement weather conditions (e.g., blowing sand, sleet, and hail). In addition, crash statistics from Arkansas and Oklahoma (Arkansas Department of Transportation, 2015; OKDOT, 2015) showed that single-vehicle crashes under rain conditions, especially those occurred in rural areas, have a probability of drivers being seriously injured approximately twice as high as that for multi-vehicle crashes occurred under the same or similar conditions. However, in most traffic safety studies, weather condition has been considered as a contributing factor in crash cause-effect analysis, and only a limited number of studies directly focused on crashes under rain conditions. Andrey and Yagar (1993) analyzed the crash risk during and after rain events in urban areas, and discovered that the overall crash risk under rain conditions is 70% higher than that in average-day clear conditions. Jung et al. (2010) developed two types of polychotomous response models to analyze rain-related crashes in Wisconsin and concluded that rain-related factors could significantly affect injury severity. However, the safety impacts of rain and other variables in rain-related crashes are found unstable among different studies. For instance, a study examining the temporal and spatial distribution of rain-related crashes in Texas suggested that rain is a contributor to fatal crashes only in few dry counties, but has no impacts on crashes in some of the wetter counties (Jackson and Sharif, 2014). Qiu and Nixon (2008) reported that rain is associated with higher injury severity and crash rates. Feng et al. (2016) concluded that severe accidents are about twice more likely to occur on curved roadways on rainy days, although straight and curved roadways have similar impacts in clear days. Shaheed et al. (2016) also reported that gender, seating position, road junction type, and other risk factors have different effects on injury severity in weather-related (rain, snow, blowing sand, etc.) and non-weather-related crashes. Whereas in the article of Lee et al. (2015), estimation results showed that injury severity is relatively lower under rain condition in all crash types since drivers tend to reduce their speeds and be more careful on a wet surface. The sophisticated influences of rain on overall traffic safety indicate that there is a need for detailed analyses regarding external weather conditions and collision types.

In contemporary traffic safety research, unobserved heterogeneity of the police-reported crash dataset has been recognized as a critical issue (Mannering et al., 2016; Mannering and Bhat, 2014). In this study, the dataset was obtained from a representative sample of police-reported motor vehicle crashes with discrete injury severity outcomes, where there are still certain elements related to crashes not recorded and remaining unobserved to researchers, even though many elements (e.g., driver age, gender, number of lines of a roadway, etc.) have been covered. Traditional models that are usually used in traffic crash data analysis, e.g., multinomial logit model (MNL), ordered logit model, etc., cannot adequately address the unobserved heterogeneity within such dataset. Therefore, models that can account for unobserved heterogeneity should be developed. One of the first practices on unobserved heterogeneity model was a study by Kim et al. (2008), where they developed a mixed logit model (MLM) to investigate the effect of driver age on driver injury severity outcome in single-vehicle crashes. A further study on driver injury severity also conducted by Kim et al. (2013) demonstrated that MLMs are superior to traditional discrete choice models in that they are more flexible and can approximate any random utility model. Wu et al. (2014) developed MLMs to analyze driver injury severities in single-vehicle crashes and compared the results with multi-vehicle crashes. Noticeably, elasticity analyses and transferability tests were applied to discuss the models' parameters estimation outcomes, and the results showed that elasticity analysis is a necessary supplement to MLMs. The MLM can account for individual unobserved heterogeneity by allowing parameters to vary across observations and therefore yield more reliable estimations (Kim et al., 2013; Milton et al., 2008; Moore et al., 2011). It should be mentioned that some recent studies tried to explicitly examine the possible heterogeneity in means and/or variance. For instance, Behnood and Mannering, (2017a; 2017b) adopted an MLM with heterogeneity in parameter means to explore the differences in driver-injury severities. Seraneeprakarn et al. (2017) developed an MLM of injury severity while allowing for heterogeneity in parameter means and variances. Models with no mean-variance related heterogeneity, and with mean related heterogeneity only, are also developed and compared with the proposed model. The estimation results showed that for their dataset, the proposed model has better performance over the other two, and some variables were found to randomly distributed with significant heterogeneity in both means and variances. Interested readers are referred to these papers and the references cited therein. However, this model also has its own drawbacks. Due to its flexible structure, the MLM requires appropriate distribution assumptions for potential random parameters; otherwise, these random effects may remain undetected. This restriction is released in another widely used method, the latent class model (LCM), where specific distributions for parameters of interest are not required. Instead, determining a proper number of classes becomes a critical step when using this approach. The unobserved heterogeneity is then identified by these different classes with homogeneous characteristics of the within-class observations (Gelman and Hill, 2007; Ma et al., 2016; Mannering et al., 2016). Xie et al. (2012) developed the LCM to deal with the single-vehicle crashes, and concluded that the LCM has the potential to overcome the problems associated with the irrelevant alternatives (IIA) property that commonly exists in multinomial logit models (Abdel-Aty, 2003). Both MLMs and LCMs were developed on a pedestrian-injury dataset to ensure reliable estimation, and the results showed that both models are appropriate to capture unobserved heterogeneity (Behnood and Mannering, 2016). However, there are only a few references have directly compared the MLM and the LCM for driver injury severity analysis, and the comparison is not always comprehensive. For instance, Cerwick et al. (2014) compared the two models by their model fit, inferences, and predicted crash severity outcome probabilities by a large sample of crash data on multiple vehicle crashes. Behnood and Mannering (2016) developed both the MLM and the LCM to study the risk factors on the pedestrian crash dataset from Chicago city. However, most of the previous studies did not explicitly conclude which model is superior to the other.

Another issue lies in the traffic safety analysis is that analysts always aggregated the crash data over a specified time period to gather sufficient observations for analyzing. However, some recent research suggests that the impact of factors affecting injury severity may not be temporal stable (Behnood and Mannering, 2015; Mannering, 2018). In our dataset, the proportion of different types of crashes each year is not constant. The occurrence of rain-related crashes may have relationships with the weather or even climate change, both of which are virtually impossible to measure with existing data sources. In addition, another potential problem is the fact that driver involved in rain-related crashes may be a non-random sample since safer drivers may choose to take other modes of travel due to compromised road friction and visibility. Ignoring possible temporal effects may adversely affect the inferences drawn from model estimations as well as their ability to be used to forecast and evaluate the effects of safety countermeasures (Mannering, 2018). However, both MLM and LCM are not able to explicitly distinguish that the unobserved heterogeneity revealed by these models are entirely due to temporal instability, or a combination of temporal shifts and other traditional sources of unobserved heterogeneity. An article by Behnood and Mannering (2016) provided some insightful technique details to determine whether there exists temporary instability or not in the estimates of unobserved heterogeneity models. A series of likelihood ratio tests were conducted to compare models developed for two time periods and examine if the parameter estimates are stable between these periods. This technique is also adopted in this study for the aim of temporal stability testing and model comparison of the MLM and the LCM.

The objective of this paper is to investigate the differences between the MLM and the LCM for exploring the relationships between driver injury severity in the rain-related rural single-vehicle crash and its corresponding risk factors. The comparison of the two models lies in model performance, inferences, estimated outcome probabilities, and likelihood ratio test results. The rest of this paper is organized as follows: Section 2 provides the details of the studied crash dataset. The two proposed models are explicitly introduced in Section 3. The intuitive implications of the models, detailed discussions of the estimation results, pseudo-elasticity effects of the best-fit model, and model comparison results are presented in Section 4. Finally, this research is concluded in Section 5.

2. Data

A three-year crash dataset including all rural single-vehicle crashes under rain conditions in four South Central states in 2012 and 2014 is utilized in this research. This dataset is obtained from Texas Department of Transportation (TxDOT), Arkansas State Highway and Transportation Department (AHTD), Oklahoma Department of Transportation (OKDOT), and Louisiana Department of Transportation and Development (LADOTD). The geographical locations of these four states are illustrated in Fig. 1. These states are concentrated in the south-central United States and have similar climatic characteristics, such as precipitation characteristics (Carter et al., 1974), annual temperature variations (Aguilar et al., 2005), etc. In addition, the similar demographic features of these states indicate that they can be studied as a whole, as evidenced by numerous peer studies in different fields (Adams et al., 2016; Miller et al., 2013; Munn et al., 2002).

