Analytic Methods in Accident Research 2 (2014) 39–51

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

Crash frequency modeling for signalized intersections in a high-density urban road network

Kun Xie [a,b,c,*], Xuesong Wang [d], Kaan Ozbay [a,b,c], Hong Yang [a,b,c]

[a] Department of Civil and Urban Engineering, New York University, Brooklyn, NY 11201, USA
[b] Center for Urban Science and Progress (CUSP), New York University, Brooklyn, NY 11201, USA
[c] Urban Mobility and Intelligent Transportation Systems (UrbanMITS) Laboratory, New York University, Brooklyn, NY 11201, USA
[d] Department of Transportation Engineering, Tongji University, Shanghai, China

ARTICLE INFO

Article history:
Received 12 October 2013
Received in revised form
10 June 2014
Accepted 10 June 2014
Available online 15 July 2014

Keywords:
Crash frequency model
Signalized intersection
Hierarchy
Spatial correlation
Hierarchical conditional autoregressive model
High-density network

ABSTRACT

Conventional crash frequency models rely on an assumption of independence among observed crashes. However, this assumption is frequently proved false by spatially related crash observations, particularly for intersection crashes observed in high-density road networks. Crash frequency models that ignore the hierarchy and spatial correlation of closely spaced intersections can lead to biased estimations. As a follow-up to our previous paper (Xie et al., 2013), this study aims to address this issue by introducing an improved crash frequency model. Data for 195 signalized intersections along 22 corridors in the urban areas of Shanghai was collected. Moran's *I* statistic of the crash data confirmed the spatial dependence of crash occurrence among the neighboring intersections. Moreover, Lagrange Multiplier test was performed and it suggested that the spatial dependence should be captured in the model error term. A hierarchical model incorporating a conditional autoregressive (CAR) effect term for the spatial correlation was developed in the Bayesian framework. A deviance information criterion (DIC) and cross-validation test were used for model selection and comparison. The results showed that the proposed model outperformed traditional models in terms of the overall goodness of fit and predictive performance. In addition, the significance of the corridor-specific random effect and CAR effect revealed strong evidence for the presence of heterogeneity across corridors and spatial correlation among intersections.

1. Introduction

Statistical modeling of the inter-relationship between crash frequencies and contributing factors associated with intersections is of great interest in intersection safety studies. Traditional crash frequency models mainly assume that crash observations are independent from each other. However, this assumption is often proved false by crash observations occurring at signalized intersections in high-density urban road networks. Take the urban areas of Shanghai as an example, where the average intersection spacing is only about 200 m. Such short spacing leads to potential, and sometimes unavoidable, dependence among crash observations at adjacent intersections. This dependence can be attributed to two factors: (a) hierarchy – signalized intersections located along the same corridor share similar traffic flow, geometric design

* Corresponding author. Tel.: +1 646 997 0547.
E-mail addresses: kun.xie@nyu.edu (K. Xie), wangxs@tongji.edu.cn (X. Wang), kaan.ozbay@nyu.edu (K. Ozbay), hong.yang@nyu.edu (H. Yang).

http://dx.doi.org/10.1016/j.amar.2014.06.001
2213-6657/

and land use, and thus characteristics of the intersection can be described by a two-level hierarchical data structure consisting of the intersection-level information and corridor-level information; and (b) spatial correlation – neighboring intersections typically interact with one another in operation, for example, signals of adjacent intersections are usually coordinated and queue spillovers may frequently occur in dense road networks. Modeling aggregated crash frequencies without properly addressing the issues of hierarchy or spatial correlation can lead to unreliable findings.

This study is a follow-up to our recent paper (Xie et al., 2013) which focuses on the analysis of contributing factors to crashes in view of the unique traffic characteristics of Chinese cities. Different from the previous one, the main objective of this study is to develop an improved crash frequency model for closely spaced signalized intersections by accounting for the hierarchy and spatial correlation of the crash observations.

