Journal of Safety Research 46 (2013) 91–97

Contents lists available at SciVerse ScienceDirect

Journal of Safety Research

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

Investigating the different characteristics of weekday and weekend crashes

Rongjie Yu [a,b,*], Mohamed Abdel-Aty [a]

[a] Department of Civil, Environmental and Construction Engineering, University of Central Florida, Orlando, FL 32816-2450, USA
[b] School of Transportation Engineering, Tongji University, 4800 Cao'an Road, 201804 Shanghai, China

ARTICLE INFO

Article history:
Received 14 December 2012
Received in revised form 19 April 2013
Accepted 20 May 2013
Available online 4 June 2013

Keywords:
Multivariate Poisson model
Weekend crash
Random effects Bayesian logistic regression
Mountainous freeway traffic safety

ABSTRACT

Introduction: This study provides a systematic approach to investigate the different characteristics of weekday and weekend crashes. *Method:* Weekend crashes were defined as crashes occurring between Friday 9 p.m. and Sunday 9 p.m., while the other crashes were labeled as weekday crashes. In order to reveal the various features for weekday and weekend crashes, multi-level traffic safety analyses have been conducted. For the aggregate analysis, crash frequency models have been developed through Bayesian inference technique; correlation effects of weekday and weekend crash frequencies have been accounted. A multivariate Poisson model and correlated random effects Poisson model were estimated; model goodness-of-fits have been compared through DIC values. In addition to the safety performance functions, a disaggregate crash time propensity model was calibrated with Bayesian logistic regression model. Moreover, in order to account for the cross-section unobserved heterogeneity, random effects Bayesian logistic regression model was employed. *Results:* It was concluded that weekday crashes are more probable to happen during congested sections, while the weekend crashes mostly occur under free flow conditions. Finally, for the purpose of confirming the aforementioned conclusions, real-time crash prediction models have been developed. Random effects Bayesian logistic regression models incorporating the microscopic traffic data were developed. Results of the real-time crash prediction models are consistent with the crash time propensity analysis. Furthermore, results from these models would shed some lights on future geometric improvements and traffic management strategies to improve traffic safety. *Impact on Industry:* Utilizing safety performance to identify potential geometric improvements to reduce crash occurrence and monitoring real-time crash risks to pro-actively improve traffic safety.

1. Introduction

Motor-vehicle crashes have drawn great attention from both the public and research sides in recent decades. According to the National Highway Traffic Safety Administration (NHTSA), in 2010, 32,885 people were killed in traffic crashes in the United States (NHTSA, 2012). Research studies have been conducted to reveal the hazardous factors related to crash occurrence. Conclusions from these studies were utilized in geometry design and traffic management fields to develop countermeasures to improve traffic safety. In addition, with the merits of advanced traffic surveillance systems and weather reporting systems, researchers are able to identify crash occurrence contributing factors at the microscopic level.

This study focuses on a mountainous freeway (I-70) in Colorado. The freeway serves commuters during weekdays, and mostly connects the ski resorts and Denver on weekends. Considering the diverse driving behaviors of weekday and weekend travelers, different levels of traffic safety analysis models were established. Firstly, safety performance functions (SPFs) were established with a correlated random effects Poisson model and a multivariate Poisson (MVP) model. Distinct sets of crash occurrence contributing factors have been obtained and the two models were compared based on DIC values. In addition to aggregate crash frequency models, this study also focused on disaggregate traffic safety analysis. Crash occurrence mechanisms of the weekday and weekend crashes were revealed with the benefits of real-time traffic data. Differences between weekday and weekend crashes were analyzed through a crash time propensity model, which followed by two real-time crash prediction models for weekday and weekend crashes separately. Finally, results of these analyses shed some lights on future traffic management strategies and geometric design improvements to reduce crash occurrence.

2. Background

2.1. Crash frequency models

Poisson regression models were first introduced to analyze crash frequency data due to their non-negative integers' characteristics (Jovanis & Chang, 1986). Researchers benefited from the Poisson

* Corresponding author at: Department of Civil, Environmental and Construction Engineering, University of Central Florida Orlando, Florida 32816-2450, USA. Tel.: +1 407 823 0300.
E-mail address: rongjie.yu@gmail.com (R. Yu).

0022-4375/$ – see front matter
http://dx.doi.org/10.1016/j.jsr.2013.05.002

regression models of the easy estimations and straightforward explanations. However, these models were blamed for lacking the ability of handling over-dispersion problems. Extensions of the Poisson regression models, such as random effects Poisson models (also called negative binomial models) (Ahmed, Huang, Abdel-Aty, & Guevara, 2011; Guo, Wang, & Abdel-Aty, 2010; Miaou & Lord, 2003; Shankar, Albin, Milton, & Mannering, 1998; Yu, Abdel-Aty, & Ahmed, 2013) and Poisson-gamma models (Lord, 2006; Lord & Miranda-Moreno, 2008; Schlüter, Deely, & Nicholson, 1997) were employed to analyze crash frequency data.

The abovementioned studies assumed a log-linear relationship between crash frequency and the explanatory variables. Other studies, such as Kononov, Lyon, and Allery (2011), have identified sigmoid or other non-linear relationships to be more appropriate approximation for crash frequency and the explanatory variables. However, these non-linear functions are quite complex and difficult to estimate (Lord & Mannering, 2010).