In view of the differences in the process of recording crash reports in the four states, only the identical variables from the sub-dataset of each of the four states were selected in the study. After careful examination, the incomplete and erroneous records in the original data set were deleted. Finally, 17,929 accurate records were retained in the study for modeling analysis. The final dataset contains detailed information on driver injury severity and potential contributing factors regarding the characteristics of the crash, vehicle, and driver, such as road geometry, vehicle type, driver demographics, etc. The driver injury severity was classified into five subtypes in the original files, including fatal injury, incapacitating injury, visible injury, complaint of injury, and no apparent injury, respectively. In this study, fatal injury and incapacitating injury are combined together as the most severe injury severity level to maintain a statistically meaningful sample size. The classifications of other injury severities are consistent with the original files. Therefore, driver injury severity is categorized into four subtypes in this study, including O (original category: no apparent injury), C (original category: complaint of injury), B (original category: visible injury), and AK (original category: fatal injury and incapacitating injury), respectively. Some continuous integer variables (including driver age, number of vehicles in the crash, etc.) are categorized as discrete variables with a finite number of exclusive values, as a constant coefficient may not fully reveal the various impacts of a continuous variable on driver injury severity when its numerical value falls into different ranges (Chen et al., 2016b). Moreover, based on our previous studies and engineering experience (Chen et al., 2015a, 2016c, 2015b), some multi-categorical variables with an excessive amount of original values are simplified to improve modeling efficiency. For example, right turn and making a right turn on red are combined together as a variable, right turn, to reduce the number of categorical values in a variable. Furthermore, variables with relatively similar impacts on driver injury severity but not having enough records of presence, such as alcohol-impaired and drug-impaired, are combined as an integrated factor for modeling simplification purpose. The descriptive statistics of the studied dataset are shown in Table 1.

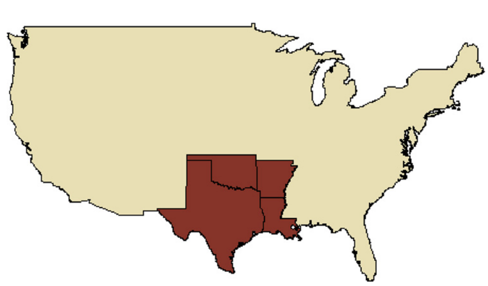

Fig. 1. Location map of study area.

3. Methodology

MLMs and LCMs are widely used to capture unobserved heterogeneity (Behnood et al., 2014; Cerwick et al., 2014; Sasidharan et al., 2015; Ye and Lord, 2014). However, both of these two models have their own drawbacks. For the MLM, there has been an ongoing debate on which distribution assumption can best capture unobserved heterogeneity, and an optimal MLM requires predefined prior distribution specific to each possible random parameter with which model parameter estimations are allowed to be random and vary across observations (Luo et al., 2008). Several continuous distributions, including normal distribution, uniform distribution, and lognormal distribution, have been widely used in previous studies for capturing random effects of parameters (Greene, 2012; Hensher and Greene, 2003). However, there are no clear conclusions of which distributions can provide the best model performance. Thus, careful selection of the predefined distributions is of practical importance in the modeling process. On the other hand, a disadvantage of the LCM is that it is difficult to determine the optimal number of subtypes, and that the unobserved heterogeneity, although reduced, might still exist within each identified latent class. In addition, previous applications also suggested that after specifying more than four subtypes, it becomes extremely difficult to achieve model convergence and obtain accurate parameter estimations (Greene, 2012). However, there is no panacea for this problem, as that the optimal number of classes is closely related to the data structure, the nature of the variables, and other characters of the dataset. In order to bridge this gap, therefore, the aforementioned issues will be carefully addressed in the following sections.

3.1. Mixed logit model

Assuming that driver injury severities are classified into K levels (in this study $K = 4$), and given the fact that the studied dataset is regarding single-vehicle crashes, the function determining the driver injury severity level k ($k \in K$) for the nth driver, Y_{kn}, is given by

$$Y_{kn} = \boldsymbol{\beta}_k \boldsymbol{X}_{kn} + \varepsilon_{kn} \tag{1}$$

where $\boldsymbol{\beta}_k$ is a vector of parameters to be estimated for driver injury severity level k which may vary across observations, $\boldsymbol{X}_{kn}$ is a vector of explanatory variables (light conditions, traffic controls, driver ages, etc.), and the disturbance term is notated as ε_{kn}, which is assumed to be generalized extreme value distributed (McFadden, 1981). Consequently, the standard multinomial logit model (neglecting for the error components) can be expressed as

$$P_n(k) = \frac{e^{\boldsymbol{\beta}_k \boldsymbol{X}_{kn}}}{\sum_{\forall k \in K} e^{\boldsymbol{\beta}_k \boldsymbol{X}_{kn}}} \tag{2}$$

where $P_n(k)$ is the probability of the nth driver having kth severity level. Supposing the random parameters that capture unobserved heterogeneity on driver injury severity outcomes are given by $f(\boldsymbol{\beta}_k|\boldsymbol{\varphi})$, where

Table 1
Variable definition and description.

Variable	Driver injury severity								Total
	O (Mean)		C (Mean)		B (Mean)		AK (Mean)		
Severity	10850	0.61	3855	0.22	2387	0.13	837	0.05	17929
Light Condition									
Dark	3909	0.61	1308	0.2	896	0.14	337	0.05	6450
Dawn	151	0.64	24	0.1	41	0.17	20	0.08	236
Daylight	6790	0.6	2523	0.23	1450	0.13	462	0.04	10982
Road Character									
No curve	8734	0.61	3302	0.23	1747	0.12	564	0.04	14347
Curve	2116	0.59	553	0.15	640	0.18	273	0.08	3582
Road Grade									
Level	9339	0.61	3345	0.22	2059	0.13	662	0.04	15405
Hillcrest	255	0.63	74	0.18	54	0.13	20	0.05	403
On grade	1256	0.59	436	0.21	274	0.13	155	0.07	2121
Road Surface Condition									
Dry	143	0.52	62	0.23	40	0.15	28	0.1	273
Wet	10629	0.61	3776	0.21	2288	0.13	786	0.05	17258
Snow	78	0.68	17	0.15	14	0.12	5	0.04	114
Traffic Control									
No Control	1489	0.62	540	0.22	334	0.14	44	0.02	2407
Stop-Yield Sign	53	0.65	12	0.15	13	0.16	3	0.04	81
Signal Control	9308	0.6	3303	0.16	2040	0.17	790	0.07	8666
Number of Lanes									
One Lane	589	0.64	165	0.18	135	0.15	29	0.03	918
Two Lanes	8738	0.59	3383	0.23	1923	0.13	670	0.05	14714
Multiple Lanes	1523	0.66	307	0.13	329	0.14	138	0.06	2297
Vehicle Type									
Passenger Car	9481	0.61	3501	0.23	1998	0.12	692	0.04	13715
Pick-up	1214	0.59	342	0.17	374	0.18	141	0.07	2071
Semi	117	0.79	12	0.08	15	0.1	4	0.03	148
Bus	38	1	0	0	0	0	0	0	38
Action									
Straight	7990	0.61	3052	0.23	1443	0.11	560	0.04	13044
Right Turn	1319	0.57	364	0.16	455	0.2	164	0.07	2302
Left Turn	1450	0.59	434	0.18	479	0.19	113	0.05	2476
U-Turn	7	1	0	0	0	0	0	0	7
Slowing	70	0.82	5	0.06	10	0.12	0	0	85
Backing	14	1	0	0	0	0	0	0	14
Seat Belt used									
Used	10288	0.61	3659	0.22	2162	0.13	694	0.04	16803
Not used	562	0.5	196	0.17	225	0.2	143	0.13	1126
Drug/Alcohol Impaired	257	0.39	142	0.21	178	0.27	86	0.13	663
Gender									
Male	6323	0.63	1937	0.19	1289	0.13	487	0.05	10036
Female	4527	0.57	7893	1	1098	0.14	350	0.04	7893
Age									
Young (< 25 years)	4468	0.61	1544	0.21	1033	0.14	259	0.04	7304
Middle (25˜64 years)	5938	0.6	2141	0.22	1246	0.13	530	0.05	9855
Old (> 64 years)	444	0.58	170	0.22	108	0.14	48	0.06	770

$\boldsymbol{\varphi}$ is a vector that representing the probability density function (PDF). According to previous studies (McFadden and Train, 2000), in the MLM, the resulting outcome probabilities $P_n(k|\boldsymbol{\varphi})$ are given by

$$P_n(k|\boldsymbol{\varphi}) = \int \frac{e^{\boldsymbol{\beta_k X_{kn}}}}{\sum_{\forall k \in K} e^{\boldsymbol{\beta_k X_{kn}}}} f(\boldsymbol{\beta}_k|\boldsymbol{\varphi}) d\boldsymbol{\beta}_k \tag{3}$$

Consequently, the individual specific variations of the impacts of the corresponding variable vector, $\boldsymbol{X}_{kn}$, is accounted by the parameter vector, $\boldsymbol{\beta}_k$. Different with the standard multinomial logit model, $\boldsymbol{\beta}_k$ is not a fixed parameter and may be randomly distributed with various mode and skewness. Normal distribution is the most familiar, simplest assumption of the distribution of $\boldsymbol{\beta}_k$, and is specified as,

$$\boldsymbol{\beta}_k = \boldsymbol{\beta}_i + \sigma_i \nu_i, \ \nu_i \sim N(0, 1) \tag{4}$$

where $\boldsymbol{\beta_i}$ is the mean, σ_i is the standard deviation of the distribution, and ν_i is the individual-specific heterogeneity, with mean equal to zero and standard deviation equal to one (Greene, 2012; Li et al., 2018b).

