Hindawi
Journal of Advanced Transportation
Volume 2018, Article ID 2702360, 9 pages
https://doi.org/10.1155/2018/2702360

Research Article

Investigating the Differences of Single-Vehicle and Multivehicle Accident Probability Using Mixed Logit Model

Bowen Dong[1], **Xiaoxiang Ma**[1], **Feng Chen**[1] **and Suren Chen**[2]

[1]*College of Traffic Engineering and Key Laboratory of Road & Traffic Engineering of the Ministry of Education, Tongji University, 4800 Cao'an Road, Shanghai 201804, China*
[2]*Department of Civil & Environmental Engineering, Colorado State University, Fort Collins, CO 80523, USA*

Correspondence should be addressed to Feng Chen; fengchen@tongji.edu.cn

Received 15 June 2017; Accepted 27 November 2017; Published 17 January 2018

Academic Editor: Helai Huang

Road traffic accidents are believed to be associated with not only road geometric feature and traffic characteristic, but also weather condition. To address these safety issues, it is of paramount importance to understand how these factors affect the occurrences of the crashes. Existing studies have suggested that the mechanisms of single-vehicle (SV) accidents and multivehicle (MV) accidents can be very different. Few studies were conducted to examine the difference of SV and MV accident probability by addressing unobserved heterogeneity at the same time. To investigate the different contributing factors on SV and MV, a mixed logit model is employed using disaggregated data with the response variable categorized as no accidents, SV accidents, and MV accidents. The results indicate that, in addition to speed gap, length of segment, and wet road surfaces which are significant for both SV and MV accidents, most of other variables are significant only for MV accidents. Traffic, road, and surface characteristics are main influence factors of SV and MV accident possibility. Hourly traffic volume, inside shoulder width, and wet road surface are found to produce statistically significant random parameters. Their effects on the possibility of SV and MV accident vary across different road segments.

1. Introduction

Given the economic costs and human casualties motor vehicle crashes continue to claim, traffic safety remains a hot topic among researchers across the world. Over the last decades, traffic safety researchers have spent tremendous effort and time to gain a better understanding of the contributory factors towards motor vehicle crash [1–6]. Despite the progress, there are many knowledge gaps yet to be filled in safety-related studies. One of these gaps pertains to the differences between single-vehicle and multivehicle crashes. As shown by previous studies [7–9], the mechanisms of single-vehicle (SV) accidents and multivehicle (MV) accidents are inherently different. Knipling [8] pointed out that single-vehicle and multivehicle crashes were related to different kinds of driver errors. Specifically, single-vehicle crashes usually resulted from loss of vehicle control that is associated with driver misbehavior. Multivehicle crashes, on the other hand, are most often caused by driver errors when interacting with other vehicles. It is therefore important to identify different contributing factors between single-vehicle and multivehicle crashes, which further offers insights for countermeasures to mitigate SV and MV crash risk, respectively. Along with this line of research, separate models were developed at first for SV and MV crashes to account for the difference between them [7, 9–12]. However, those models largely ignored the shared unobserved effects between SV and MV crashes, which is problematic and leads to inconsistent estimates [13]. To account for these shared heterogeneities between SV and MV crashes, researchers have proposed advanced models such as bivariate Poisson-gamma/lognormal models to study SV and MV accidents jointly [14–17]. These previous studies mainly employed count data models and had undoubtedly provided many useful finding which contributed to the overall understanding of the characteristics of SV and MV accidents.

Though many safety studies have already been conducted, the investigations on the mechanism of SV and

MV crashes especially those using disaggregate data are still lacking. Besides, methods other than Poisson-based frequency models are yet to be explored to bring a new understanding on SV and MV crashes. The objective of this study is mainly to investigate the difference of contributing factors between SV and MV accidents using disaggregate data. To this end, a comprehensive database is first established which includes road geometric features, traffic status, and environmental conditions that are processed on an hourly basis. Then, mixed logit models where the response variable is categorized as no accidents, SV accidents, and MV accidents are employed to address unobserved heterogeneity. Unlike previous aggregated studies that suffer from loss of time-varying information [4], this paper adopts refined-scale panel data to capture the time-varying information as well as to make the short-term prediction. To the authors' knowledge, there is rarely reported study so far on using the mixed logit model to analyze short-term SV and MV accidents risk. By using mixed logit models to examine SV and MV crash risk in short period, this study can address the unobserved heterogeneity and potentially provide new insights regarding the mechanisms of SV and MV crashes.

The remainder of this paper is organized as follows. Previous studies on SV and MV crashes and mixed logit model are briefly summarized in Section 2. In Section 3, a description of data is presented, followed by Section 4, where a detailed explanation of the mixed logit model structure used in this study is outlined. Section 5 presents the model results. Finally, Section 6 summarizes the conclusions and future research directions.

2. Previous Research

2.1. Studies on SV and MV Crashes. SV and MV crashes often refer to different types of accidents. To be specific, SV usually involves run-off-road crash and hitting objects, while MV usually relates to accidents such as rear-end crashes and sideswipes. Therefore, traffic safety researchers have long established that the etiologies of SV and MV crashes are different. Past studies have examined the different features of SV and MV crashes. For example, Mensah and Hauer [9] were among the first to investigate single-vehicle and multivehicle crashes. They developed separate models for SV and two-vehicle crashes and concluded that two separate models for SV and two-vehicle crashes outperformed the model that aggregated SV and two-vehicle crashes together.

