Analytic Methods in Accident Research 3-4 (2014) 28–43

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

A Full Bayesian multivariate count data model of collision severity with spatial correlation

Sudip Barua, Karim El-Basyouny [1], Md Tazul Islam *

Department of Civil and Environmental Engineering, University of Alberta, 4-110 NREF Edmonton, AB, Canada T6G 2W2

ARTICLE INFO

Article history:
Received 15 April 2014
Received in revised form
6 September 2014
Accepted 7 September 2014
Available online 4 November 2014

Keywords:
Multivariate count data modeling
Spatial correlation
Heterogeneous effects
Full Bayesian (FB) estimation
Markov Chain Monte Carlo (MCMC)
Poisson lognormal regression

ABSTRACT

This study investigated the inclusion of spatial correlation in multivariate count data models of collision severity. The models were developed for severe (injury and fatal) and no-injury collisions using three years of collision data from the city of Richmond and the city of Vancouver. The proposed models were estimated in a Full Bayesian (FB) context via Markov Chain Monte Carlo (MCMC) simulation. The multivariate model with both heterogeneous effects and spatial correlation provided the best fit according to the Deviance Information Criteria (DIC). The results showed significant positive correlation between various road attributes and collision severities. For the Richmond dataset, the proportion of variance for spatial correlation was smaller than the proportion of variance for heterogeneous effects. Conversely, the spatial variance was greater than the heterogeneous variance for the Vancouver dataset. The correlation between severe and no-injury collisions for the total random effects (heterogeneous and spatial) was significant and quite high (0.905 for Richmond and 0.945 for Vancouver), indicating that a higher number of no-injury collisions is associated with a higher number of severe collisions. Furthermore, the multivariate spatial models were compared with two independent univariate Poisson lognormal (PLN) spatial models, with respect to model inference and goodness-of-fit. Multivariate spatial models provide a superior fit over the two univariate PLN spatial models with a significant drop in the DIC value (35.3 for Richmond and 116 for Vancouver). These results advocate the use of multivariate models with both heterogeneous effects and spatial correlation over univariate PLN spatial models.

1. Introduction

Collision modeling is widely considered the key method for estimating the safety levels of different road entities (i.e., intersections and road segments). Collision models are mathematical models statistically developed to link collision occurrence or frequency to a roadway's traffic and geometric characteristics. To develop a collision model, it is useful to classify the likely random effects on the distribution of collisions into four categories: i) Poisson variation, as collisions are a discrete, nonnegative, and random event; ii) heterogeneity (extra-variation), due to within-site effects reflecting individual characteristics; iii) spatial correlation (spatial effects), since neighboring sites typically have similar environmental and

* Corresponding author. Fax: +1 780 492 0249.
E-mail addresses: barua1@ualberta.ca (S. Barua), karim.el-basyouny@ualberta.ca (K. El-Basyouny), mdtazul@ualberta.ca (M.T. Islam).
[1] Fax: +1 780 492 0249.

http://dx.doi.org/10.1016/j.amar.2014.09.001
2213-6657/

geographical characteristics and may share unobserved effects; and iv) temporal correlation (temporal effects), since data collected over successive time periods may have unobserved effects.

Most of the literature related to the development of collision models accounts only for Poisson variation and heterogeneity. Despite the evident spatial nature of collisions, little road safety research has been conducted to account for spatial correlations. Recently, however, the need to include spatial correlation in the development of collision models for both intersections and road segments (Abdel-Aty and Wang, 2006; Aguero-Valverde and Jovanis, 2008, 2010; El-Basyouny and Sayed, 2009a; Mitra, 2009) and the area-wide (e.g., wards, neighborhoods, county) (Aguero-Valverde and Jovanis, 2006; Aguero-Valverde, 2013; Amoros et al., 2003; Flask and Schneider IV, 2013; Noland and Quddus, 2004; Quddus, 2008) has been gaining attention in the literature. Congdon (2006) suggested that ignoring spatial dependence leads to an underestimation of variability. Furthermore, according to Aguero-Valverde and Jovanis (2010), the inclusion of spatially correlated random effects significantly improves the precision of the estimates of the expected collision frequency for road segments. The inclusion of spatial correlation has two main advantages: i) spatial correlation sites estimate the “pool strength” from neighboring sites, thereby improving model parameter estimation (Aguero-Valverde and Jovanis, 2008); and ii) spatial dependence can be a surrogate for unknown and relevant covariates, thereby reflecting unmeasured confounding factors (Cressie, 1993; Dubin, 1988).