In this study, not only the normal distribution but also other types of distribution are chosen to limit the parameter values to a specific range based on engineering experience. For example, as shown in Eq. (5), the lognormal distribution is used to maintain the values of certain parameters (e.g., drug-impaired, seatbelt used, etc.) to be either positive or negative

$$\boldsymbol{\beta}_k = e^{\boldsymbol{\beta}_i + \sigma_i \nu_i}, \ \nu_i \sim N(0, 1) \tag{5}$$

In addition, to ensure a reasonable range of the variation of a parameter, uniform distribution is selected on certain parameters to replace the normal distribution that assumes the variation range is infinite, and is given by

$$\boldsymbol{\beta}_k = \boldsymbol{\beta_i} + \sigma_i \nu_i, \ \nu_i \sim unif(-1, 1) \tag{6}$$

Normal distribution, lognormal distribution, and uniform distribution are separately assumed for each potential random parameter. The model is finalized based on the characters of these parameters as well as some parsimony indices, including Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). AIC and BIC are defined as follows, respectively,

$$AIC = -2\ln(L) + 2p \tag{7}$$

$$BIC = -2\ln(L) + p \times ln(N) \tag{8}$$

where $\ln(L)$ is the model likelihood, p is the number of estimated model parameters in the model, and N is the total number of observations. Lower AIC or BIC value of a candidate model generally indicates the model has better fit and is considered to be closer to the “true” model.

McFadden Pseudo R-squared (McFadden and Zarembka, 1974) statistic is also applied to demonstrate the model fitness. The formula is given as

$$R^2 = 1 - \frac{ln\hat{L}(M_{Full})}{ln\hat{L}(M_{Constant})} \tag{9}$$

where $\hat{L}$ is the estimated likelihood, M_{Full} is the full model with the constant term and all predicting variables, $M_{Constant}$ is the intercept model only including the constant term. The log likelihood of the intercept model is treated as a total sum of squares (TSS), and the log likelihood of the full model is treated as the sum of squared errors of prediction (SSE). The ratio of the likelihoods suggests the level of improvement over the intercept model offered by the full model. According to previous research (Domencich and McFadden, 1975), the McFadden Pseudo R-squared value in this study indicates that the full model is a significantly better fit compared with the intercept model.

3.2. Latent class model

The LCM has some similarities with the MLM, but embodies several critical differences as well. This model is generally used to discover potential subtypes or confirm hypothesized subtypes based on multivariate categorical data (Lazarsfeld et al., 1968). The underlying theory of the LCM posits that individual behavior depends on observable attributes and on latent heterogeneity that varies with factors that are unobserved by the analyst. In an LCM framework, a discrete distribution (e.g. multinomial distribution) is selected to model unobserved heterogeneity across observations. The latent classes that are not revealed to analyst could be treated as different bins where an individual resides in based on its own characteristics.

Assuming there are C distinct latent classes in the model, the probability of the nth crash record belonging to class c ($c \in C$) specified by an MNL model is defined as:

$$P_n(c) = \frac{e^{\theta_c \gamma_n}}{\sum_{\forall C} e^{\theta_c \gamma_n}} \tag{10}$$

where γ_n is a vector of characteristics that determine class c probabilities for n th crash, θ_c is the corresponding vector of estimable parameters. In addition, the conditional probability of in the nth driver in class c having level k injury severity is given by

$$P_n(k|c) = \frac{e^{\beta_{kc} \boldsymbol{X}_{knc}}}{\sum_{\forall k \in K} e^{\beta_{kc} \boldsymbol{X}_{knc}}} \tag{11}$$

Finally, the unconditional probability of driver in the nth crash having level k injury severity is defined as:

$$P_n(k) = \sum_{\forall C} P_n(c) \times P_n(k|c) \tag{12}$$

As mentioned above, significant research effort has been made in search of an optimal number of classes when an LCM is developed. In this study, a statistical accrual searching process is utilized to find the optimal number of latent classes, starting with 2 latent classes and increasing by 1 in each step up to the maximum plausible number of latent classes, and AIC and BIC are selected to assess model fitness. Estimation of the variables in this study is conducted with an iterative numerical method, the maximum likelihood estimation (MLE) algorithm. The estimated asymptotic covariance matrix is based on the second derivatives of the specific utility functions. The Berndt–Hall–Hall–Hausman (BHHH) estimator is used in case that the matrix fails to be positive because of rounding error (Berndt et al., 1974).

3.3. Pseudo-elasticity analysis

Extensive studies have proved that the signs of parameter estimation results are not always consistent with the real impacts of these parameters when a multinomial response variable is applied in model design (Kim et al., 2013; Osman et al., 2016; Wu et al., 2014). Therefore, an elasticity analysis is necessary to assess the influences of statistically significant variables in each of the proposed models, given the multi-level driver injury severity outcome. The elasticity is calculated in the form of the partial derivative for each observation (Washington et al., 2011),

$$E_{X_{kni}}^{P_{kn}} = \frac{\partial P_{kn}}{\partial X_{kni}} \frac{X_{kni}}{P_{kn}} \tag{13}$$

where $E_{X_{kni}}^{P_{kn}}$ is the elasticity outcome for the driver of the n th crash with severity level k, X_{kni} is the value of the ith variable for the nth crash in the propensity function with respect to the kth injury severity level. However, Eq. (12) is not applicable for this study since the variables have already been transformed into binary forms (with 0/1 outcome), and the probabilities are not differentiable with respect to indicator variables. To address this issue, direct pseudo-elasticity is defined in Eq. (13) by modifying Eq. (12) to calculate the influence of each significant indicator variable (Kim et al., 2007),

$$E_{(p)X_{kni}}^{P_{kn}} = \frac{P_{kn}[given\ X_{kni} = 1] - P_{kn}[given\ X_{kni} = 0]}{P_{kn}[given\ X_{kni} = 0]} \tag{14}$$

where $E_{(p)X_{kni}}^{P_{kn}}$ is the pseudo-elasticity of the probability and is defined as the percentage change in probability when an indicator variable is switched (i.e., from 0 to 1 or from 1 to 0); P_{kn} is the probability the driver of nth crash having an injury severity level k for the given value of the variable X_{kni} while holding other variables constant. The direct pseudo-elasticity in Eq. (14), $E_{(p)X_{kni}}^{P_{kn}}$, is calculated for each record in the dataset, and the average pseudo-elasticity is calculated based on all data records to measure variable influence.

3.4. Temporal stability test

In order to test and compare the temporal stability of MLM and LCM, a series of likelihood ratio tests are conducted. These tests are used to compare models developed for two different years and examine if the parameter estimates are stable between the two years. The test statistic follows a χ^2 distribution with degrees of freedom equal to the number of estimated parameters and can be written as (Washington et al., 2011)

$$\chi^2 = -2[LL(\boldsymbol{\beta}_{t_2 t_1}) - LL(\boldsymbol{\beta}_{t_1})] \tag{15}$$

where $LL(\boldsymbol{\beta}_{t_2 t_1})$ is the log-likelihood at convergence of a model containing converged parameters based on using year $t_2's$ data, while using data from year t_1, and $LL(\boldsymbol{\beta}_{t_1})$ is the log-likelihood at the convergence of the model using year $t_1's$ data. It should be noted that the parameters are no longer restricted to using year $t_2's$ converged parameters as is the case for $LL(\boldsymbol{\beta}_{t_2 t_1})$. This test is also reversed such that year t_1 above becomes year t_2 and year t_2 above becomes subset year t_1. The resulting χ^2 statistic can be used to determine if the null hypothesis that the parameters are equal in the two year can be rejected.

