Analytic Methods in Accident Research 16 (2017) 104–116

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

Exploring spatio-temporal effects in traffic crash trend analysis

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, USA
[b] Department of Statistics, Iowa State University, USA

A R T I C L E   I N F O

Article history:
Received 25 June 2017
Received in revised form 25 September 2017
Accepted 25 September 2017

Keywords:
Spatio-temporal modeling
Bayesian
Integrated nested Laplace approximation
Conditional autoregressive
Unobserved heterogeneity

A B S T R A C T

Unobserved heterogeneity produced by spatial and temporal correlations of crashes often needs to be captured in crash frequency modeling. Although many studies have included either spatial or temporal effects in crash frequency modeling, only a limited number of studies have considered both. This study addresses the limitations of existing studies by exploring multiple models that best fit the spatial and temporal correlations. In this study, we used Bayesian spatio-temporal models to investigate regional crash frequency trends, and explored the effects of omitting spatial or temporal trends in spatio-temporal correlated data. The fast Bayesian inference approach, integrated nested Laplace approximation, was used to estimate parameters. It was found that fatal crashes showed decreasing trends in all Iowa counties from 2006 to 2015, but the decreasing rates varied by counties. Among all the covariates investigated, only vehicle miles traveled (VMT) was significant. None of the socio-economic or weather indicators were found to be significant in the presence of VMT. Both spatial and temporal effects were found to be important, and they were responsible for both over dispersion and zero inflation in the crash data. In addition, spatial effects played a more important role than did temporal effects in the studied dataset, but temporal component selection was still important in spatio-temporal modeling.

1. Introduction

Traffic crashes have been one of the major sources of fatalities and injuries in the United States. Crash frequency models often are used to identify the factors influencing the propensity of traffic crashes. The most common crash frequency model is the Poisson model. When crashes show over dispersion, quasi-Poisson, Poisson log-normal model (PLN), and negative binomial (NB) models are often adopted. Unobserved heterogeneity is often an issue in crash frequency analysis, because many crash-related factors are often not observed by the analyst (Mannering et al., 2016). The excess zeros in crash data can be a result of unobserved heterogeneity (Mullahy, 1997), often causing zero-inflated and hurdle models to be adopted (Lord et al., 2005; Lord and Mannering, 2010; Malyshkina and Mannering, 2010; Mannering et al., 2016; Mannering and Bhat, 2014). In addition, the zero-state Markov switching model, which allows observations to switch between zero and normal-count states over time, has been proven to be a viable alternative to zero-inflated models (Malyshkina and Mannering, 2010). Because crash data are often aggregated over time and space, spatial and temporal correlations are often also responsible for a portion of 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; Mannering et al., 2016; Mannering

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

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

https://doi.org/10.1016/j.amar.2017.09.002
2213-6657/

and Bhat, 2014; Savolainen et al., 2011). However, these spatial and temporal correlations are often overlooked in existing studies, and neglecting them may produce inefficient or biased estimated results (Mannering et al., 2016; Mannering and Bhat, 2014; Savolainen et al., 2011).

The spatial correlation of traffic crashes may exist on a macro- or microscopic spatial scale. At a macroscopic level, factors such as census tract (Wang and Kockelman, 2013), traffic analysis zone (Matkan and Mohaymany, 2013), ZIP code level (Ponicki et al., 2013), census block group (Noland et al., 2013), electoral ward (Boulieri et al., 2017), census ward (Quddus, 2008a), county (Aguero-Valverde and Jovanis, 2006; Eckley and Curtin, 2013; Song et al., 2006), and state/province (Erdogan, 2009; Truong et al., 2016), as well as similarity of economic and social activities, culture, land use, and enforcements within a given region, may explain the spatial correlation in traffic crashes. At a microscopic level, crashes occurring at nearby intersections (Abdel-Aty and Wang, 2006; Ahmed and Abdel-Aty, 2015; Guo et al., 2010; Liu et al., 2015; Mitra et al., 2007; Pulugurtha and Sambhara, 2011; Wang and Abdel-Aty, 2006; Xie et al., 2014) or adjacent road segments (Aguero-Valverde, 2011; Aguero-Valverde and Jovanis, 2008; Jiang et al., 2014; Wang et al., 2011, 2009; Zeng and Huang, 2014) may be correlated as a result of geometric or traffic flow similarities (Levine et al., 1995).

