Analytic Methods in Accident Research 5-6 (2015) 28–42

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

Effects of spatial correlation in random parameters collision count-data models

Sudip Barua [a,1], Karim El-Basyouny [a,2], Md. Tazul Islam [b,*]

[a] Department of Civil and Environmental Engineering, University of Alberta, Edmonton, AB, Canada T6G 2W2
[b] Department of Civil and Environmental Engineering, University of Alberta, 6-106 NREF, Edmonton, AB, Canada T6G 2W2

ARTICLE INFO

Article history:
Received 29 July 2014
Received in revised form
9 February 2015
Accepted 9 February 2015
Available online 17 March 2015

Keywords:
Collision count-data modeling
Random parameters model
Spatial correlation
Heterogeneous effects
Full Bayesian (FB) estimation
Poisson-lognormal regression

ABSTRACT

This study investigated the inclusion of spatial correlation in random parameters collision count-data models. Three different modeling formulations were applied to measure the effects of spatial correlation in random parameters models using three years of collision data collected from two cities, Richmond and Vancouver (British Columbia, Canada). The proposed models were estimated in a Full Bayesian (FB) context using a Markov Chain Monte Carlo (MCMC) simulation. The Deviance Information Criteria (DIC) values and chi-square statistics indicated that all the models were comparable to one another. According to the parameter estimates, a variety of traffic and road geometric covariates were found to significantly influence collision frequencies. For the Richmond dataset, only 38.3% of the total variability was explained by spatial correlation under model with both heterogeneous effects and spatial correlation (Model C), as most of the variations were likely captured by heterogeneous effects and site variation. For the Vancouver dataset, the effects of spatial correlation were much clearer, with a high percentage of the total variability (83.8%) explained by spatial correlation under Model C. Moreover, model estimation results showed that the precision of parameter estimates slightly improved by inclusion of spatial correlation when the sample size was small. However, parameter estimations did not change significantly and goodness of fit did not improve which indicate that it cannot be substantiated with the current datasets that the random parameters model with both heterogeneous effects and spatial correlation is better than other models investigated. Therefore, further studies with different datasets are needed to get more clear understanding of the added benefits of incorporating spatial correlation in random parameters model.

1. Introduction

Most of the previous research related to the development of collision models focuses on accounting for Poisson variation (Jovanis and Chang, 1986; Joshua and Garber, 1990; Miaou, 1994; Miaou and Lum, 1993) and heterogeneity (El-Basyouny and Sayed, 2006; Hauer et al., 1988; Lord and Bonneson, 2007; Maher and Summersgill, 1996; Maycock and Hall, 1984; Milton and Mannering,

* Corresponding author. Tel.: +1 780 695 5877; fax: +1 780 492 0249.
E-mail addresses: barua1@ualberta.ca (S. Barua), karim.el-basyouny@ualberta.ca (K. El-Basyouny), mdtazul@ualberta.ca (Md.T. Islam).
[1] Tel.: +1 780 399 8073.
[2] Tel.: +1 780 492 9564; fax: +1 780 492 0249.

http://dx.doi.org/10.1016/j.amar.2015.02.001
2213-6657/

1998; Persaud, 1994) in collision data. The parameters of these traditional collision models were assumed to be fixed when they can actually vary across observations (road segments or intersections). Further, due to unobserved heterogeneity, the effect of an explanatory variable on collisions may vary for different observations. For instance, a two-lane road segment with high traffic volume may have a high collision frequency compared to a similar road configuration with lower traffic volume. Therefore, constraining the parameters may not incorporate the site-specific effects, leading to an underestimation of standard errors and inconsistent, biased, and erroneous inference (Washington et al., 2003). Given the magnitude of this potential issue, several researchers successfully applied random parameters to collision modeling. For instance, Milton et al. (2008), Gkritza and Mannering (2008), Anastasopoulos and Mannering (2009, 2011), El-Basyouny and Sayed (2009a), Anastasopoulos et al. (2012), and Russo et al. (2014) all demonstrated that random parameters models can provide more accurate inference than traditional fixed parameters models, as well as account for heterogeneity across sites due to unobserved road geometrics, traffic characteristics, environmental factors, driver behavior, and other confounding factors.