Most studies used spatial models to examine collision frequency or type independently. However, collision data is multivariate in nature, and it is necessary to account for the likely correlation between collision counts at different levels of classification. While a number of studies explored multivariate crash modeling to capture the heterogeneous correlations among different collision types or severities (Bijleveld, 2005; Ma and Kockelman, 2006; Ma et al., 2008; Park and Lord, 2007; El-Basyouny and Sayed, 2009b, El-Basyouny et al., 2014), multivariate spatial correlations were rarely investigated. Furthermore, univariate spatial modeling of different types of collision counts may lead to biased results, because collision types or severities are not spatially independent of one another. With the link of multivariate spatial correlation, a collision model can estimate the associated safety risk and spatial correlation of different collision severities and types in the same spatial unit. However, only two studies (Aguero-Valverde, 2013; Song et al., 2006) focused on area-wide multivariate spatial models for different collision severities and types. Further, Wang and Kockelman (2013) and Narayanamoorthy et al. (2013) used a multivariate spatial modeling approach for pedestrian and bicyclist collision analysis at the census tracts level. Multivariate spatial modeling for road segments or intersections has rarely been explored in the literature. To this end, there are two main objectives of this study: i) use a multivariate spatial modeling approach to develop spatial models for road segments in order to assess spatial correlation in different collision severity levels and their influence on the collision analysis of urban arterials; and ii) compare multivariate spatial models with independent (separate) univariate spatial models for each collision severity in terms of model inference and goodness-of-fit. To accomplish the objectives, three years (1994–1996) of collision data and other geometric and non-geometric road data were used from the cities of Richmond and Vancouver, British Columbia, Canada.

From a methodological perspective, various approaches, such as Moving Average (Congdon, 2006), Simultaneous Auto-regressive (SAR) (Quddus, 2008), Spatial Error Model (SEM) (Anselin, 1988; Quddus, 2008), Multiple Membership (MM) (El-Basyouny and Sayed, 2009a; Goldstein, 1995; Langford et al., 1999), Extended Multiple Membership (EMM) (El-Basyouny and Sayed, 2009a), Geographic Weighted Regression (GWR) (Hadayeghi et al., 2003), Geographic Weighted Poisson Regression (GWPR) (Hadayeghi et al., 2010), and Generalized Estimating Equations (GEE) (Abdel-Aty and Wang, 2006) have been advocated by other researchers to assess spatial effects or spatial correlation. Each approach has its own pros and cons. However, almost all of the earlier studies used Gaussian Conditional Auto-regressive (CAR) (Besag et al., 1991) distribution for modeling spatial correlation (Aguero-Valverde and Jovanis, 2006, 2008, 2010; Ahmed et al., 2011; Mitra, 2009; Siddiqui et al., 2012). In addition, Quddus (2008) advocated that CAR distribution under a Bayesian framework can provide more appropriate and better inference over classical spatial models, because the Bayesian CAR models with heterogeneous effects are able to accurately take into account both spatial correlation and unobserved heterogeneity of the collision data. Therefore, in this study, univariate and multivariate spatial models were applied using CAR distribution. The models were estimated in a Full Bayesian (FB) context via Markov Chain Monte Carlo (MCMC) simulation (Gilks et al., 1996). As WinBUGS (Lunn et al., 2000) is a flexible platform for the Bayesian analysis of complex statistical models using MCMC methods, this open-source statistical software was used for the development of the proposed spatial models.

2. Previous research

Conventional collision models with Poisson-gamma or Poisson lognormal (PLN) distribution assume that sites are independent of one another and, hence, can be regarded as non-spatial models. However, as collision data are collected with reference to location, which is measured as points in space (Quddus, 2008), a spatial correlation exists between observations (LeSage, 1998). Ignoring spatial correlations may lead to a biased estimation of the model parameters. Hence, a number of road element-specific (i.e., intersection or road segment) and area-wide accident studies have incorporated spatial correlation in modeling collision data (Aguero-Valverde and Jovanis, 2006, 2008, 2010; Aguero-Valverde, 2013).

In the context of intersection-based spatial models, Abdel-Aty and Wang (2006) used GEE to address spatial correlation between signalized intersections. The authors determined that signalized intersections, especially ones close together along a certain corridor, are spatially correlated and influence one another. Similarly, Guo et al. (2010) applied Poisson and