In addition, crash frequency data were typically analyzed by severity levels and crash types. In order to consider the cross-model correlation for the dependent variables, multivariate regression models were employed. Ye, Pendyala, Washington, Konduri, and Oh (2009) proposed using multivariate Poisson regression model to simultaneously analyze different crash types. Common unobserved heterogeneity have been accounted for through the error covariance. Furthermore, multivariate Poisson-lognormal (MVPLN) models were frequently adopted to analyze correlated count variables to consider the over-dispersion issue. Park and Lord (2007) utilized MVPLN model to analyze crash frequency by severity. The MVPLN model was argued to be capable of accounting for over-dispersion; the general correlation structure is superior to the MVP model. El-Basyouny and Sayed (2009) also employed the MVPLN model to jointly investigate crash frequency by severity levels. MVPLN model was compared to the univariate models regarding the goodness-of-fit and hazardous location identification. Results demonstrated that MVPLN model is better than the univariate models.

Apart from the diverse model formulations utilized to analyze crash frequency data, Bayesian inference technique has become popular recently in traffic safety studies. Compared to the conventional frequentist inference approach, the Bayesian inference method provides a complete and coherent way to balance the empirical data and prior expectations, which would enhance the model's goodness-of-fit (Yu & Abdel-Aty, 2013). Another merit of the Bayesian inference approach is the advantage to deal with multilevel traffic safety data (Huang & Abdel-Aty, 2010). For example, Guo et al. (2010) developed Bayesian models to analyze spatial correlation effects of crash frequency between neighboring intersections. In this study, the Bayesian inference technique was utilized to develop a MVP model and correlated random effects Poisson model for weekday and weekend crash frequency data.

2.2. Real-time crash prediction models

Real-time crash prediction models were estimated with the purpose of identifying crash precursors. With advanced traffic surveillance system (loop detectors, speed radars, and automatic vehicle identification systems), traffic states prior to the crash occurrence could be achieved. Based on the crash risk prediction models, Variable Speed Limit (VSL) systems were developed to test the effectiveness of Advanced Traffic Management System (ATMS) on improving traffic safety (Abdel-Aty et al., 2007).

In the real-time crash prediction models, crash occurrence was analyzed as a binary variable with outcomes of crash and non-crash. Regarding the binary dependent variable, a variety of discrete choice models were utilized. The most frequently adopted approach is to develop logistic regression models (Abdel-Aty, Uddin, Pande, Abdalla, & Hsia, 2004; Ahmed & Abdel-Aty, 2011); results from this kind of model are easy to be interpreted. Additionally, other non-linear models like neural network models (Pande & Abdel-Aty, 2006a, 2006b; Pande, Das, Abdel-Aty, & Hassan, 2011) and support vector machine model (Yu & Abdel-Aty, 2013) have also been introduced. Machine learning modeling techniques provided better model goodness-of-fits compared to the logistic regression models. However, the relationships between crash occurrence and the explanatory variables are hard to detect. Considering the ordinary logistic regression models failed to account for the heterogeneity across the crashes, in this study, random effects logistic regression models were utilized to account for the heterogeneity across segments and analyze crash likelihood.

2.3. Random effects logistic regression model

Hierarchical logit models were frequently used to account for the unobserved heterogeneity. Huang, Chin, and Haque (2008) introduced hierarchical Bayesian logistic regression models to perform the multi-vehicle crash injury severity analysis. By considering driver-vehicle units' correlations in multi-vehicle crashes and the unobserved heterogeneity at the crash units' level, a better model fit was achieved. Moreover, by calculating the intra-class correlation coefficient, it was concluded that 28.9% of the unexplained variations resulting from between-crash variance were accounted. In addition, Huang et al. (2010) utilized a multilevel ordered logistic model to analyze crash injury severity for the fog and smoke related crashes in the Central Florida area. Random effects were introduced to accommodate the cross-segment heterogeneities. Recently, Yu and Abdel-Aty (2013) employed random parameter logistic regression models to develop crash risk evaluation models. As a benefit from the ability of hierarchical models to account for unobserved heterogeneity, better goodness-of-fits have been achieved.

Besides the traffic safety studies, random effects logistic regression models have been used in many studies of other fields. Mok, Sohn, and Ju (2010) employed a random effects logistic regression model to perform anomaly detection. High classification accuracy has been achieved since the model considers not only the usual explanatory variables, but also the uncertainty. Sohn and Kim (2007) utilized random effects logistic regression models for default prediction, a higher classification accuracy was acquired by comparing the random effects logistic regression model to the ordinal logistic regression model. Singh, Bartolucci, and Bae (2010) applied the Bayesian random effects logistic regression model in the area of clinical trial. The random effects were assigned to each sample to account for the unobserved heterogeneity. Moreover, according to Gibbons and Hedeker (1997), random effects logistic regression models can be extremely useful in the analysis of discrete clustered multivariate data.

3. Data preparation

A 15-mile mountainous freeway on I-70 in Colorado was chosen as the study area. This freeway section starts from the Mile Marker (MM) 205 and ends at MM 220. Different data sets have been organized and prepared for the aggregate and disaggregate analyses.

3.1. Aggregate analysis

For the crash frequency models, two data sets were utilized in this study: (a) crash data (from Jan 2006 to Apr 2011) provided by the Colorado Department of Transportation (CDOT) and; (b) roadway geometric characteristics data obtained from the Roadway Characteristics Inventory (RCI). The 15-mile section has been split into 120 homogenous segments (60 in each direction); detailed homogenous segmentation method can be found in a previous study (Ahmed et al., 2011).

A total of 1,239 crashes were documented within the studied period; crashes have been assigned to each homogenous segment according to the MM. Crashes that happened between Friday 9 p.m. and Sunday 9 p.m. were labeled as weekend crashes while the other crashes