Shankar et al. [18] investigated the effects of road geometric design and environmental and seasonal characteristics on SV and MV crashes. Different types of accident data and the various risk factors such as curve number and precipitation from a 61-km highway section were collected. They found that the separate models for SV and MV crashes can better explain the data than the model that pools all the crash data together.

F. Chen and S. Chen [19] examined the injury severities of truck-involved crashes on rural highways based on the distinct models for different crash types. Mixed logit models were used to investigate different risk factors including the driver, temporal, environmental, and roadway characteristics. Their results showed that SV and MV models have their respective contributing factors. The likelihood ratio test was conducted to verify the significance of separate models over the combined model, and the results indicated that separate models are superior.

These past studies have shown that it is beneficial to develop separate models for SV and MV crashes. However, the models adopted in those studies failed to account for the dependence between SV and MV crashes. By developing separate models, possible unobserved effects shared by SV and MV crashes were typically ignored [4]. To account for the dependence between crash types, researchers have proposed multivariate models to study SV and MV accidents jointly. For instance, Yu and Abdel-Aty [17] employed Bayesian bivariate Poisson-lognormal model and hierarchical Poisson models to examine the different characteristic of SV and MV crashes.

Geedipally and Lord [15] investigated the difference of confidence intervals between disaggregated models and the combined model of SV and MV crashes. Five-year crash data on multilane undivided highways were used to develop bivariate Poisson-gamma models for crash prediction. They discovered that the univariate models provide narrower confidence intervals than the bivariate model.

Ma et al. [16] proposed a random effect bivariate Poisson-lognormal model to investigate the effect of geometric features, weather, and traffic conditions on crash occurrence. Their results indicated that the proposed model could address the different levels of correlations between SV and MV crashes.

These abovementioned studies have contributed to the general understanding of SV and MV crashes. Most of these studies adopted Poisson-based models such as Poisson-gamma and Poisson-lognormal models to predict crash frequency. In this study, the difference of single-vehicle and multivehicle crashes will be reexamined from a different perspective. Advanced discrete choice model, that is, mixed logit model, is developed using real-time crash-related information that is processed into hourly records.

2.2. Mixed Logit Model. Over the past two decades, researchers have developed various methods to analyze the risk factors related to traffic crash frequency. Count data models such as Poisson, Negative Binomial, and Poisson-lognormal models are predominantly employed for such purposes [4]. Discrete choice models, on the other hand, are mainly used to investigate injury severity levels. For example, Barua and Tay [20] developed an ordered logit model to study the injury severities of bus crashes in Bangladesh. Xu et al. [21] used spatial logit model to examine the impact of possible risk factor on the injury severity of pedestrians in the crashes which occurred at signalized intersections.

Among various discrete choice models, mixed logit model, that is, the random parameter logit model, has become popular in injury severity studies [19, 22, 23]. It relaxes the independence of irrelevant alternatives assumption for multinomial logit model and offers great capability to capture unobserved heterogeneity in crash data. For instance, Haleem

and Gan [23] developed a mixed logit model to investigate the injury severities of urban freeway crashes in Florida. The role of vehicle types, driver's age, and sides of impact on each injury severity are assessed to unfold their respective effects. Based on the results, two major strategies were suggested to reduce the impacts of adverse factor. Hao and Kamga [24] use ten-year crash data which occurred at highway-rail grade crossings to analyze the effect of lighting on driver injury severities based on mixed logit models. The authors established separate models for lighted intersection and unlighted intersection and found that there are common and different significant attributes for the two situations and suggested that it is necessary to focus more on how drivers react to emergencies at unlighted highway-rail intersections. These studies have all demonstrated the great potential of mixed logit models in crash analyses. By allowing the parameters to vary across observations, mixed logit models enable analysts to discover complex relationships between injury severity and its contributing factors.

2.3. Real-Time Crash Prediction. Despite the preponderance of literature in safety research, most studies are focused on crash predictions on aggregate levels based on yearly records [1, 4, 14, 15, 18, 25]. Those studies employed highly aggregated data, thus being unable to provide guidance for proactive intervention. Real-time crash risk prediction which seeks to identify crash precursor, on the other hand, shows a great appeal for proactive traffic management. It has become a hot topic and been frequently examined by researchers in recent years [17, 26, 27].

However, the literature on real-time crash estimation is not without limitation. To begin with, the real-time safety analysis often requires traffic turbulence measures 5–10 minutes before the crash. Therefore, one key assumption for real-time risk evaluation requires the error of reported crash time to be small. Imprialou and Quddus [28] investigated in detail the quality of police-reported crash data. The results revealed that the reported crash time which ended at zero or five minutes over the course of 1 hour was unproportionally high, with an even higher spike at the thirtieth minute. It is therefore possible that some crashes occurred earlier than the reported time. They concluded that such inaccuracy in reported crash time might significantly compromise the validity of real-time safety studies. Schlögl and Stütz [29] summarized important issues associated with data uncertainty in road safety studies. They pointed out that it is untenable to use time unit smaller than 1 hour due to rounding errors in reported crash time and called for more hourly based studies. Moreover, Roshandel et al. [27] summarized the opportunities and challenges facing real-time risk prediction. They reviewed real-time safety literature and revealed problems such as inconsistent results from different studies and poor predictive performance. Another shortcoming relates to the nature of matched case-control design, which is a major tool for real-time crash prediction. The readers are referred to Roshandel et al. [27] for a detailed discussion.

