Accident Analysis and Prevention xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Crash risk analysis for Shanghai urban expressways: A Bayesian semi-parametric modeling approach

Rongjie Yu[a,b,c], Xuesong Wang[a,b,c,*], Kui Yang[a,b,c], Mohamed Abdel-Aty[d]

[a] The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Shanghai 201804, China
[b] Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, SiPaiLou #2, Nanjing 210096, China
[c] College of Transportation Engineering, Tongji University, 4800 Cao'an Road, Shanghai 201804, China
[d] Department of Civil, Environmental and Construction Engineering, University of Central Florida Orlando, FL 32826-2450, United States

ARTICLE INFO

Article history:
Received 31 January 2015
Received in revised form 9 November 2015
Accepted 22 November 2015
Available online xxx

Keywords:
Bayesian inference
Semi-parametric model
Crash risk analysis
Random effects logistic regression model

ABSTRACT

Urban expressway systems have been developed rapidly in recent years in China; it has become one key part of the city roadway networks as carrying large traffic volume and providing high traveling speed. Along with the increase of traffic volume, traffic safety has become a major issue for Chinese urban expressways due to the frequent crash occurrence and the non-recurrent congestions caused by them. For the purpose of unveiling crash occurrence mechanisms and further developing Active Traffic Management (ATM) control strategies to improve traffic safety, this study developed disaggregate crash risk analysis models with loop detector traffic data and historical crash data. Bayesian random effects logistic regression models were utilized as it can account for the unobserved heterogeneity among crashes. However, previous crash risk analysis studies formulated random effects distributions in a parametric approach, which assigned them to follow normal distributions. Due to the limited information known about random effects distributions, subjective parametric setting may be incorrect. In order to construct more flexible and robust random effects to capture the unobserved heterogeneity, Bayesian semi-parametric inference technique was introduced to crash risk analysis in this study. Models with both inference techniques were developed for total crashes; semi-parametric models were proved to provide substantial better model goodness-of-fit, while the two models shared consistent coefficient estimations. Later on, Bayesian semi-parametric random effects logistic regression models were developed for weekday peak hour crashes, weekday non-peak hour crashes, and weekend non-peak hour crashes to investigate different crash occurrence scenarios. Significant factors that affect crash risk have been revealed and crash mechanisms have been concluded.

1. Introduction

With the rapid development of Chinese cities, residents' trip lengths have significantly increased while the trip modes have changed substantially from non-motorized to motorized. Along with the longer and motorized trips, the fast increase of vehicle ownership has resulted in more frequent traffic congestions. For the purpose of solving the congestion issue and providing higher traveling speeds, starting from the 1990s, Chinese cities began to construct the urban expressway systems. As defined in the design standards, urban expressways featured controlled access, more than four lanes (two directions), and divided (Ministry of Construction, 1990). By the year of 2011, Shanghai had a total mileage of 193 km urban expressways in operation.

Along with the fast development of urban expressways in China, traffic volume on urban expressways increased dramatically. In 2011, the urban expressways in Shanghai have taken up 38.5% of the total traffic (only 5% of the total roadway mileage) (Shanghai Urban Construction and Transportation Commission, 2011). In addition, according to the Roadway Traffic Accident Statistical Annual Report (Ministry of Public Security Traffic Management Bureau, 2007), crashes occurred on urban expressways have accounted for 14.51%, 17.66%, and 14.23% of the annual roadway traffic crashes with respective to total, fatal, and injury crashes, respectively in 2006. Considering the large traffic demand and frequent crash occurrence, it is essential to provide smoother and safer driving conditions for urban expressway users.

Metropolitans in the US and Europe recently utilize Active Traffic Management (ATM) Systems to manage their urban expressways proactively. However, in order to effectively prevent crash occurrence through control strategies such as variable speed limits (VSL) and queue warning (QW), it is vital to understand the

* Corresponding author.
E-mail address: wangxs@tongji.edu.cn (X. Wang).

http://dx.doi.org/10.1016/j.aap.2015.11.029

crash mechanisms on urban expressways. For the purpose of investigating crash contributing factors on urban expressways in Shanghai, this study conducts disaggregate traffic safety analyses based on historical crash data and traffic flow data captured by loop detectors. Crash risk analysis models were estimated for total crashes, weekday peak/non-peak hour crashes, and weekend non-peak hour crashes; Bayesian random effects logistic regression models were utilized to identify the relationships between traffic flow characteristics and crash risk.

One key part of the Bayesian random effects logistic regression models is the inference of the random effects term. Although previous studies focused on developing safety performance functions using random parameter models have tested various random effects distribution (Tulu et al., 2015; Xu and Huang, 2015), previous crash risk analysis studies subjectively assigned normal distributions with vague prior information (zero means and large variances) for random effects, which may be incorrect or unable to capture the unobserved heterogeneity. In order to construct more flexible and robust random effects distributions, Bayesian semi-parametric inference approach was adopted in this study. Modeling results and goodness-of-fits were compared between the parametric setting and semi-parametric approach.

