Analytic Methods in Accident Research 19 (2018) 1–15

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

# Developing a grouped random parameters multivariate spatial model to explore zonal effects for segment and intersection crash modeling

Qing Cai [a,*], Mohamed Abdel-Aty [a], Jaeyoung Lee [a], Ling Wang [a,b], Xuesong Wang [b,c]

[a] *Department of Civil, Environment and Construction Engineering, University of Central Florida, Orlando, FL, USA*
[b] *College of Transportation Engineering, Tongji University, Shanghai 201804, China*
[c] *Key Laboratory of Road and Traffic Engineering and the Engineering Research Center for Road Traffic Safety and Environment, Tongji University, Shanghai 201804, China*

ARTICLE INFO

*Article history:*
Received 12 December 2017
Received in revised form 20 May 2018
Accepted 21 May 2018

*Keywords:*
Segment and intersection crash analysis
Zonal effects
Multivariate spatial autocorrelation
Grouped random parameters model

ABSTRACT

It is acknowledged that crash occurrence on segments and intersections could be affected by multilevel factors. Omission of important explanatory variables could result in biased and inconsistent parameter estimates. This paper contributes to the literature by examining the zonal effects which are always excluded or ignored in traffic safety research for segments and intersections. A grouped random parameters multivariate spatial model is proposed to identify both observable zonal effects and unobserved heterogeneity at the zonal level by considering the heterogeneous and spatial correlations. The proposed model is evaluated by comparing it with its three counterparts: a fixed parameters univariate spatial model without zonal factors, a random parameters univariate spatial model without zonal factors, and a random parameters univariate spatial model with zonal factors. The results indicate that the three random parameters models could consistently provide better performance than the fixed parameters model and the models including zonal factors outperform the models without zonal factors. Besides, the proposed model has the optimal model performance compared with its counterparts, which validates the concept of adopting the multivariate modeling framework to identify the heterogeneous and spatial correlations of zonal effects. The results confirm the significantly correlated heterogeneous residuals in modeling zonal factors on crash occurrence on segments and intersections. However, the spatial correlations of zonal effects on different types of road entities (segments and intersections) in adjacent zones are not statistically significant. Furthermore, the proposed model provides more valuable insights about the crash occurrence on segments and intersections by revealing segment-/intersection-level factors together with zonal factors.



## 1. Introduction

To ensure efficient safety countermeasures, it is essential to understand the effects of exogenous factors on crash occurrence (Wang et al., 2015). Segments and intersections are two major parts of the road network that carry traffic demands. In the previous literature, numerous traffic crash analysis models have been developed for the two types of road entities. The

* Corresponding author.
*E-mail address:* qingcai@knights.ucf.edu (Q. Cai).

https://doi.org/10.1016/j.amar.2018.05.001
2213-6657/

choice of appropriate analytical models and the selection of explanatory variables would be the two crucial elements to build reliable relationships between safety and its contributing features. With various advanced statistical methodologies such as spatial and temporal autocorrelation, finite mixture, latent class, random effects, random parameters, and multilevel approaches, etc., the effects of segment-/intersection-level factors on the segment and intersection crashes have been recognized (Barua et al., 2016; Bhat et al., 2014; Castro et al., 2012; Fu et al., 2018; Huang et al., 2017; Lee et al., 2017; Liu and Sharma, 2018; Ma et al., 2017; Osama and Sayed, 2017; Wang et al., 2018a; Xie et al., 2014). The segment-/intersection-level factors refer to variables aggregated at the segment/intersection level including traffic data (e.g., average annual daily traffic (AADT)), geometric data (e.g., number of lanes, road functional classification), etc. Beside the segment-/intersection-level factors, the road entities should share certain zonal factors, which may affect driving behaviors, travel modes, and further affect crash occurrence. In this study, the zonal factors indicate the information based on zones including traffic data (e.g., vehicle miles travelled (VMT)), socio-demographic information (e.g., population, median household income), etc. Although the crash studies based on zones have suggested that the zonal factors have significant effects on traffic safety, only few studies have included zonal data for the segment/intersection safety analysis. Omission of important explanatory variables at the zonal level may result in biased and inconsistent parameter estimates for crash analysis on segments and intersections (Mannering et al., 2016; Wang et al., 2017a, Cai et al., 2018). In addition to the observed factors from zones, there should exist unobserved heterogeneity at the zonal level that would affect the crash frequency outcomes on the segments (or intersections) in the same zones.

This study aims to investigate the observed and unobserved effects of zones on crashes of segments and intersections. Toward this end, a grouped random parameters multivariate spatial model, which allows investigating the potential observable zonal effects as well as unobserved heterogeneity at the zonal level on crash occurrence on segments and intersections. In addition, the proposed model considers the potential correlation of zonal effects on crash occurrence on segments and intersections due to the heterogeneous and spatial effects.

This paper is organized as follows: A brief discussion in earlier literature about spatial effects, zonal effects, and modeling methodologies is provided in the next section. Section 3 introduces the grouped random parameters multivariate spatial model to explore the zonal effects on crashes of both segments and intersections. Section 4 describes the collected data for the empirical analysis of the proposed model. The model estimation results of the proposed model and three other counterparts are presented in Section 5. Finally, the last section summarizes and concludes this paper.

## 2. Literature review

In the previous traffic safety literature, Poisson-based models (e.g., Poisson, negative binomial, Poisson lognormal models) have been widely used and proven to be appropriate to deal with the non-negative crash count data. Nevertheless, the models assume that observations are independent from each other, which might violate the fact that crash data based on segments and intersections might share possible spatial correlation effects and zonal effects. A comprehensive literature review about such effects in crash count modeling is presented in the following two subsections. In addition, the relevant modeling methodologies are discussed.

### *2.1. Spatial effect*

Road entities/traffic zones in close spatial proximity may possess common unobserved features, thereby inducing a spatial autocorrelation between crash counts of road entities/traffic zones within neighborhoods. A vast body of studies has shown that accounting for spatial autocorrelation can contribute to unbiased parameter estimation in identifying the exogenous variables (Lord and Mannering, 2010; Mannering and Bhat, 2014). At both road entity- and zone-levels, two approaches to incorporate spatial autocorrelation are considered: (1) spatial spillover effects (observed exogenous variables at one location have impacts on crash occurrence at both the targeted and neighboring location) and (2) spatial error correlation effects (unobserved exogenous variables at one location affect crash occurrence at the targeted and neighboring locations) (Cai et al., 2016). Researchers have considered a spatially lagged dependent variable at neighboring locations for the spatial spillover effects (LaScala et al., 2000; Ha and Thill, 2011; Quddus, 2008). On the other hand, several research efforts have accommodated for spatial error in the safety literature. In particular, the Bayesian spatial Poisson lognormal model with Gaussian Conditional Autoregressive (CAR) prior for modeling spatial error correlation effects has been widely applied in road safety analysis (Ahmed et al., 2011; Cai et al., 2016; Xie et al., 2014).

### *2.2. Zonal effect*

Crash data are often characterized a multilevel (hierarchical) structure since there are many factors from different levels can affect crash occurrence (Dupont et al., 2013; Huang et al., 2010; Mitra and Washington, 2012). At both segment- and intersection-level, the crash data would be nested in various spatial units. Hence, it would be reasonable to claim that the segments or intersections located in the same zone should share certain zonal factors, which may affect crash occurrence through regional traffic regulations, driving behaviors, climate-related features, etc. (Heydari et al., 2016; Wu et al., 2017). Omission of important zonal features may result in biased and inconsistent parameter estimates (Mannering et al., 2016;