Given that the aggregated models are incapable of guiding proactive traffic management and that the real-time safety studies suffer from abovementioned shortcomings, this paper tries to find a middle ground that can balance both sides. By employing the crash-related data processed into hourly records, this study can be far less sensitive to the inaccuracy in reported crash time yet still being able to provide short-term (1 hour) crash prediction for proactive traffic management.

3. Data Description

The selected highway stretch is a part of I-25 in Colorado, which starts at Mile Marker 188.49 and ends at Mile Marker 221.03. The overall length of this stretch is 55.93 miles. The data set used in this research mainly consists of the following four sources: (1) one-year detailed crash data obtained from Colorado State Patrol; (2) highway geometric characteristic and pavement condition data obtained from Roadway Characteristic Inventory; (3) refined-scale (in 20-minute intervals) weather and surface condition data from Road Weather Information System; (4) real-time (in 2-minute intervals) traffic data detected by the traffic monitoring stations on I-25.

In previous studies, crash data are usually processed into relatively large time interval. Such aggregation suffers from loss of the time-varying information and leads to estimation bias. To avoid these problems, the crash-related data are processed into relatively short time intervals (one hour) in the current study. The road segments are divided based on the location of traffic stations and further segmented according to the variation of geometric characteristics. For instance, if one of the main characteristics, such as speed limit, changes, this segment will be divided into two new segments. In this way, this study developed 57 road segments, 29 of which are northbound and the others are southbound. The crash data were mapped to roadway segments and processed into hourly records. Then they were matched with the traffic and weather data. The response variable resulted in four possible outcomes: (1) no accident; (2) SV accident; (3) MV accident; and (4) SV mixed with MV accident. However, there were only five out of 328,398 total observations which ended with the fourth outcome (SV mixed with MV). Due to its scarcity, the fourth outcome does not warrant a standalone category in the mixed logit model. Besides, the SV mixed with MV accident resembles the MV accident more than SV accident in terms of etiology. The resulting response variable is therefore defined as three categories, which are no accident, SV accidents, and MV accidents.

There are many detailed geometric variables in the dataset including segment length (miles), number of lanes, number of merging ramps per lane per mile, rutting condition, curvature (degree), and inside shoulder width (feet). Some important traffic control information like speed limit is also collected. There are five weather stations on the study stretch of I-25, which can provide road surface and weather data at a twenty-minute interval. The weather data of each segment are evaluated from the closest station. There are more than 20 monitoring stations that are almost evenly distributed on the road section. The traffic speed and volume data recorded by these stations at 2-minute interval are processed

Table 1: Descriptive statistics of response and significant explanatory variables.

Variable	Mean	Std. dev.	Minimum	Maximum
Crash type	0.008	0.121	0	2
Temporal characteristic				
Night	0.431	na	0	1
Weekend	0.271	na	0	1
November	0.095	na	0	1
Time of sunset	0.062	na	0	1
4 a.m.-5 a.m.	0.040	na	0	1
Traffic characteristic				
Low speed limit indicator (1 if speed limit is less than 55 miles per hour, 0 otherwise)	0.369	na	0	1
Speed gap (speed limit minus traffic speed) (mph)	2.642	5.533	0	69.18
Hourly traffic volume (in thousands per hour)	2.916	2.101	0.03	14.988
Truck percentage (%)	6.215	1.922	2.8	10.7
Road characteristic				
Number of entering ramps per lane per mile	0.252	0.215	0	0.926
Number of lanes	4.159	0.562	3	5
Curvature (degree)	0.947	0.681	0	2.260
length of segment (miles)	1.014	0.769	0.236	4.5
Long remaining service life of rutting (1 if the value of rut is higher than 99, 0 otherwise)	0.225	na	0	1
Inside shoulder width (in feet)	9.006	2.583	5	15
Environmental and road surface characteristic				
Cross wind speed	4.146	3.906	0	31.98
Visibility	1.075	0.136	0	1.1
Wet road surface	0.082	na	0	1
Chemical wet road surface	0.037	na	0	1

into hourly record. More details about how the data were processed can be found in a study by Chen et al. [30]. The descriptive statistics of response and explanatory variables are summarized in Table 1.

4. Methodology

For traditional crash prediction model, the effects of all the explanatory variables were assumed to be fixed across observations. Therefore, the unobserved heterogeneities were ignored. To address the problem, this study adopted mixed logit models to examine the risk factors and their degree of influence on the SV and MV accident. The model structure of mixed logit model is specified in the following section.

Since the data set used in this study is processed to panel data structure with multiple hourly observations from the same roadway segment, the number of all the observations is expressed as N:

$$N = \sum_{j=1}^{J} t_j = \sum_{t=1}^{T} j_t, \tag{1}$$

where t_j means the total number of observations in the site of j; j_t means the total number of observations in the time period of t; J and T mean the number of segments and time periods, respectively. In contrast to previous studies that used cross-sectional data [31], this study adopts panel data structure and specifies the random parameter on road segment level.