4. Estimation results and discussions

The NLOGIT 5 software is utilized for model estimation. It should be noted that although we introduced temporal instability, the detailed estimation results of each year are omitted due to the high complexity. Instead, the three-year overall parameter estimation results and corresponding pseudo-elasticity results are presented for discussing their

Table 2
Comparison results of MLMs with different distributions.

Model No. Distribution	1 Normal	2 Lognormal	3 Uniform
Significant random parameters (Injury Severity)	Curve (C), Male (C), Young driver (AK), Male (AK)	None	Curve (C), Male (C), Young driver (AK), Male (AK)
Log likelihood	−18138.98	−18226.73	−18130.34
Number of estimated model parameters	26	18	26
AIC	36329.96	36489.46	36312.68
BIC	36532.61	36629.76	36515.33

impacts on driver injury severity.

4.1. Mixed logit model estimation results

A simulation-based maximum likelihood estimation (MLE) method is used to estimate model parameters in the MLMs. Three models with different assumptions on parameter distributions, including normal, lognormal, and uniform distributions, are examined respectively. By balancing the computational cost-efficiency and model goodness-of-fit, simulations with 1000 Halton draws are applied in each model to provide an efficient estimation (Train, 2000). In addition, the O level is selected as the reference level. The comparison results of the three models are provided in Table 2.

As illustrated in Table 2, the results indicate that using lognormal distribution as the prior distribution assumption (Model 2) is not appropriate for this crash dataset since there is no significant random parameter found with this assumption. In contrast, several parameters (e.g., curve, male, etc.) are found to be randomly distributed in the other two models with the uniform and normal distributions assumptions (Model 1 and Model 3), indicating that the two distributions are both applicable for analyzing the dataset. In addition, Model 2 has the highest AIC and BIC values, also demonstrating that the lognormal distribution works inferior to the other two. Based on the rule-of-thumb of AIC and BIC (Schermelleh-Engel et al., 2003), using the uniform distribution as prior shows much better performance than using the normal distribution. Thus, Model 3 is selected as the final model in this study. The estimation results using MLM with uniform distribution simulated random parameters (Model 3) are illustrated in Table 3.

A variety of variables is found significantly associated with driver injury severity. The variable, *male driver*, is found as a random parameter affecting both possible injury (C) and more severe injury (AK), with a statistically significant mean and standard deviation with respect to each injury level. Besides, the variable, *curve*, is also a random parameter that has an influence on possible injury (C), although its mean is not significant based on the level of significance. This issue is not critical because whether a parameter is random or not is primarily based on the significance indication of its distribution of the standard deviation instead of its mean. The t-stat result of the standard deviation indicates that the variable, *curve*, is a random parameter. In addition, the variable, *young age*, is also found as a random parameter in the utility function of AK injury severity.

4.2. Latent class model estimation results

For the LCM, three different class numbers, 2–4, are separately tested with the same dataset, and their examination results are illustrated in Table 4. As the class number increases, the AIC and BIC values of the model also slightly increase, indicating that the performance becomes degraded. In addition, another two indices, the class probability and level of significance for each class, are also adopted for model selection. When the dataset is classified into two classes, the two classes involve 70% and 30% of the total data, respectively. When assuming the whole dataset contains three latent classes, the larger sub-dataset is then approximately evenly divided and the smaller one remains almost the same, as shown by the percentages 36%, 33% and 31%, respectively. While in the four-class scenario, approximately both the two sub-datasets in the two-class model are more delicately divided, and the final four classes contain 66%, 4%, 18%, and 12% of the whole data, respectively. However, the delicate division is not beneficial for improving model performance. It shows in Table 4 that the class probabilities in the three-class model are not statistically significant. In

Table 3
Estimation results of MLM.

Variable	Coefficient	Standard error	t-stat	95% Confidence interval	
				Lower	Upper
Constants					
C	2.43*	0.20	12.15	2.04	2.82
B	1.38*	0.18	7.67	1.03	1.73
AK	1.72*	0.19	9.05	1.35	2.09
Mean of random parameters					
Male (C)	−1.04*	0.27	−3.85	−1.57	−0.51
Curve (C)	−0.52	0.49	−1.06	−1.48	0.44
Male (AK)	−1.61*	0.53	−3.04	−2.65	−0.57
Young (AK)	−1.63*	0.51	−3.20	−2.63	−0.63
Spread scale of random parameters					
Male (C)	2.96*	0.32	9.26	2.33	3.59
Curve (C)	3.91*	0.71	5.51	2.52	5.31
Male (AK)	3.69*	0.44	8.38	2.83	4.55
Young (AK)	3.00*	0.45	6.66	2.11	3.88
Fixed parameters					
Semi (C)	−0.91*	0.39	−2.33	−1.67	−0.15
Straight (C)	−0.85*	0.07	−12.14	−0.99	−0.71
Drug/Alcohol Impaired (C)	1.86*	0.16	11.63	1.55	2.17
Signal Control (B)	−0.47*	0.04	−11.75	−0.55	−0.39
Multiple Lanes (B)	−0.36*	0.07	−5.14	−0.50	−0.22
Semi (B)	−0.84*	0.31	−2.71	−1.45	−0.23
Drug/Alcohol Impaired (B)	0.41*	0.11	3.73	0.19	0.63
Male (B)	−0.31*	0.04	−7.75	−0.39	−0.23
Curve (AK)	0.88*	0.11	8.00	0.66	1.10
On grade (AK)	0.57*	0.13	4.38	0.32	0.82
Wet (AK)	−0.61*	0.20	−3.05	−1.00	−0.22
Pick-up (AK)	0.31*	0.13	2.38	0.06	0.56
Semi (AK)	−1.48*	0.68	−2.18	−2.81	−0.15
Seatbelt not used (AK)	1.88*	0.17	11.06	1.55	2.21
Drug/Alcohol Impaired (AK)	2.08*	0.21	9.90	1.67	2.49
Model statistics					
Number of observations (*N*)	17,929.00				
Log-likelihood at constant	−34854.87				
Log-likelihood at convergence	−18130.34				
McFadden Pseudo R-squared	0.47				

* Level of significance ≤ 5%.

Table 4
Comparison results of LCMs with different numbers of class.

The number of latent classes	2	3	4
Log likelihood	−18038.60	−18196.38	−18203.31
Class probability	70%*/30%*	36%/33%/31%	66%*/4%*/18%*/12%*
Number of estimated model parameters (p)	17	18	12
AIC	36111.20	36428.76	36430.62
BIC	36243.70	36569.06	36524.15

* Level of significance ≤ 0.05.

Table 5
Estimation results of LCM.

Variable	Latent Class 1					Latent Class 2				
	Coef.[a]	S.E.[b]	t-stat	95% CI[c]		Coef.[a]	S.E.[b]	t-stat	95% CI[c]	
				Lower	Upper				Lower	Upper
Constant										
C	3.62*	0.61	5.93	2.42	4.82	0.13	0.36	0.36	−0.57	0.83
B	1.67*	0.62	2.69	0.46	2.89	0.70*	0.31	2.26	0.1	1.31
AK	0.26	1.36	0.19	−2.41	2.93	3.28*	0.55	5.96	2.21	4.35
Non-constant Parameter										
Curve (C)	0.58*	0.13	4.46	0.32	0.84	–	–		–	–
Straight (C)	–	–		–	–	−1.33*	0.23	−5.78	−1.78	−0.88
Drug/Alcohol Impaired (C)	1.50*	0.13	11.54	1.25	1.77	–	–		–	–
Male (C)	–	–		–	–	−0.70*	0.22	−3.18	−1.14	−0.28
Signal Control (B)	2.52*	1.12	2.25	0.34	4.71	−4.86*	1.32	−3.68	−7.45	−2.28
Multiple Lanes (B)	−0.26*	0.11	−2.36	−0.48	−0.05	−1.71*	0.43	−3.98	−2.55	−0.87
Male (B)	−0.34*	0.1	−3.40	−0.55	−0.14	−1.51*	0.36	−4.19	−2.21	−0.81
On grade (AK)	1.16*	0.35	3.31	0.47	1.86	–	–		–	–
Wet (AK)	−1.09*	0.4	−2.73	−1.88	−0.32	–	–		–	–
Pick-up (AK)	1.55*	0.4	3.88	0.78	2.33	–	–		–	–
Young (AK)	−0.73*	0.3	−2.43	−1.33	−0.14	−0.65*	0.11	−5.91	−0.88	−0.43
Seatbelt not used (AK)	2.50*	0.47	5.32	1.58	3.43	1.01*	0.23	4.39	0.57	1.46
Drug/Alcohol Impaired (AK)	3.50*	0.54	6.48	2.51	4.61	–	–		–	–
Male (AK)	−0.88*	0.29	−3.03	−1.45	−0.31	−0.35*	0.17	−2.06	−0.69	−0.01
Class probability	0.70*					0.30*				
Model statistics										
Number of observations (*N*)	17929									
Log-likelihood at constant	−34854.87									
Log-likelihood at convergence	−18038.6									
McFadden Pseudo R-square	0.48									