The following parts of the paper were divided into five sections. Firstly, previous studies related to crash risk analysis and its modeling techniques, such as Bayesian logistic regression models and Bayesian semi-parametric models, were discussed. Section 3 provides a brief description of the data preparation procedures, followed by a description of the methodologies employed in this study. Section 5 presents the modeling results and discussion about the crash mechanisms on urban expressways at different time periods. Finally, summaries and conclusions of the work were given.

2. Background

2.1. Crash risk analysis

Crash risk analysis was developed to identify crash occurrence precursors through investigating different traffic patterns between crash and non-crash scenarios; different methods have been utilized to formulate them. Among the various analysis methods, logistic regression models were the most frequently adopted one. Abdel-Aty et al. (2004) employed the matched case–control logistic regression modeling technique to analyze crash risk based on traffic flow data from loop detectors on I-4 in Orlando, Florida. Later on, Abdel-Aty and Pemmanaboina (2006) further incorporated weather data along with traffic data to assess the real-time crash risk on the same roadway with logistic regression models. Xu et al. (2012) utilized conditional logistic regression models with traffic data from I-880 in California to identify the traffic flow characteristics' effects on crash risk. Apart from the logistic regression models, data mining analysis methods such as neural network (2006) and support vector machine (Yu and Abdel-Aty, 2013a; Sun et al., 2014) were also employed to analyze crash risk. Although data mining approaches provided comparative or even superior goodness-of-fits as the logistic regression models did; the analysis procedure is argued to be a black-box which is not useful for crash occurrence mechanism investigation (Roshandel et al., 2015).

Recently, as the Bayesian inference technique became popular in traffic safety analysis studies due to its flexible model structure and the advantage of incorporating prior knowledge, Bayesian logistic regression models were developed to evaluate the crash risk on both freeway (Ahmed et al., 2012a) and urban expressway (Ahmed et al., 2012b). And it was claimed that the Bayesian analysis approach provided better fit and reduced uncertainty for parameter estimations (Ahmed et al., 2012b). In addition, in order to account for the unobserved heterogeneity between crashes, Bayesian random effects logistic regression models were applied (Xu et al., 2013; Yu and Abdel-Aty, 2013b; Yu et al., 2013); better model goodness-of-fits have been achieved. In this study, Bayesian random effects logistic regression models were developed to analyze crash occurrence mechanisms for urban expressways in China.

2.2. Bayesian semi-parametric models

The Bayesian inference technique treats parameters as random variables which are characterized by prior distributions, which is the major difference with the Frequentist inference approach (Ntzoufras, 2009). Within the Bayesian inference framework, parametric Bayesian inference assumes the prior distributions can be indexed by finite-dimensional parameters, which would limit the inference scope and affect the model flexibility (Müller and Quintana, 2004). For the purpose of enriching the model flexibility and robustness and avoid mis-specifying parametric models, Bayesian semi-parametric or nonparametric models were introduced (Jara et al., 2011). Shively et al. (2010) adopted a Bayesian semi-parametric approach to analyze the relationships between crash counts and roadway characteristics. The nonparametric estimation procedure was employed for the model's link function and the results showed that the semi-parametric model was more robust compared to the standard parametric assumption.

One key advantage of Bayesian semi-parametric models is to model random effects distributions in hierarchical models (Müller and Quintana, 2004). Random effect terms were frequently introduced to account for the unobserved heterogeneity in crash risk analysis as mentioned above; since the true forms or distributions of these terms are unknown, semi-parametric approach is more reasonable for model estimation. However, to the best of our knowledge, no semi-parametric model has been adopted in crash risk analysis models. This study fills the gap as introducing semi-parametric Bayesian models to analyze crash risk and comparing the modeling results with parametric Bayesian models.

3. Data preparation

Data from the Shanghai urban expressway system were utilized; the roadway system includes Yan'an elevated road, North-South elevated road, Middle ring elevated road, Inner ring elevated road, and Yixian elevated road. As shown in Fig. 1, the Yan'an elevated road has an east–west trajectory, the North–South elevated road is north–south, and Yixian elevated road is radial; the Inner ring and the Middle ring elevated roads are two loops that cover mostly the urban areas of Shanghai. Three datasets were used to form the data: (i) crash data in January and March of 2009; (ii) roadway segment geometry data obtained from Shanghai Electronic Information Science Research Institute; (iii) traffic data detected by 1646 LDs at 518 detector sections (the LDs are lane specified) located on both directions along Shanghai urban expressways. LD data provide lane-based average speed and total volume at 5-min interval.

For each specific crash, stake number was utilized to describe its location (stake numbers are marked along the Shanghai urban expressway; generally consisted of letters or Chinese characters and numbers, and is ordered with non-repetitive numbers of designed foundation piles when constructed). The stake numbers of LDs on Shanghai urban expressway were collected for the purpose of identifying and matching traffic data with crash data; ArcMap (ESRI, 2006) were employed to identify the upstream and downstream LDs for each crash.

Based on the location of each crash, four nearest detector stations (two stations upstream and two stations downstream) were identified to monitor the traffic conditions prior to crash