Let $P_{n_{jt}}(i)$ be the probability of crash category I (no accident, single-vehicle accident, and multivehicle accident) which occurred on observation n_{jt}:

$$P_{n_{jt}}(i) = P\left(\beta_i X_{n_{jt}} + \varepsilon_{n_{jt}i} \geq \beta_{i'} X_{n_{jt}} + \varepsilon_{n_{jt}i'}\right) \quad \forall i' \in I,\ i' \neq i, \tag{2}$$

where $n_{jt} = 1, \ldots, N$, which means the observation on the jth road segment at the tth hour. I is the set of all the possible crash categories which are mutually exclusive. i and i' are different crash categories. β_i and $\beta_{i'}$ mean the parameter vectors of crash categories i and i'. $X_{n_{jt}}$ is the vector of all the contribution variables for the observation n_{jt}, which have an influence on the possibilities of crash categories i and i'. $\varepsilon_{n_{jt}i}$ and $\varepsilon_{n_{jt}i'}$ are random components (also called error terms) that explain the unobserved effects on crash categories of the jth road segment at tth hour.

Assuming that $\varepsilon_{n_{jt}i}$ follows a type I extreme-value distribution [32], this results in a multinomial logit model which can be defined as

$$P_{n_{jt}}(i) = \frac{e^{\beta_i X_{n_{jt}}}}{\sum_{\forall i' \in I} e^{\beta_{i'} X_{n_{jt}}}}, \quad (3)$$

where the parameter β_i can be estimated using the maximum likelihood method.

The mixed logit model is introduced by relaxing the parameters β_i of the multinomial logit model to be variable across hours t. The distribution of random parameter is specified as follows:

$$\beta_{ki} \sim N\left(\overline{\beta_{ki}}, \sigma_{\beta_{ki}}\right) \quad \text{if } \beta_{ki} \text{ is a random parameter,} \quad (4)$$

where k is the index of explanatory variables for the crash category of level i; β_{ki} is the kth parameter in β_i at crash level i; $N(\overline{\beta_{ki}}, \sigma_{\beta_{ki}})$ means that β_{ki} obeys a normal distribution which varies across different hours; $\overline{\beta_{ki}}$ and $\sigma_{\beta_{ki}}$ are the mean and standard deviation of β_{ki}. In this case, the mixed logit model is specified on a panel data structure where multiple observations are nested within each segment under different hours. The resulting mixed logit model is given as follows:

$$P_{n_{jt}}(i) = \int \frac{e^{\beta_i X_{n_{jt}}}}{\sum_{\forall i' \in I} e^{\beta_i X_{n_{jt}}}} f(\beta \mid \varphi)\, d\beta, \quad (5)$$

where $f(\beta \mid \varphi)$ is the density function of β with parameter vector φ. The likelihood function of mixed logit model was programmed using the NLMIXED procedure in SAS software. In previous studies [19, 22], the normal distribution was found to best fit the data compared to other distributions, including lognormal, triangular, and uniform distribution. Therefore, only normal distribution is considered herein.

5. Results

In this study, the no accident category is chosen as the base category. Hence, the estimated parameters of mixed logit model indicate the difference between the base category (no accident) and the corresponding category (SV accidents or MV accidents). To examine whether it is reasonable to divide the crash type into three categories, models with three crash categories and four crash categories are both established, respectively, for comparing. Detailed model estimation results are summarized in Tables 2 and 3. Many risk factors from different aspects (road geometric, traffic status, and environment characteristics) are shown to have significant influence on the SV and MV crash risk.

AIC (Akaike information criterion) and BIC (Bayesian information criterion), which weigh model fit against model complexity, are used to compare the performance of the two models. The model of three crash categories has relatively lower AIC and BIC than the model of four categories. This result somehow provides empirical evidence that three crash categories are better than four crash categories. Therefore, the following analysis is mainly based on the results shown in Table 2.

Table 2: The estimated results of mixed logit model with three crash categories.

Variable	Estimated parameter	t-statistic
Constant [SV]	−7.255	−16.38
Constant [MV]	−8.608	−9.9
Temporal characteristic		
Weekend [MV]	0.135	2.07
November [MV]	0.322	3.43
4 a.m.-5 a.m. [MV]	−1.698	−2.92
Traffic characteristic		
Low speed limit indicator [MV]	0.654	3.2
Speed gap [SV]	0.045	4.68
Speed gap [MV]	0.084	31.62
Hourly traffic volume [MV]	0.823	12.61
Truck percentage [MV]	−0.056	−1.94
Road characteristic		
Number of entering ramps per lane per mile [MV]	−1.138	−3.04
Curvature [MV]	0.360	2.43
Length of segment [SV]	0.787	6.01
Length of segment [MV]	0.823	5.5
Long remaining service life of rutting [MV]	0.422	2.44
Environmental and road surface characteristic		
Visibility [SV]	−0.821	−2.18
Wet road surface [SV]	0.676	2.56
Wet road surface [MV]	−0.330	−2.04
Chemical wet road surface [MV]	0.429	3.01
Random parameters		
Mean of log of hourly traffic volume [MV]	0.823	12.61
Std. dev. of log of hourly traffic volume (normal distribution)	0.160	2.29
Mean of inside shoulder width [MV]	0.032	0.95
Std. dev. of inside shoulder width (normal distribution)	0.047	5.89
Mean of wet road surface [MV]	−0.330	−2.04
Std. dev. of wet road surface (normal distribution)	0.608	3.35
Fit statistics		
−2 log likelihood	15476	
AIC (smaller is better)	15532	
BIC (smaller is better)	15589	

According to model estimation results, three explanatory variables are found to be better treated as random parameters (significant at the level of 95% with t-statistics 2.29, 5.89, and 3.35, respectively). From Table 2, the parameter associated with the hourly traffic volume of MV crash category is normally distributed with mean 0.8228 and standard deviation

Table 3: The estimated results of mixed logit model with four crash categories.