[a] Coefficient.
[b] Standard error.
[c] 95% confidence interval of estimation results.
* Level of significance ≤ 0.05.

addition, although all the class probabilities in the four-class scenario are significant, the deficient parameter estimation results show that this kind of division is less meaningful since considerably fewer parameters (12 versus 17/18) are found to be significantly related to driver injury severity. Therefore, the two-class model is selected as the final model for studying the variables' impacts on driver injury severity outcomes in rural single-vehicle crashes under rain conditions.

The detailed estimation results of this model are listed in Table 5. It shows that remarkable differences exist between the two latent classes, since the variables that significantly influence driver injury severity to distribute quite diversely in the two classes. For instance, Drug/Alcohol-Impaired (C) and Drug/Alcohol-Impaired (AK) are found significantly affecting driver injury severities in Latent Class 1, whereas insignificant in Latent Class 2. These distinctive outcomes suggest that the data has a multivariate categorical nature and demonstrate the latent class logit model is appropriate for analyzing the crash dataset.

4.3. Pseudo-elasticity analysis results

In order to explain the impacts of these variables accurately, the average pseudo-elasticity is adopted on both the MLM and the LCM, and the results are listed in Table 6. Twelve variables regarding road geometric characteristics, road surface conditions, vehicle types, and driver demographic information and behavior, are found significant in the MLM while the elasticity analysis results of the two models are quite similar in terms of colored magnitude category, although the estimation values are not exactly same. The detailed discussions of these variables are presented in the following sections.

Pseudo-elasticity analysis results of both the two models show that the variable, *curve*, has an influence on increasing driver injury severity in rural single-vehicle crashes under rain conditions. In the MLM, this variable increases the probabilities of driver injury serveries, B level and AK level injuries, by 35.87% and 84.14%, respectively, while the corresponding values in the LCM are 34.69% and 20.13%, respectively. These results are consistent with previous studies where it was found that severe injuries are more likely to happen on the curve roads (Holdridge et al., 2005; Ye and Lord, 2014). Considering the impacts of rain, driving on curve roads becomes more challenging. It may take more time for vehicles to decelerate to a safe speed when running on a curve roadway because of the low friction on wet road surface. In addition, also due to the low friction, vehicles may easily lose control when sudden changes are applied to speed or steering while running at high speed on wet curve roads. Accordingly, possible countermeasures on this issue include increasing the radius of the curve where possible, installing a speed indicator at the beginning of the curve, paving the curve road with materials with high resistance, etc.

The variable, *road grade*, is also significantly associated with driver injury severity by showing that it can increase the likelihood of the driver being severe injured (AK level) by over 50% in both two models. The reasons of these results are complicated, one of the important factors is that vehicle brakes are more frequently used to maintain the vehicle stable while driving on the graded roadway, which may increase the risk of brake failure and then lead to the vehicle losing control. Similar findings have also been discovered by previous studies (Khattak, 2001; Quddus et al., 2009; Li et al., 2018a), and the same influence exists not only in single-vehicle crashes but also in multi-vehicle crashes. Enhanced delineation treatments on the roadway can alert drivers in advance of grade roads and vary depending on the severity of the grade and the driving speed. In addition, high friction surface materials and treatments also can be implemented to help the drivers to maintain speeds when driving on grade roadways (Li et al., 2018a).

The variable, *wet*, is found to significantly decrease the possibilities of AK level injuries in both two models. The results seem to be contrary

Table 6
Average pseudo-elasticity analysis for MLM and LCM.

Variable	MLM				LCM			
	O	C	B	AK	O	C	B	AK
Curve	−3.62%	−24.33%	35.87%	84.14%	−3.36%	−9.43%	34.69%	20.13%
On Grade	−2.55%	−2.37%	−2.37%	52.77%	−1.52%	−1.84%	−2.93%	59.88%
Wet	2.81%	2.34%	2.63%	−35.23%	2.13%	1.83%	4.69%	−43.94%
Signal Control	9.08%	−27.20%	7.20%	6.51%	9.06%	−41.38%	21.11%	48.08%
Multiple Lanes	7.62%	−25.39%	5.92%	5.40%	8.21%	−38.20%	11.31%	23.88%
Pick-up	−1.39%	−1.11%	−1.12%	28.22%	−1.60%	−2.58%	−2.90%	68.28%
Semi	30.52%	−47.83%	−39.72%	−61.11%	–	–	–	–
Straight	10.62%	10.20%	−43.42%	9.93%	6.56%	3.99%	−29.60%	11.79%
Drug/Alcohol Impaired	−42.37%	−10.58%	152.74%	204.72%	−43.73%	−15.46%	136.33%	502.67%
Seatbelt not used	−11.73%	−10.77%	−11.34%	265.43%	−7.63%	−6.50%	−13.89%	318.38%
Male	11.19%	−18.72%	−11.49%	−5.27%	12.27%	−24.98%	−8.93%	−16.81%
Young	2.47%	1.77%	2.22%	−38.24%	1.42%	0.66%	3.94%	−42.76%

to everyday experience, however, in fact many previous studies have already obtained similar conclusions (Lee et al., 2015; Wu et al., 2014). The reason may be that the drivers tend to adapt their speeds to the adverse road conditions to some degree while driving on the wet roads. This behavior is due to the driver's active behavioral adjustments to adverse external conditions to reduce and maintain low perceived driving risk, which is an example of risk compensation behavior. Interested readers are referred to the paper by Mannering and Bhat (2014) and the references cited therein.

It is not surprising that the variable, *signal control*, can aggravate the driver injury severity in both two models. Signal control devices are mainly located at the intersections that are among the locations with the most complex traffic conditions in a road network (Wang and Abdel-Aty, 2008). In the U.S., although only around 10% of all intersections were signalized, nearly 30% of intersection-related fatalities occurred at signalized intersections (Rice, 2007). Traffic signals, especially for the left-turn and through traffic, may increase the number of conflict points as well as accident potential on a roadway. This challenge becomes more serious under rain conditions. Due to the limited visibility and low pavement friction under rain conditions, the available response distances for drivers at intersections is significantly reduced, and therefore crash risks and severe injury possibilities are notably increased.

As illustrated in Table 6, the variable, *multiple lanes*, slightly increases the possibilities of AK level crashes. A probable explanation is that multiple lanes are always associated complex roadway and traffic conditions, e.g., more exclusive turning lanes, frequent lane changing behaviors, variable speed limits across lanes, etc., and thus may pose more challenges to the drivers in the rain. This finding is consistent with the results of previous studies (Aziz et al., 2013; Wang et al., 2006), where it was found that crashes on multi-lane roads have a higher probability of fatality. On the contrary, single lane roads were found to have a lower probability of leading to severe injuries and fatalities.

The variable, *pickup*, is found significantly associated with driver injury severity in the rural single-vehicle crashes under rain conditions in the two models. More specifically, pickup drivers are more likely to suffer serious injuries and fatalities in rain-related single-vehicle crashes since the pseudo-elasticity analysis results showed that this variable could increase the probabilities of AK level injuries by 28.22% and 68.28% in the two models, respectively. The reason is understandable that driving a pickup requires more driving skills and experiences than driving a passenger car. Besides, the higher inertia, which results from the larger mass of pickups comparing to passenger vehicles, also makes it more difficult to maintain safe driving, especially on the slippery road surface. Due to the difficulties in vehicle operation, rollovers, collisions with fixed objects, and other crash types with severe outcomes are more likely to occur in pickup related accidents.