Spatial correlation is another key issue that has been gaining attention in the development of collision models (Abdel-Aty and Wang, 2006; Aguero-Valverde, 2013; Aguero-Valverde and Jovanis, 2006, 2008, 2010; Amoros et al., 2003; El-Basyouny and Sayed, 2009c; Flask and Schneider, 2013; Mitra, 2009; Noland and Quddus, 2004; Quddus, 2008). As collision data are collected with respect to location, which is measured as points in space (Quddus, 2008), a spatial correlation exists between observation sites (LeSage, 1998). Ignoring this spatial correlation may lead to a biased estimation of the model parameters. According to Aguero-Valverde and Jovanis (2010), the inclusion of spatially correlated random effects in collision models significantly improves the precision of the estimated expected collision frequency for road segments. The inclusion of spatial effects 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 previous studies have used random parameters and spatial correlation separately; rarely in the literature have these two factors been incorporated together in collision models. Nevertheless, it is necessary to investigate whether the inclusion of spatial correlation in a random parameters model can improve the model's goodness-of-fit and precision of parameter estimation. This approach may help gain new insight in collision analysis and allow for a more precise assessment of the safety risks of collision sites. Thus, the main objective of this study is to investigate the effects of spatial correlation in random parameters collision count-data models. To accomplish the objective, three years (1994–1996) of collision data and other geometric and non-geometric road data were used for two cities in British Columbia, Canada: (i) Richmond and (ii) Vancouver.

In determining the model formulation for this study, several different approaches in the literature were reviewed for suitability with respect to random parameters and spatial correlation. From a methodological perspective, a wide variety of modeling approaches, such as the negative binomial model (Anastasopoulos and Mannering, 2009; Chen and Tarko, 2014; Garnowski and Manner, 2011; Ukkusuri et al., 2011; Venkataraman et al., 2011, 2013; Wu et al., 2013), logit model (Anastasopoulos and Mannering, 2011), mixed logit model (Gkritza and Mannering, 2008; Milton et al., 2008), Tobit model (Anastasopoulos et al., 2012), Poisson-lognormal (PLN) model (El-Basyouny and Sayed, 2009a), bivariate ordered probit model (Russo et al., 2014), and finite mixture model (Xiong and Mannering, 2013) were used to employ random parameters in collision analysis. To account for spatial correlation, 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) (Goldstein, 1995; El-Basyouny and Sayed, 2009c; Langford et al., 1999), extended multiple membership (EMM) (El-Basyouny and Sayed, 2009c), 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. Each approach has its own pros and cons, but almost all 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; Guo et al., 2010; Mitra, 2009; Siddiqui et al., 2012). Quddus (2008) also advocated that CAR distribution under a Bayesian framework can provide more appropriate and better inference over traditional spatial models because the Bayesian CAR models are able to accurately take into account both spatial correlation and the unobserved heterogeneity of the collision data.

Given the support in the literature, random parameters spatial models in this study were developed using CAR distribution. As the PLN modeling approach is more flexible than the traditional Poisson-gamma or negative binomial model to handle over-dispersion, the lognormal distribution was used for the heterogeneous effects, which leads to PLN posterior distribution. The models were estimated in a Full Bayesian (FB) context via Markov Chain Monte Carlo (MCMC) simulation (Gilks et al., 1996), which has been recognized as an alternative way to estimate complex discrete-outcome model structures. 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 random parameters spatial models.

2. Previous research

2.1. Random parameters models

Despite the fact that random parameters models have outperformed traditional fixed parameter models, a limited number of studies have used this approach in safety research. The key reasons behind this are (i) random parameters