Variable	Estimated parameter	t-statistic
Constant [SV]	−7.254	−16.38
Constant [MV]	−8.587	−9.79
Constant [SV and MV]	−12.398	−16.62
Temporal characteristic		
Weekend [MV]	0.134	2.05
November [MV]	0.319	3.35
4 a.m.-5 a.m. [MV]	−1.678	−2.89
Traffic characteristic		
Low speed limit indicator [MV]	0.653	3.17
Speed gap [SV]	0.045	4.68
Speed gap [MV]	0.083	31.29
Speed gap [SV and MV]	0.140	5.56
Hourly traffic volume [MV]	0.822	10.92
Truck percentage [MV]	−0.058	−2.01
Road characteristic		
Number of entering ramps per lane per mile [MV]	−1.140	−3.02
Curvature [MV]	0.363	2.43
Length of segment [SV]	0.786	6
Length of segment [MV]	0.812	5.39
Long remaining service life of rutting [MV]	0.427	2.44
Environmental and road surface characteristic		
Visibility [SV]	−0.822	−2.18
Wet road surface [SV]	0.674	2.55
Wet road surface [MV]	−0.342	−2.1
Chemical wet road surface [MV]	0.419	2.91
Random parameters		
Mean of log of hourly traffic volume [MV]	0.823	10.92
Std. dev. of log of hourly traffic volume (normal distribution)	0.162	2.3
Mean of inside shoulder width [MV]	0.032	0.91
Std. dev. of inside shoulder width (normal distribution)	0.047	5.91
Mean of wet road surface [MV]	−0.342	−2.10
Std. dev. of wet road surface (normal distribution)	0.614	3.38
Fit statistics		
−2 log likelihood	15536	
AIC (smaller is better)	15598	
BIC (smaller is better)	15661	

0.1602. The distribution of hourly traffic volume of MV crash is shown in Figure 1. This indicates the increase of hourly traffic volume will raise the possibility of multivehicle crash in nearly all (99.99%) of the road segments. This result is consistent with people's perceptions/experience and is in line with previous studies [16, 33]. Besides, the random parameter indicates that the impact of hourly traffic volume on MV crash possibility is different across road segments. The same variable is not significant in SV crash category, which means hourly traffic volume has no significant effect on single-vehicle crashes. The difference of significant parameters reveals that some essential differences do exist between single-vehicle crashes and multivehicle crashes.

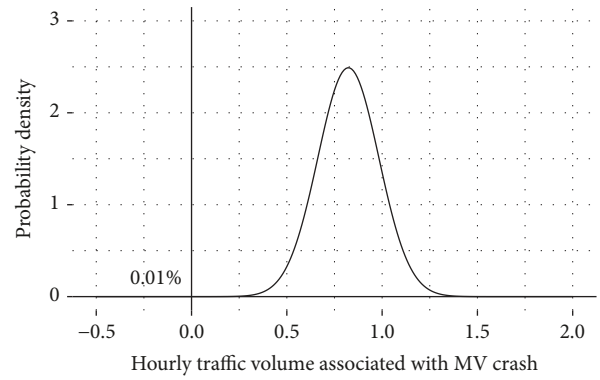

Figure 1: Distribution of hourly traffic volume associated with MV crash.

Inside shoulder width is also found to have random effect on different road segments for MV accidents. According to the results, the estimated mean of the parameter is not statistically significant, which may be considered as a problem. Nevertheless, a recent study by Behnood and Mannering [34] pointed out that when the standard deviation of a random parameter is significant, the mean of the random parameter does not need to be significant as well. From Table 2, the parameter of the inside shoulder width of MV crash category is normally distributed with mean 0.0318 and standard deviation 0.0472. The corresponding distribution is shown in Figure 2, which shows that 75% of the distribution is greater than zero and 25% is less than zero. This means that wider inside shoulder is associated with higher probability of MV accidents in 75% of the road segments and lower probability of MV accidents in the rest 25% of the road segments. This finding can possibly be explained by the tradeoff between forgiving geometric design and risky driving. On the one hand, wider inside shoulder tolerates more driver errors. On the other hand, it is possible that when inside shoulder exceeds a certain threshold, drivers may be more likely to take risky actions such as passing and speeding, according to risk compensation theory [35].

The parameter of the wet road surface of MV crash category is normally distributed with mean −0.3303 and standard deviation 0.6076. As shown in Figure 3, 29% of the distribution is greater than zero and 71% is less than zero. When the road surface gets wet, nearly three-quarters of the hours are related to lower likelihood of MV accidents, while the other hours are related to higher risk of MV accidents. Moreover, wet road surface is positively correlated to SV accidents, which means that wet road surface usually leads to more SV accidents. Such phenomena may be caused by some

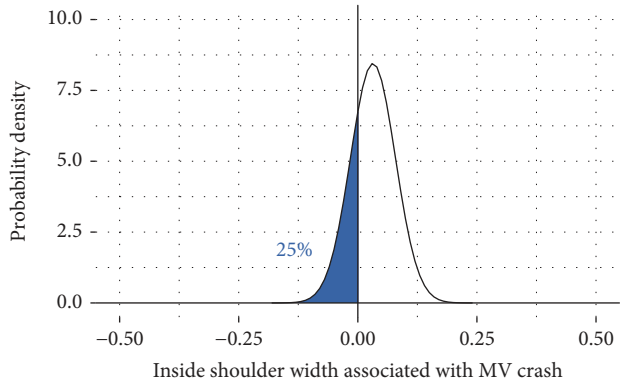

Figure 2: Distribution of insider shoulder width associated with MV crash.