The variable, *Semi-trailers*, can reduce the possibilities of the driver being seriously injured according to their low pseudo-elasticity results in serious crashes in the MLM. However, this variable is not significant in the LCM. Previous studies also implied that semi-trailer is a variable that has adverse effects on driver injury severity (Carson and Mannering, 2001; Celik and Oktay, 2014; Chen et al., 2016a). For instance, Chen et al. (2016a) found that semi-trailers are less related to severe injuries, indicated by the negative estimated coefficient. Carson and Mannering (2001) concluded that semi-trailers could increase the probability of fatal injuries due to its relatively large size and weight. Therefore, more efforts should be made to figure out the underlying reason for these aforementioned impacts.

This variable, *drug/alcohol-impaired*, contains two aspects, i.e., *drug* and *alcohol*. However, they have quite similar impacts in compromising drivers' sobriety and reasonable judgment, but did not have enough records of presence in the studied dataset, and therefore were combined together in this study. The combined variable, representing the drivers' state of consciousness, is expected to aggravate driver injury severity significantly. As shown in Table 6, *driver alcohol/drug impairment* can increase the potential for injuries and fatalities (B and AK levels) by over 150% in the two models. The results are reasonable since drug and alcohol can easily affect drivers' physical and psychological functions, e.g., body balance, vision, sobriety, reaction time, etc., and bring about a series of consequences, including misjudgment, short-term memory loss, reduced information processing capability, and impaired perception. Therefore, engineering, enforcement and educational (3E) traffic safety-related countermeasures are needed, including highway safety patrol and sobriety check on the random and timely basis, enforced punishment for driving under the influence (DUI), and related defensive driving programs.

As shown in Table 6, when the seatbelt is not used, the likelihood of drivers suffering AK level injury dramatically increases in the two models, suggesting that using a seatbelt is an effective way of protecting the driver in a rural single-vehicle crash under rain condition. The favorable effects of seatbelt usage have also been evaluated by sufficient previous research (Abay et al., 2013; Chu, 2014; Yasmin et al., 2014). Seatbelt, which is an important part of vehicle design, can secure the occupant of a vehicle against harmful movement during a collision or a sudden stop. When faced with crashes or other urgent circumstances or in the rain, it is harder for drivers to maintain normal driving on the slippery roadway, and thus seatbelts become more necessary to secure the driver against fierce movement and potential collision impact. Therefore, 3E efforts are also needed to ensure seatbelt usage on each occupant in every vehicle ride. For instance, a useful countermeasure could be video recognition techniques through roadway cameras applied to identify seatbelt usage status on vehicle occupants and issue traffic violation tickets and fines to those not wearing seatbelts, without violating people's rights of privacy.

The elasticity analysis results suggest that *male* drivers have less likelihood of serious injury and fatality when comparing with female drivers. The reason for this finding may be that male drivers have relatively experienced operation skills and additional physiological strength, and can better handle complex road and environment situations under rain conditions. Other scholars have also found similar results that male drivers demonstrate better driving performance and safety levels in the areas of the complex external environment that require additional driving skills than during average-day driving (Staff et al., 2014; Yasmin et al., 2014).

The *age* of drivers involved in the crashes is found to be a significant variable on driver injury severity in both two models. The pseudo-elasticity analysis results indicate that young drivers are associated with reductions in the possibilities of driver serious injuries and fatalities in rural single-vehicle crashes under rain conditions. This is because young drivers have faster reactions than drivers in other age groups due to their physical flexibility (Castro et al., 2013; Xie et al., 2012). The relative lower average speeds in the rain may also contribute to reducing the injury severity of young drivers who are more likely to conduct speeding or reckless driving (Bolderdijk et al., 2011; Ulleberg, 2001). Furthermore, it is found in the dataset that 7.64% of young drivers drive pickup trucks, while the numbers in mid-age and old drivers are 15.30% and 10.66%, respectively. Given various impacts of vehicle types in the previous section, it is also evident to conclude that on average young drivers are safer than drivers of other age groups in rural single-vehicle crashes under rain conditions.

4.4. Comparison between mixed logit model and latent class model

A statistical comparison can better demonstrate which model is more proper for this dataset. As shown in Table 2, AIC and BIC of the MLM are 36312.68 and 36515.33, respectively. The same indices are also presented in Table 4 for the LCM, and are 36111.20 and 36243.70, respectively. The relatively lower AIC and BIC values of the LCM indicate the model has slightly better performance than the MLM. In addition, the prediction success index, McFadden pseudo r-squared (0.47 for the MLM, 0.48 for the LCM), suggests that the LCM has slightly better predictive capability than the MLM. As shown in Table 7, it can also be observed that the LCM predicted probabilities for the O and C levels (contain over 80% observations) are closer to the observations than the ones predicted by the MLM.

Tables 8 and 9 present the likelihood ratio tests between different years of the MLM and the LCM, respectively. The results show that both models are not temporally stable. This finding is in line with some recent research (Behnood and Mannering, 2015, 2016). It is impossible to determine whether this temporal instability is due to some underlying influence of factors that affecting driver injury severity, or the result of changes induced by variations in economic conditions. However, the significant differences indicate that the effect of explanatory variables on driver injury severity of rain-related crashes has shifted over the years studied. Moreover, the LCM shows that it tends to provide lower χ^2 values in the temporal stability tests. In addition, some χ^2 values are not significant, indicating the model is temporal stable in some extent.

Table 8
Likelihood ratio test results between different years based on MLM (χ^2 values with degrees of freedom in parenthesis and confidence level in brackets).

t_1	t_2 2012	2013	2014
2012	–	47.332 (24) [> 99.70%]	54.631 (26) [> 99.91%]
2013	48.350 (26) [> 99.50%]	–	60.232 (25) [> 99.99%]
2014	55.253 (26) [> 99.99%]	53.282 (24) [> 99.99%]	–

Table 9
Likelihood ratio test results between different years based on LCM (χ^2 values with degrees of freedom in parenthesis and confidence level in brackets).

t_1	t_2 2012	2013	2014
2012	–	34.332 (16) [> 99.50%]	22.218 (17) [> 82.22%]
2013	27.622 (17) [> 99.50%]	–	40.335 (17) [> 99.99%]
2014	24.333 (17) [> 89.00%]	33.513 (16) [> 99.40%]	–

It should also be noted that the temporal instability indicates the estimation results of these two models may not be able to fully reveal the actual effects of the variables, and thus more advanced models that can account this issue are recommended in the future studies.

Therefore, with regard to the comparison between the MLM and the LCM, it could be concluded that the LCM is slightly superior to the MLM in the modeling process on the studied dataset regarding rural single-vehicle crashes under rain conditions. However, it is well noteworthy that the differences between the two models are rather modest, as the estimated parameters of the two models and the corresponding pseudo-elastic results are not significantly different.

5. Conclusions

A three-year crash dataset including all rural single-vehicle crashes under rain conditions in four South Central states, i.e., Texas, Arkansas, Oklahoma, and Louisiana, from 2012 to 2014 is selected in this paper to analyze the impact factors on driver injury severity. The MLM and the LCM are both developed in this study on the identical dataset. This paper has both methodological and empirical contributions to the field of traffic safety analysis. Several ongoing debates, including distributions of random parameters and efficient iterations in the MLM, the optimal class number of LCM, and comparison of the two models, are all discussed in the study. Statistical parsimony indices, including AIC, BIC, as well as McFadden pseudo r-squared, are calculated for each model to evaluate their respective performance. In addition, their abilities to capture temporal instability are also calculated via

Table 7
MLM and LCM estimated outcome probabilities compared to observed severity outcomes.

Components	Mixed Logit Model	Latent Class Model: Latent Class 1	Latent Class 2	Overall	Observed
Crash population share		0.70	0.30		
Crash injury severity					
O	0.594 (−1.82%)	0.604	0.595	0.601 (−0.66%)	0.605
C	0.223 (3.72%)	0.213	0.226	0.217 (0.93%)	0.215
B	0.131 (−1.50%)	0.122	0.135	0.126 (−5.26%)	0.133
AK	0.052 (10.64%)	0.061	0.044	0.056 (19.15%)	0.047

likelihood ratio tests. Results show that choosing uniform distribution as prior for random parameters could better increase goodness-of-fit of the MLM than using normal and lognormal distributions. In addition, the two-class LCM also shows superiority when compared to the three- and four-class models. Moreover, a careful comparison between these two best models of their kinds is also conducted, and the results indicate that the LCM works slightly better in analyzing the aforementioned dataset in this study.

