Analytic Methods in Accident Research 17 (2018) 14–31

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

Using the multivariate spatio-temporal Bayesian model to analyze traffic crashes by severity

Chenhui Liu [a,b,*], Anuj Sharma [a]

[a] Department of Civil, Construction, and Environmental Engineering, Iowa State University, InTrans, 2711 South Loop Drive, Suite 4700, Ames, IA 50010-8664, United States

[b] Department of Statistics, Iowa State University, United States

ARTICLE INFO

Article history:
Received 23 September 2017
Received in revised form 3 February 2018
Accepted 4 February 2018

Keywords:
Multivariate spatio-temporal
Bayesian
Crash frequency
Posterior expected rank
Crash cost rate

ABSTRACT

Unobserved heterogeneity across space, time, and crash type is often non-negligible in crash frequency modeling. When multiple crash types with spatial and temporal features are analyzed, multivariate spatio-temporal models should be considered. For this study, we analyzed the yearly county-level fatal, major injury, and minor injury crashes in Iowa from 2006 to 2015 using a multivariate spatio-temporal Bayesian model. The model adopted a multivariate spatial structure, a multivariate temporal structure, and a multivariate spatio-temporal interaction structure to account for possible correlations across injury severities over space, time, and spatio-temporal interaction, respectively. Income and weather indicators were found to have no significant effects on crash frequencies in the presence of vehicle miles traveled and unemployment rate. Both spatial and temporal effects were found to be important, and they played nearly the same roles for all three crash types in the studied dataset. Counties located in north and southwest Iowa were found to tend to have fewer crashes than the remaining counties. All three crash types generally showed descending trends from 2006 to 2015. They also had significantly positive correlations between each other in space but not in time. The crude crash rates and predicted crash rates were generally consistent for major injury and minor injury crashes but not for low-count fatal crashes. High-risk counties were identified using the posterior expected rank by the predicted crash cost rate, which was more able to truly represent the underlying traffic safety status than the rank by the crude crash cost rate.

1. Introduction

Traffic crashes have been one of the major sources of fatalities and injuries in the United States. Crash frequency analysis is often used to identify key factors influencing the propensity of crashes, which is important for policymakers as they propose interventions to prevent road traffic crashes. However, unobserved heterogeneity is often an issue in crash frequency modeling, because many crash-related elements are often unavailable. Neglecting unobserved heterogeneity may produce biased and inefficient results (Mannering et al., 2016).

Unobserved heterogeneity may come from many sources. Crashes are usually classified into multiple types by different criteria, and their underlying correlations may produce some unobserved heterogeneity across observations when they are

* Corresponding author at: Department of Civil, Construction, and Environmental Engineering, Iowa State University, InTrans, 2711 South Loop Drive, Suite 4700, Ames, IA 50010-8664, United States.

E-mail addresses: cliu9@iastate.edu (C. Liu), anujs@iastate.edu (A. Sharma).

https://doi.org/10.1016/j.amar.2018.02.001
2213-6657/

analyzed simultaneously (Mannering and Bhat, 2014; Mannering et al., 2016). Thus, multivariate models, such as the multivariate Poisson log-normal (MVPLN) model, are often adopted (Ma et al., 2008; El-Basyouny and Sayed, 2009; Aguero-Valverde and Jovanis, 2010; El-Basyouny et al., 2014; Zhao et al., 2017). In addition, crash frequency data are always aggregated over space and time, which may also produce unobserved heterogeneity, as crashes that occur close in space or time are very likely to share some unobserved characteristics (Lord et al., 2005; Lord and Mannering, 2010; Savolainen et al., 2011; Mannering and Bhat, 2014; Mannering et al., 2016). Previous studies have shown that spatial correlations of traffic crashes may exist across states/provinces (Erdogan, 2009; Truong et al., 2016), counties (Aguero-Valverde and Jovanis, 2006; Song et al., 2006; Eckley and Curtin, 2013), census tracts (Wang and Kockelman, 2013), traffic analysis zones (Matkan and Mohaymany, 2013), intersections (Ahmed and Abdel-Aty, 2015; Liu et al., 2015), and segments (Aguero-Valverde and Jovanis, 2008; Wang et al., 2009, 2011; Aguero-Valverde, 2011; Jiang et al., 2014; Zeng and Huang, 2014). The similarity of economy, culture, land use, weather, traffic laws, and driving behavior within a given region may explain the spatial correlations in traffic crashes. When multiple crash types with spatial correlations need to be analyzed, multivariate spatial models have been proved to be more powerful than univariate spatial models, as multivariate spatial models can account for correlations across crash types in space in addition to spatial correlations (Miaou and Song, 2005; Song et al., 2006; Aguero-Valverde, 2013; Wang and Kockelman, 2013; Aguero-Valverde et al., 2016; Barua et al., 2016). Temporal correlations of traffic crashes may exist across year (Wang and Abdel-Aty, 2006; Brijs et al., 2008; Andrey, 2010; Wang et al., 2011; Yannis et al., 2011; Matkan and Mohaymany, 2013; El-Basyouny et al., 2014), month (Quddus, 2008a; Hu et al., 2013), week (Kilamanua et al., 2011; Sukhai et al., 2011; Liu et al., 2015), and day (Brijs et al., 2008). Temporal correlations occur because many traffic-related factors, such as economy, weather, environment, travel mode, and travel demand, often exhibit some temporal features. Similarly, when multiple crash types with temporal correlations need to be analyzed, multivariate temporal models should be considered, as they can account for correlations across crash types in time in addition to temporal correlations (Serhiyenko et al., 2014; Michalaki et al., 2016).