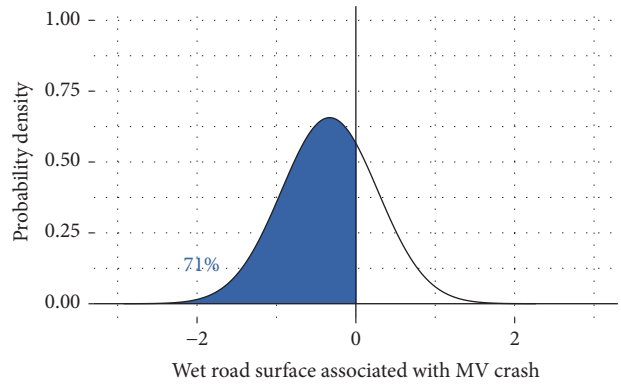

Figure 3: Distribution of wet road surface associated with MV crash.

unobserved heterogeneity of driver behavior. On wet road surface, the skid resistance decreases which increases crash risks. At the same time, wet road surface is often concurrent with rainy condition when drivers tend to maintain a longer distance between vehicles. The crash risk is therefore a result of increased driver attentiveness and reduced skid resistance. As a result, wet road surface leads to mixed effects on MV crash.

5.1. Temporal Characteristics. It can be found from Table 2 that the temporal factors do not have a strong impact on the possibility of SV accidents. As for MV accidents, the results indicate that the MV crashes are more likely to occur on weekends and less likely to occur during 4 a.m. to 5 a.m. Compared to other months, the likelihood of MV crashes is larger in November. This may be associated with some harmful impacts caused by the sudden storm and temperature change in November, 2010 [30].

5.2. Traffic Characteristic. Speed limit indicator is used to evaluate the effect of speed limit on traffic safety. This study uses a dummy indicator to express the speed limit. If the legal speed limit is smaller than 55, then the value equals 1, otherwise equals 0. The results show that low speed limit will increase the possibility of multivehicle crashes. Some researchers also found that low speed limit will increase crash possibility [36], but they failed to reveal its different effects on SV and MV accidents.

The speed gap variable denotes the difference between the average speed and the speed limit, which can represent the congestion level to some extent. As shown in Table 2, the results indicate that both SV and MV vehicle crashes are more likely to occur when the speed gap gets larger. This finding is partially similar to those in some previous studies [17]. In addition, the increase of truck ratio will decrease the likelihood of multivehicle crashes, which is in accordance with the conclusions drawn by Anastasopoulos and Mannering [37].

5.3. Road and Pavement Characteristic. Several road characteristics are found to have significant influence on SV and MV crash risks. The length of road segment tends to increase the likelihood of both SV and MV crash, which is consistent with the research by Venkataraman et al. [38]. More merging ramps per lane per mile will decrease the likelihood of multivehicle crashes, which may be attributed to the careful driving behavior on roads consisting of more merging ramps. The same indicator has also been investigated in previous studies. Pei et al. [39] also found that an increase in the merging and diverging ramp number leads to fewer accidents, while some other researchers [37] made opposite conclusions. The difference of their conclusion may be due to their aggregate model structure, which does not consider the different mechanism between SV and MV accidents. This inconsistency among past findings points to the very need to investigate SV and MV crashes separately and uncover their respective risk factors.

A similar result is also found for the curvature variable. In this study, the results imply that higher curvature will cause higher multivehicle crash risks. Albeit some researchers found that the degree of curvature can be beneficial for traffic safety [17, 39], other researchers found it positively correlated with the crash likelihood [36, 40].

As for pavement conditions, it can be found that the possibility of multivehicle crashes will decrease on segments with longer remaining service live of rut. This is probably because drivers have a tendency to drive carefully on the road with deeper rut and is consistent with past studies [30].

5.4. Environmental and Surface Characteristic. Turning to surface characteristic, wet road surface and chemical wet road surface are both shown to be associated with increased possibility of single-vehicle crash. Chemical wet road surface leads to more multivehicle crashes, while the effect of wet road surface will change across road segments because of its random nature, which has been discussed above.

From Table 2, higher visibility is related to decreased possibility of single-vehicle crashes. According to the results, other environment characteristics, such as cross wind, temperature, and humidity, are not significant. This phenomenon is plausible because the selected stretch of I-25 is relatively flat and spans across Denver metro. On highways located at mountainous terrains subjected to complex weather, the

environment characteristics may impose significant influence on traffic safety.

6. Discussion and Conclusion

In this study, mixed logit models are developed to examine the difference of SV and MV accident probability using hourly based disaggregated crash data. One-year accident data, detailed traffic data, weather condition, road geometry, and surface condition data on I-25 from the state of Colorado were collected to establish a refined-scale panel data structure. The refined-scale is used to capture the potential information lost in aggregate data. Many risk factors are found to have varying effects on SV and MV accident probability.