Temporal correlation captures the variability of traffic crashes with temporal scales such as year (Andrey, 2010; Boulieri et al., 2017; Brijs et al., 2008; El-Basyouny et al., 2014; Malyshkina and Mannering, 2010; Matkan and Mohaymany, 2013; Wang et al., 2011; Wang and Abdel-Aty, 2006; Yannis et al., 2011), month (Hu et al., 2013; Quddus, 2008b), week (Kilamanua et al., 2011; Liu et al., 2015; Malyshkina et al., 2009; Sukhai et al., 2011), day (Brijs et al., 2008), and hour (Kilamanua et al., 2011; Liu et al., 2015). Temporal correlation reflects the influence of different traffic-related factors, such as economy, weather, environment, law, and travel demand, which often exhibit some temporal trends or periodicities.

Depending on the study site, one of three scenarios is feasible: (a) the crash data may show both spatial and temporal effects, (b) these effects may exist individually, or (c) neither of them may exist. When spatial and temporal effects coexist, their interaction (i.e. spatio-temporal effects) also needs to be considered. Although many studies have included either spatial effects or temporal effects in crash frequency modeling, only a limited number of studies have considered both of them. Miaou et al. (2003) first introduced the spatio-temporal modeling approach to traffic crash modeling in analyzing yearly county-level crash rates in Texas from 1992 to 1999 using multiple spatio-temporal models. Wang and Abdel-Aty (2006) analyzed spatial and temporal correlations for rear-end crashes at signalized intersections in Florida. However, they built separate models for spatial effects and temporal effects. Jiang et al. (2014) considered both spatial and temporal correlations in analyzing the crashes on urban four-lane divided arterial segments in the central Florida area. However, they assumed that the spatial and temporal effects followed normal distributions without presenting any data-driven evidence to support their assumption. Truong et al. (2016) analyzed yearly crash fatalities of 63 provinces in Vietnam from 2012 to 2014 using the conditional autoregressive (CAR) spatio-temporal autocorrelation technique. The CAR spatio-temporal model performed better than the random effects NB model and random parameters NB model did in terms of both goodness of fit and crash prediction. Aguero-Valverde and Jovanis (2006) had similar findings.

The CAR model (Besag, 1974; Besag et al., 1991) often is used for modeling areal data in spatial statistics. Several researchers (Aguero-Valverde and Jovanis, 2006; Boulieri et al., 2017; Truong et al., 2016; Wang et al., 2011) have used the CAR model to illustrate spatial correlations paired with different temporal models. However, they all showed only one temporal model, despite the fact that the choice of a particular temporal model was also very important (Miaou et al., 2003). In this study, we used the spatio-temporal crash frequency model to identify the long-term county-level fatal crash frequency trends in Iowa. Multiple temporal components were built and contrasted to choose the most appropriate model. A fast Bayesian estimation tool, integrated nested Laplace approximation (INLA), was used to estimate these spatio-temporal models.

The workflow of the data analysis is as follows:

- First, we discuss whether crashes have over dispersion and zero inflation.
- Second, we examine spatial correlations and temporal correlations of crashes.
- Third, we evaluate the necessity of including the spatial component, temporal component, and spatio-temporal component in modeling, and we also discuss the temporal component selection.
- Finally, after determining the final model, the estimation results are discussed.

The rest of paper is organized as follows. Section 2 comprises a discussion of the traffic crash data used for this study. Section 3 presents the statistical models and estimation methods used in this study. Section 4 includes the analyses and discussions of the observed results. A conclusion and future recommendations are provided in Section 5.

2. Data description

Traffic crash data for Iowa's 99 counties from 2006 to 2015 were obtained from the Iowa Department of Transportation. Based on their severity, the crashes were divided into five categories: fatal, major injury, minor injury, possible injury/unknown, and property damage only. Fatal crashes were analyzed for this study, as they usually cause much more severe outcomes than do other types of crashes. The vehicle miles traveled (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, population and unemployment rate data were downloaded from the website of Iowa Community Indicators Program (2016), and per