A series of significant contributing factors in terms of road geometric characters, traffic compositions and dynamics, driver demographic features, etc., are identified and compared with the two models. To better explain the model estimation results, pseudo-elasticity analyses of the significant factors are conducted. The results reveal that *curve*, *on grade*, *signal control*, *multiple lanes*, *pickup*, *straight*, *drug/alcohol impaired*, and *seat belt not used* can increase driver injury severity in the two models. On the other hand, *wet*, *male*, *semi*, and *young* are found to decrease driver injury outcomes. These results are not only useful for understanding the underlying risk factors of rural single-vehicle crashes under rain conditions, but can also provide meaningful references developing appropriate countermeasures, strategies, and policies to mitigate the driver injury severity of relative crashes worldwide. In addition, effective management and planning, technical implementation guide of specific countermeasures, and political support and leadership are necessary and should be fulfilled together to improve traffic safety.

Acknowledgment

The authors sincerely thank the Texas Department of Transportation, Arkansas State Highway and Transportation Department, Oklahoma Department of Transportation, and Louisiana Department of Transportation and Development for the data support on this research. The interpretations are only those of the authors and do not necessarily reflect the views of any organizations with which the authors have an affiliation.

References

Abay, K.A., Paleti, R., Bhat, C.R., 2013. The joint analysis of injury severity of drivers in two-vehicle crashes accommodating seat belt use endogeneity. Transp. Res. Part B: Methodol. 50, 74–89.
Abdel-Aty, M., 2003. Analysis of driver injury severity levels at multiple locations using ordered probit models. J. Safety Res. 34, 597–603. https://doi.org/10.1016/j.jsr.2003.05.009.
Adams, K.J., Drenner, R.W., Chumchal, M.M., Donato, D.I., 2016. Disparity between state fish consumption advisory systems for methylmercury and US Environmental Protection Agency recommendations: a case study of the south central United States. Environ. Toxicol. Chem. 35, 247–251.
Aguilar, E., Peterson, T.C., Obando, P.R., Frutos, R., Retana, J.A., Solera, M., Soley, J., García, I.G., Araujo, R.M., Santos, A.R., 2005. Changes in precipitation and temperature extremes in Central America and northern South America, 1961–2003. J. Geophys. Res.: Atmos. 110.
Andrey, J., Yagar, S., 1993. A temporal analysis of rain-related crash risk. Accid. Anal. Prev. 25, 465–472.
Aziz, H.M.A., Ukkusuri, S.V., Hasan, S., 2013. Exploring the determinants of pedestrian–vehicle crash severity in New York City. Accid. Anal. Prev. 50, 1298–1309.
Behnood, A., Mannering, F., 2015. The temporal stability of factors affecting driver-injury severities in single-vehicle crashes: some empirical evidence. Anal. Methods Accid. Res. 8, 7–32.
Behnood, A., Mannering, F., 2016. An empirical assessment of the effects of economic recessions on pedestrian-injury crashes using mixed and latent-class models. Anal. Methods Accid. Res. 12, 1–17. https://doi.org/10.1016/j.amar.2016.07.002.
Behnood, A., Mannering, F., 2017a. Determinants of bicyclist injury severities in bicycle-vehicle crashes: a random parameters approach with heterogeneity in means and variances. Anal. Methods Accid. Res. 16, 35–47.
Behnood, A., Mannering, F., 2017b. The effect of passengers on driver-injury severities in single-vehicle crashes: a random parameters heterogeneity-in-means approach. Anal. Methods Accid. Res. 14, 41–53. https://doi.org/10.1016/j.amar.2017.04.001.
Behnood, A., Roshandeh, A.M., Mannering, F.L., 2014. Latent class analysis of the effects of age, gender, and alcohol consumption on driver-injury severities. Anal. Methods Accid. Res. 3, 56–91.
Berndt, E.R., Hall, B.H., Hall, R.E., Hausman, J.A., 1974. Estimation and inference in nonlinear structural models. Ann. Econ. Soc. Meas. 3 (4), 653–665 NBER.
Bolderdijk, J.W., Knockaert, J., Steg, E.M., Verhoef, E.T., 2011. Effects of pay-as-you-drive vehicle insurance on young drivers' speed choice: results of a Dutch field experiment. Accid. Anal. Prev. 43, 1181–1186.
Carson, J., Mannering, F., 2001. The effect of ice warning signs on ice-accident frequencies and severities. Accid. Anal. Prev. 33, 99–109.
Carter, C.E., Greer, J.D., Braud, H.J., Floyd, J.M., 1974. Raindrop characteristics in south central United States. Trans. ASAE 17, 1033–1037.
Castro, M., Paleti, R., Bhat, C.R., 2013. A spatial generalized ordered response model to examine highway crash injury severity. Accid. Anal. Prev. 52, 188–203.
Celik, A.K., Oktay, E., 2014. A multinomial logit analysis of risk factors influencing road traffic injury severities in the Erzurum and Kars Provinces of Turkey. Accid. Anal. Prev. 72, 66–77.
Cerwick, D.M., Gkritza, K., Shaheed, M.S., Hans, Z., 2014. A comparison of the mixed logit and latent class methods for crash severity analysis. Anal. Methods Accid. Res. 3, 11–27.
Chang, L.-Y., Chien, J.-T., 2013. Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model. Saf. Sci. 51, 17–22.
Chen, C., Zhang, G., Tarefder, R., Ma, J., Wei, H., Guan, H., 2015a. A multinomial logit model-Bayesian network hybrid approach for driver injury severity analyses in rear-end crashes. Accid. Anal. Prev. 80, 76–88.
Chen, C., Zhang, G., Tian, Z., Bogus, S.M., Yang, Y., 2015b. Hierarchical Bayesian random intercept model-based cross-level interaction decomposition for truck driver injury severity investigations. Accid. Anal. Prev. 85, 186–198.
Chen, C., Zhang, G., Huang, H., Wang, J., Tarefder, R.A., 2016a. Examining driver injury severity outcomes in rural non-interstate roadway crashes using a hierarchical ordered logit model. Accid. Anal. Prev. 96, 79–87.
Chen, C., Zhang, G., Qian, Z., Tarefder, R.A., Tian, Z., 2016b. Investigating driver injury severity patterns in rollover crashes using support vector machine models. Accid. Anal. Prev. 90, 128–139. https://doi.org/10.1016/j.aap.2016.02.011.
Chen, C., Zhang, G., Yang, J., Milton, J.C., 2016c. An explanatory analysis of driver injury severity in rear-end crashes using a decision table/Naïve Bayes (DTNB) hybrid classifier. Accid. Anal. Prev. 90, 95–107.
Chu, H.-C., 2014. Assessing factors causing severe injuries in crashes of high-deck buses in long-distance driving on freeways. Accid. Anal. Prev. 62, 130–136.
Domencich, T.A., McFadden, D., 1975. Urban Travel Demand-A Behavioral Analysis.
Feng, S., Li, Z., Ci, Y., Zhang, G., 2016. Risk factors affecting fatal bus accident severity: their impact on different types of bus drivers. Accid. Anal. Prev. 86, 29–39. https://doi.org/10.1016/j.aap.2015.09.025.
Gelman, A., Hill, J., 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge 651. https://doi.org/10.2277/0521867061.
Greene, W., 2012. NLOGIT, Version 5. Econometric Software Inc., Plainview, NY.
Hensher, D.A., Greene, W.H., 2003. The mixed logit model: the state of practice. Transportation 30, 133–176.
Holdridge, J.M., Shankar, V.N., Ulfarsson, G.F., 2005. The crash severity impacts of fixed roadside objects. J. Safety Res. 36, 139–147.
Islam, S., Brown, J., 2017. A comparative injury severity analysis of motorcycle at-fault crashes on rural and urban roadways in Alabama. Accid. Anal. Prev. 108, 163–171.
Jackson, T.L., Sharif, H.O., 2014. Rainfall impacts on traffic safety: rain-related fatal crashes in Texas. Geomat. Nat. Hazards Risk 7, 1–18. https://doi.org/10.1080/19475705.2014.984246.
Jung, S., Qin, X., Noyce, D.A., 2010. Rainfall effect on single-vehicle crash severities using polychotomous response models. Accid. Anal. Prev. 42, 213–224. https://doi.org/10.1016/j.aap.2009.07.020.
Khattak, A., 2001. Injury severity in multivehicle rear-end crashes. Transp. Res. Record: J. Transp. Res. Board 59–68.
Kim, J.-K., Kim, S., Ulfarsson, G.F., Porrello, L.A., 2007. Bicyclist injury severities in bicycle–motor vehicle accidents. Accid. Anal. Prev. 39, 238–251.
Kim, J.-K., Ulfarsson, G.F., Shankar, V.N., Kim, S., 2008. Age and pedestrian injury severity in motor-vehicle crashes: a heteroskedastic logit analysis. Accid. Anal. Prev. 40, 1695–1702. https://doi.org/10.1016/j.aap.2008.06.005.
Kim, J.-K., Ulfarsson, G.F., Kim, S., Shankar, V.N., 2013. Driver-injury severity in single-vehicle crashes in California: a mixed logit analysis of heterogeneity due to age and gender. Accid. Anal. Prev. 50, 1073–1081. https://doi.org/10.1016/j.aap.2012.08.011.
Lazarsfeld, P.F., Henry, N.W., Anderson, T.W., 1968. Latent Structure Analysis. Houghton Mifflin, Boston.
Lee, J., Nam, B., Abdel-Aty, M., 2015. Effects of pavement surface conditions on traffic crash severity. J. Transp. Eng. 141 4015020.
Li, Z., Chen, C., Ci, Y., Zhang, G., Wu, Q., Liu, C., Qian (Sean), Z., 2018a. Examining driver injury severity in intersection-related crashes using cluster analysis and hierarchical Bayesian models. Accid. Anal. Prev. 120, 139–151. https://doi.org/10.1016/J.AAP.2018.08.009.
Li, Z., Chen, C., Wu, Q., Zhang, G., Liu, C., Prevedouros, P.D., Ma, D.T., 2018b. 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. 20, 1–14. https://doi.org/10.1016/j.amar.2018.08.001.
Luo, L., Zhou, S., Cai, W., Yoke, M., Low, H., Tian, F., Wang, Y., Xiao, X., Chen, D., 2008. Agent-based human behavior modeling for crowd simulation. Comput. Animat. Virtual Worlds 19, 271–281. https://doi.org/10.1002/cav.
Ma, L., Wang, G., Yan, X., Weng, J., 2016. A hybrid finite mixture model for exploring heterogeneous ordering patterns of driver injury severity. Accid. Anal. Prev. 89, 62–73.
Mannering, F., 2018. Temporal instability and the analysis of highway accident data. Anal. Methods Accid. Res. 17, 1–13.
Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: methodological frontier and future directions. Anal. Methods Accid. Res. 1, 1–22. https://doi.org/10.1016/j.amar.2013.09.001.
Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the