These findings contribute to the literature on the risk factors that are associated with the different mechanisms of SV and MV crash. Different from the former researches which model the SV and MV accidents frequencies separately or jointly, this study uses mixed logit model to study the risk of both SV and MV crashes. Therefore, the results of this paper can provide guidance to develop more rational and effective segment management measures and accident prevention strategies. In addition, some findings are also helpful for the evaluation and improvement of the designs of existing transportation infrastructure.

The main conclusions associated with the risk factors and their different effects on SV and MV accident are summarized as follows.

(1) Speed gap, length of segment, and wet road surface are found to have significant effects on both SV and MV accident possibility. In addition to these indicators, most of other variables including time of weekends, November, low speed limit indicator, hourly traffic volume, truck percentage, and chemical wet road surface are significant for only MV accidents. Visibility indicator is significant for only SV accidents.

(2) For I-25, the main influence factors of SV and MV accident possibility are traffic, road, and surface characteristics. Other temporal and environment characteristics like weekends, special period, and visibility also have certain effects on the possibility of SV and MV accidents, respectively.

(3) The model results indicate that hourly traffic volume, inside shoulder width, and wet road surface are random parameters with normal distribution for multivehicle crash probability. So the impacts of these variables on MV accident are proved to be different across road segments.

Without doubt, there are also some limitation existing in the present study. The conclusions conducted here are mainly based on the data from part of I-25, which may be not suitable for other highways. In order to get more precise and universal rules on the possibility of SV and MV accidents, further studies should be conducted on different types of highways.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant nos. 51508409 and 71531011, the Program for Young Excellent Talents in Tongji University, and the Fundamental Research Funds for the Central Universities. The authors would like to express their gratitude to the United States Department of Transportation and Colorado Department of Transportation and Colorado State Patrol for the detailed data.

References

[1] C. Y. Chan, B. Huang, X. Yan, and S. Richards, "Investigating effects of asphalt pavement conditions on traffic accidents in Tennessee based on the pavement management system (PMS)," *Journal of Advanced Transportation*, vol. 44, no. 3, pp. 150–161, 2010.

[2] J. Weng and Q. Meng, "Rear-end crash potential estimation in the work zone merging areas," *Journal of Advanced Transportation*, vol. 48, no. 3, pp. 238–249, 2014.

[3] S. M. Rifaat and H. C. Chin, "Accident severity analysis using ordered probit model," *Journal of Advanced Transportation*, vol. 41, no. 1, pp. 91–114, 2007.

[4] D. Lord and F. Mannering, "The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives," *Transportation Research Part A: Policy and Practice*, vol. 44, no. 5, pp. 291–305, 2010.

[5] R. Tay and S. M. Rifaat, "Factors contributing to the severity of intersection crashes," *Journal of Advanced Transportation*, vol. 41, no. 3, pp. 244–265, 2007.

[6] L. Mussone and K. Kim, "The analysis of motor vehicle crash clusters using the vector quantization technique," *Journal of Advanced Transportation*, vol. 44, no. 3, pp. 162–175, 2010.

[7] T. Jonsson, J. N. Ivan, and C. Zhang, "Crash prediction models for intersections on rural multilane highways: Differences by collision type," *Transportation Research Record*, no. 2019, pp. 91–98, 2007.

[8] R. R. Knipling, "Car-truck crashes in the national motor vehicle crash causation study," in *Proceedings of the Transportation Research Board (TRB) 92nd Annual Meeting*, vol. 13, 2013.

[9] A. Mensah and E. Hauer, "Two problems of averaging arising in the estimation of the relationship between accidents and traffic flow," *Transportation Research Record*, vol. 1635, pp. 37–43, 1998.

[10] X. Ye, R. M. Pendyala, S. P. Washington, K. Konduri, and J. Oh, "A simultaneous equations model of crash frequency by collision type for rural intersections," *Safety Science*, vol. 47, no. 3, pp. 443–452, 2009.

[11] J. Ivan, "New approach for including traffic volumes in crash rate analysis and forecasting," *Transportation Research Record*, vol. 1897, pp. 134–141, 2004.

[12] M. Griffith, "Safety evaluation of rolled-in continuous shoulder rumble strips installed on freeways," *Transportation Research Record*, vol. 1665, no. 1, pp. 28–34, 1999.

[13] F. L. Mannering and C. R. Bhat, "Analytic methods in accident research: methodological frontier and future directions," *Analytic Methods in Accident Research*, vol. 1, pp. 1–22, 2014.

[14] S. R. Geedipally and D. Lord, "Identifying hot spots by modeling single-vehicle and multivehicle crashes separately," *Transportation Research Record*, no. 2147, pp. 97–104, 2010.

[15] S. R. Geedipally and D. Lord, "Investigating the effect of modeling single-vehicle and multi-vehicle crashes separately on confidence intervals of Poisson-gamma models," *Accident Analysis & Prevention*, vol. 42, no. 4, pp. 1273–1282, 2010.

[16] X. Ma, S. Chen, and F. Chen, "Correlated random-effects bivariate poisson lognormal model to study single-vehicle and multivehicle crashes," *Journal of Transportation Engineering*, vol. 142, no. 11, 2016.