Crashes often have both spatial and temporal features. When only one crash type is analyzed, the univariate spatio-temporal modeling has been proved in some studies to be superior (Miaou et al., 2003; Aguero-Valverde and Jovanis, 2006; Truong et al., 2016; Liu and Sharma, 2017). When multiple crash types need to be analyzed, a multivariate spatio-temporal model may be needed. Ma et al. (2017) used the bivariate spatio-temporal model to analyze the daily non-injury and injury crash rates on 100 roadway segments of I70 in one year at the micro level, and Boulieri et al. (2017) used the bivariate spatio-temporal model to analyze the yearly low severity and high severity accidents of 7932 electoral wards in England from 2005–2013 considering only vehicle miles traveled (VMT). Both studies showed the superiority of the bivariate spatio-temporal model to the univariate spatio-temporal model in terms of goodness of fit.

In this study, we used the multivariate spatio-temporal Bayesian model to analyze the yearly county-level fatal, major injury, and minor injury crash frequencies in Iowa. The goal of this study was to accurately identify the long-term effects of economy and weather on crash frequency in Iowa and to explore the spatial and temporal correlations of crashes. Additionally, the counties were ranked to identify high-risk areas for safety improvement programs, as funding available for safety improvements are often limited and proper ranking can significantly influence the appropriate distribution of safety funding toward areas with more critical needs. Raw crash data-based ranking is easy to use but crude and inefficient (Miaou and Song, 2005). In Bayesian cases, one statistical ranking method is the posterior expected rank (PER), i.e. the posterior mean of the rank by ranking indicators (Miaou and Song, 2005). When rankings are the main interest, the PER method is recommended (Shen and Louis, 1998). The most common ranking indicator is crash rate, but crash rate considering crash cost by injury severity, called the "crash cost rate" in the following analysis, is strongly recommended when injury severity and associated costs are the main concerns (Miaou and Song, 2005). Thus, the PER of the crash cost rate would be used to rank the studied areas based on the predicted results of the multivariate spatio-temporal Bayesian model in this study.

2. Data description

Traffic crash data from Iowa's 99 counties from 2006 to 2015 were obtained from the Iowa Department of Transportation. Crashes were divided into five categories by severity: fatal, major injury, minor injury, possible injury/unknown, and property damage only. Fatal crashes, major injury crashes, and minor injury crashes were analyzed in this study, as these three types of crashes often lead to significant economic loss and casualties. VMT data for each county in each year from 2006 to 2015 were downloaded from the website of the Iowa Department of Transportation (2016). In addition, unemployment rate data were downloaded from the website of Iowa Community Indicators Program (2016), and per capita personal income data were downloaded from the website of the U.S. Bureau of Economic Analysis (2016) of the U.S. Department of Commerce. Meanwhile, weather data regarding rainfall, snowfall, and the number of days with minimum temperature exceeding 32°F (TH32) were downloaded from the website of the Iowa Environmental Mesonet (2017). These weather data are collected based on the daily climate observations from the National Weather Service's Cooperative Observer Program. A summary of the variables is given in Table 1. All three crash types have over-dispersion, as their variances are much larger than their means. Additionally, the highest correlation among the covariates was -0.338 (between snowfall and TH32). Thus, no explanatory variables showed strong positive or negative correlations.