statistical analysis of highway accident data. Anal. Methods Accid. Res. 11, 1–16. https://doi.org/10.1016/j.amar.2016.04.001.
McFadden, D., 1981. Econometric models of probabilistic choice. In: Manski, C., McFadden, D. (Eds.), Structural Analysis of Discrete Data with Econometric Application. MIT Press, Cambridge, pp. 198–272.
McFadden, D., Train, K., 2000. Mixed mnl models of discrete response. Journal of Applied Econometrics 15, 447–470.
McFadden, D., Zarembka, P., 1974. Frontiers in Econometrics. Conditional Logit Analysis of Qualitative Choice Behavior. pp. 105–142.
Miller, S.M., Wofsy, S.C., Michalak, A.M., Kort, E.A., Andrews, A.E., Biraud, S.C., Dlugokencky, E.J., Eluszkiewicz, J., Fischer, M.L., Janssens-Maenhout, G., 2013. Anthropogenic emissions of methane in the United States. Proc. Natl. Acad. Sci. 110, 20018–20022.
Milton, J.C., Shankar, V.N., Mannering, F.L., 2008. Highway accident severities and the mixed logit model: an exploratory empirical analysis. Accid. Anal. Prev. 40, 260–266.
Moore, D.N., Schneider IV, W.H., Savolainen, P.T., Farzaneh, M., 2011. Mixed logit analysis of bicyclist injury severity resulting from motor vehicle crashes at intersection and non-intersection locations. Accid. Anal. Prev. 43, 621–630.
Munn, I.A., Barlow, S.A., Evans, D.L., Cleaves, D., 2002. Urbanization's impact on timber harvesting in the south central United States. J. Environ. Manage. 64, 65–76.
NHTSA, 2017. Rural/Urban Comparison of Traffic Fatalities, Traffic Safety Facts.
OKDOT, 2015. Oklahoma Traffic Crashes Annual Report (2014).
Osman, M., Paleti, R., Mishra, S., Golias, M.M., 2016. Analysis of injury severity of large truck crashes in work zones. Accid. Anal. Prev. 97, 261–273.
Qiu, L., Nixon, W., 2008. Effects of adverse weather on traffic crashes: systematic review and meta-analysis. Transp. Res. Record: J. Transp. Res. Board 2055, 139–146. https://doi.org/10.3141/2055-16.
Quddus, M.A., Wang, C., Ison, S.G., 2009. Road traffic congestion and crash severity: econometric analysis using ordered response models. J. Transp. Eng. 136, 424–435.
Rice, E., 2007. Taking action to reduce intersection fatalities. Safety Compass 1, 1–3.
Rusli, R., Haque, M.M., King, M., Voon, W.S., 2017. Single-vehicle crashes along rural mountainous highways in Malaysia: an application of random parameters negative binomial model. Accid. Anal. Prev. 102, 153–164.
Sasidharan, L., Wu, K.-F., Menendez, M., 2015. Exploring the application of latent class cluster analysis for investigating pedestrian crash injury severities in Switzerland. Accid. Anal. Prev. 85, 219–228.
Schermelleh-Engel, K., Moosbrugger, H., Müller, H., 2003. Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. Methods Psychol. Res. Online 8, 23–74.
Seraneeprakarn, P., Huang, S., Shankar, V., Mannering, F., Venkataraman, N., Milton, J., 2017. Occupant injury severities in hybrid-vehicle involved crashes: a random parameters approach with heterogeneity in means and variances. Anal. Methods Accid. Res. 15, 41–55.
Shaheed, M.S., Gkritza, K., Carriquiry, A.L., Hallmark, S.L., 2016. Analysis of occupant injury severity in winter weather crashes: a fully Bayesian multivariate approach. Anal. Methods Accid. Res. 11, 33–47.
Staff, T., Eken, T., Wik, L., Røislien, J., Søvik, S., 2014. Physiologic, demographic and mechanistic factors predicting New Injury Severity Score (NISS) in motor vehicle accident victims. Injury 45, 9–15.
Train, K., 2000. Halton Sequences for Mixed Logit. Department of Economics, UCB.
TxDOT, 2016. Weather Conditions for Crashes (2015).
Ulleberg, P., 2001. Personality subtypes of young drivers. Relationship to risk-taking preferences, accident involvement, and response to a traffic safety campaign. Transp. Res. Part F: Traffic Psychol. Behav. 4, 279–297.
Wang, X., Abdel-Aty, M., 2008. Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models. Accid. Anal. Prev. 40, 1674–1682.
Wang, X., Abdel-Aty, M., Brady, P., 2006. Crash estimation at signalized intersections: significant factors and temporal effect. Transp. Res. Record: J. Transp. Res. Board 10–20.
Wang, K., Ivan, J.N., Ravishanker, N., Jackson, E., 2017. Multivariate poisson lognormal modeling of crashes by type and severity on rural two lane highways. Accid. Anal. Prev. 99, 6–19.
Washington, S., Karlaftis, M., Mannering, F., 2011. Statistical and Econometric Methods for Transportation Data Analysis. CRC press.
Wu, Q., Chen, F., Zhang, G., Liu, X.C., Wang, H., Bogus, S.M., 2014. Mixed logit model-based driver injury severity investigations in single-and multi-vehicle crashes on rural two-lane highways. Accid. Anal. Prev. 72, 105–115.
Xie, Y., Zhao, K., Huynh, N., 2012. Analysis of driver injury severity in rural single-vehicle crashes. Accid. Anal. Prev. 47, 36–44. https://doi.org/10.1016/j.aap.2011.12.012.
Yasmin, S., Eluru, N., Bhat, C.R., Tay, R., 2014. A latent segmentation based generalized ordered logit model to examine factors influencing driver injury severity. Anal. Methods Accid. Res. 1, 23–38.
Ye, F., Lord, D., 2014. Comparing three commonly used crash severity models on sample size requirements: multinomial logit, ordered probit and mixed logit models. Anal. Methods Accid. Res. 1, 72–85.