[17] R. Yu and M. Abdel-Aty, "Multi-level Bayesian analyses for single- and multi-vehicle freeway crashes," *Accident Analysis & Prevention*, vol. 58, pp. 97–105, 2013.

[18] V. Shankar, F. Mannering, and W. Barfield, "Effect of roadway geometrics and environmental factors on rural freeway accident frequencies," *Accident Analysis & Prevention*, vol. 27, no. 3, pp. 371–389, 1995.

[19] F. Chen and S. Chen, "Injury severities of truck drivers in single- and multi-vehicle accidents on rural highways," *Accident Analysis & Prevention*, vol. 43, no. 5, pp. 1677–1688, 2011.

[20] U. Barua and R. Tay, "Severity of urban transit bus crashes in Bangladesh," *Journal of Advanced Transportation*, vol. 44, no. 1, pp. 34–41, 2010.

[21] X. Xu, S. Xie, S. C. Wong, P. Xu, H. Huang, and X. Pei, "Severity of pedestrian injuries due to traffic crashes at signalized intersections in Hong Kong: a Bayesian spatial logit model," *Journal of Advanced Transportation*, vol. 50, no. 8, pp. 2015–2028, 2016.

[22] J. C. Milton, V. N. Shankar, and F. L. Mannering, "Highway accident severities and the mixed logit model: an exploratory empirical analysis," *Accident Analysis & Prevention*, vol. 40, no. 1, pp. 260–266, 2008.

[23] K. Haleem and A. Gan, "Effect of driver's age and side of impact on crash severity along urban freeways: A mixed logit approach," *Journal of Safety Research*, vol. 46, pp. 67–76, 2013.

[24] W. Hao and C. Kamga, "The effects of lighting on driver's injury severity at highway-rail grade crossings," *Journal of Advanced Transportation*, vol. 50, no. 4, pp. 446–458, 2016.

[25] C. Xu, P. Liu, W. Wang, and Y. Zhang, "Real-time identification of traffic conditions prone to injury and non-injury crashes on freeways using genetic programming," *Journal of Advanced Transportation*, vol. 50, no. 5, pp. 701–716, 2016.

[26] M. Abdel-Aty, N. Uddin, A. Pande, F. Abdalla, and L. Hsia, "Predicting freeway crashes from loop detector data by matched case-control logistic regression," *Transportation Research Record*, vol. 1897, no. 1, pp. 88–95, 2004.

[27] S. Roshandel, Z. Zheng, and S. Washington, "Impact of real-time traffic characteristics on freeway crash occurrence: Systematic review and meta-analysis," *Accident Analysis & Prevention*, vol. 79, pp. 198–211, 2015.

[28] M. Imprialou and M. Quddus, "Crash data quality for road safety research: Current state and future directions," *Accident Analysis & Prevention*, 2016.

[29] M. Schlögl and R. Stütz, "Methodological considerations with data uncertainty in road safety analysis," *Accident Analysis & Prevention*, 2016.

[30] F. Chen, X. Ma, and S. Chen, "Refined-scale panel data crash rate analysis using random-effects tobit model," *Accident Analysis & Prevention*, vol. 73, pp. 323–332, 2014.

[31] F. L. Mannering, V. Shankar, and C. R. Bhat, "Unobserved heterogeneity and the statistical analysis of highway accident data," *Analytic Methods in Accident Research*, vol. 11, pp. 1–16, 2016.

[32] D. McFadden and K. Train, "Mixed MNL models for discrete response," *Journal of Applied Econometrics*, vol. 15, no. 5, pp. 447–470, 2000.

[33] Y. Qi, B. L. Smith, and J. Guo, "Freeway accident likelihood prediction using a panel data analysis approach," *Journal of Transportation Engineering*, vol. 133, no. 3, pp. 149–156, 2007.

[34] A. Behnood and F. Mannering, "The effect of passengers on driver-injury severities in single-vehicle crashes: A random parameters heterogeneity-in-means approach," *Analytic Methods in Accident Research*, vol. 14, pp. 41–54, 2017.

[35] C. Winston, V. Maheshri, and F. Mannering, "An exploration of the offset hypothesis using disaggregate data: the case of airbags and antilock brakes," *Journal of Risk and Uncertainty*, vol. 32, no. 2, pp. 83–90, 2006.

[36] Y. Kweon and K. Kockelman, "Safety effects of speed limit changes: use of panel models, including speed, use, and design variables," *Transportation Research Record*, vol. 1908, pp. 148–158, 2005.

[37] P. C. Anastasopoulos and F. L. Mannering, "A note on modeling vehicle accident frequencies with random-parameters count models," *Accident Analysis & Prevention*, vol. 41, no. 1, pp. 153–159, 2009.

[38] N. S. Venkataraman, G. F. Ulfarsson, and V. N. Shankar, "Random parameter models of interstate crash frequencies by severity, number of vehicles involved, collision and location type," *Accident Analysis & Prevention*, vol. 59, pp. 309–318, 2013.

[39] X. Pei, S. C. Wong, and N. N. Sze, "The roles of exposure and speed in road safety analysis," *Accident Analysis & Prevention*, vol. 48, pp. 464–471, 2012.

[40] L. Ewan, A. Al-Kaisy, and F. Hossain, "Safety effects of road geometry and roadside features on low-volume roads in Oregon," *Transportation Research Record*, vol. 2580, pp. 47–55, 2016.