2. Literature review

Poisson models (Jones et al., 1991; Miaou and Lum, 1993) and negative binomial (NB) models (Miaou, 1994; Poch and Mannering, 1996; Abdel-Aty and Radwan, 2000) have been widely used to capture the relationship between traffic crashes and contributing factors. It is widely recognized that Poisson models outperform the standard regression models in handling the nonnegative, random and discrete features of crash counts (Joshua and Garber, 1990; Maher and Summersgill, 1996). Despite the improved performance, however, the constraint of the mean being equal to the variance in Poisson models is often violated by over-dispersed crash data. Alternatively, NB models are used to accommodate this over-dispersion issue by incorporating an independently distributed error term. However, with the assumption of independent observations, neither the Poisson models nor the NB models address any inherent correlation of crash data.

To complement the Poisson models and NB models, random effect models have been proposed in previous studies to account for the potential heterogeneity across homogeneous groups (Shankar et al., 1998; Chin and Quddus, 2003; Wang et al., 2014). In addition, random parameter models which can be viewed as extensions of random effect models are developed to incorporate the variability of both the intercept and the variable coefficients across observations and thus provide more flexibility for handling the heterogeneity (Anastasopoulos and Mannering, 2009; El-Basyouny and Sayed, 2009a; Venkataraman et al., 2011, 2013, 2014). More recently, hierarchical models have become the preferred method to accommodate a multilevel data structure (Jones and Jorgensen, 2003; Lenguerrand et al., 2006; Kim et al., 2007; Huang and Abdel-Aty, 2010; Ahmed et al., 2011; Xie et al., 2013; Chen and Persaud, 2014). Hierarchical models can accommodate the heterogeneity among different groups and have the ability to incorporate variables at the specific levels where impacts of specific variables occur (Gelman and Hill, 2007).

Spatial dependence is primarily modeled in two ways: using a spatially lagged dependent variable and using an error term. (Anselin, 1988b). The former way is denoted the spatial lag specification which allows spatial dependence through both spatial spillover effects (observed variables at one location can affect the dependent variable of itself and its neighboring locations) and spatial error correlation effects (omitted variables at one location can affect the dependent variable of itself and its neighboring locations) (Narayanamoorthy et al., 2013; Chiou et al., 2014). The latter is referred to as the spatial error specification that assumes the spatial dependence is only due to spatial error correlation effects. To model spatially correlated observations, generalized estimating equations (GEEs) are often the go-to method (Lord and Persaud, 2000; Hutchings et al., 2003; Abdel-Aty and Wang, 2006; Wang and Abdel-Aty, 2006). However, GEEs have the limitation of setting the same correlation matrix for all intersection groups, and thus cannot reflect the discrepancies in correlations among different groups of intersections. Conditional autoregressive (CAR) models can provide more flexibility in specifying the magnitude of correlation and have been recommended in many recent studies (Song et al., 2006; Aguero-Valverde and Jovanils, 2008; El-Basyouny and Sayed, 2009b; Guo et al., 2010). The CAR models capture the spatial dependence using the spatial error specification (Narayanamoorthy et al., 2013). To address the hierarchy and spatial correlation of crash data, a hierarchical model incorporating CAR effect terms is proposed in this study.

3. Data description

For this study, a total of 195 signalized intersections located along 22 corridors was selected in the urban areas of Shanghai. These intersections were limited to either 3-legged or 4-legged designs so that the analysis of geometric features could be simplified. To ensure the independence of intersections across corridors, efforts were made to avoid choosing intersecting corridors.

Variables at both intersection and corridor levels were extracted from different sources. First, the crash data for the year of 2009 were geocoded on a GIS map using the crash location description, and then linked to selected intersections. Second, traffic volumes for each intersection were acquired from loop detectors. Third, the Sydney coordinated adaptive traffic system (SCATS) provided the geometric design and traffic control data. In addition, for each intersection, the presence or absence of an elevated road over the intersection was observed and distances among adjacent intersections were measured from the projected coordinates of intersections on the GIS map.

In addition, floating car data (FCD) was obtained from over 40,000 GPS equipped taxis in Shanghai, which provided a continuous position data of each taxi operating in the city every ten seconds. The locations of the taxis were matched with the GIS road network, and only taxi samples that passed through the study corridors were chosen. The mean speed and speed variance of each corridor were acquired from the FCD of taxis traveling along the corridor.