Analytic Methods in Accident Research 9 (2016) 1–15

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

Multivariate random parameters collision count data models with spatial heterogeneity

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

[a] *Graduate Research Assistant Department of Civil and Environmental Engineering University of Alberta, Edmonton, AB, Canada T6G 2W2*
[b] *City of Edmonton Assistant Professor of Urban Traffic Safety Department of Civil and Environmental Engineering University of Alberta, Edmonton, AB, Canada T6G 2W2*
[c] *Department of Civil and Environmental Engineering University of Alberta, 4-110 NREF, Edmonton, Alberta, Canada T6G 2W2*

ARTICLE INFO

Article history:
Received 21 September 2014
Received in revised form
31 October 2015
Accepted 5 November 2015
Available online 11 December 2015

Keywords:
Collision count data modeling
Random parameters model
Spatial correlation
Heterogeneous effects
Full Bayesian (FB) estimation
Multivariate modeling.

ABSTRACT

This study investigated the effects of including spatial heterogeneity in multivariate random parameters models and their influence on different collision severity levels. The models were developed for severe (injury and fatal) and no-injury collisions using three years of collision data from the city of Vancouver. Three different modeling formulations were applied to measure the effects of spatial heterogeneity in a multivariate random parameters model. The proposed models were estimated in a Full Bayesian (FB) context using Markov Chain Monte Carlo (MCMC) simulation. The Deviance Information Criteria (DIC) values indicated that all the models were comparable to one another. Therefore, no particular model can be distinctly preferred over others. According to parameter estimates, a variety of traffic and road geometric covariates were found to significantly influence collision severities. The variance for spatial heterogeneity was higher than the variance for heterogeneous effects. The correlation between severe and no-injury collisions for the total random effects (heterogeneous and spatial) was significant and quite high, indicating that higher no-injury collisions are associated with higher severe collisions. These results support the incorporation of spatial heterogeneity in multivariate random parameters models. Furthermore, the multivariate random parameters spatial models were compared with two independent univariate random parameters spatial models with respect to model inference and goodness of fit. The multivariate spatial models outperformed the two univariate spatial models with a very significant drop in the DIC value.

1. Introduction

Over the last two decades, considerable research efforts have been devoted to developing and applying sophisticated methodological approaches to address several collision data-related issues (e.g., over-dispersion, under-dispersion, omitted variable bias, fixed parameters, functional form), thereby improving the precision of the parameter estimates and the model's predictability. Despite these methodological innovations and developments, several complex issues may still exist

---

[*] 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.2015.11.002

(e.g., unobserved heterogeneity, endogeneity, spatial and temporal correlation, correlated collision types—for more information, readers can refer to Lord and Mannering (2010) and Mannering and Bhat (2014)), which can substantially influence the inference, precision, and findings from the collision data analysis. Fortunately, over the past few years, substantial methodological developments have emerged to address these potential issues, including the following:

- Use of random parameters in collision models to capture unobserved heterogeneity across observations (Anastasopoulos and Mannering, 2009; 2011; Anastasopoulos et al., 2012a; Chen and Tarko, 2014; Dinu and Veeraragavan, 2011; El-Basyouny and Sayed, 2009a; El-Basyouny et al., 2014a, Garnowski and Manner, 2011; Gkritza and Mannering, 2008; Milton et al., 2008; Russo et al., 2014; Ukkusuri et al., 2011; Venkataraman et al., 2011; 2013; Wu et al., 2013; Xiong and Mannering, 2013).
- Application of a multivariate modeling approach in collision analysis at different levels of classification (Aguero-Valverde and Jovanis, 2009; Anastasopoulos et al., 2012b; Bijleveld, 2005; El-Basyouny et al., 2014a, 2014b; Maher, 1990; Ma and Kockelman, 2006; Ma et al., 2008; Park and Lord, 2007, Wang et al., 2011);
- Use of two-state Markov Switching and finite-mixture or latent class models to analyze collision frequencies (Malyshkina et al., 2009; Malyshkina and Mannering, 2010; Park and Lord, 2009; Park et al., 2010; Shaheed and Gkritza, 2014; Zou et al., 2013; 2014;);
- Inclusion of spatial correlation in collision models to capture unobserved effects, as neighboring sites typically have similar environmental and geographical characteristics (Abdel-Aty and Wang, 2006; Aguero-Valverde, 2013; Aguero-Valverde and Jovanis, 2006, 2008, 2010; Amoros et al., 2003; Castro et al., 2012, 2013; Chiou et al., 2014; El-Basyouny and Sayed, 2009c; Flask and Schneider, 2013; Quddus, 2008; Mitra, 2009; Noland and Quddus, 2004).
- Inclusion of temporal correlation in collision models to capture effects due to the collection of collision data over successive time periods (Lord and Persaud, 2000; Wang and Abdel-Aty, 2006; Wang et al., 2006).
- Application of the zero-inflated modeling technique to overcome the excessive zeroes observed in collision data (Shankar et al., 1997; Lord et al., 2005, 2007; Dong et al., 2014).

Due to the fact that random parameters models provide better parameter estimates and inferences compared to traditional fixed parameters models, the use of random parameters in collision modeling has been gaining attention over the past few years. For instance, Milton et al. (2008), Gkritza and Mannering (2008), Anastasopoulos and Mannering (2009, 2011), El-Basyouny and Sayed (2009a), Anastasopoulos et al. (2012a), and Russo et al. (2014) all demonstrated that the random parameters model can provide better inference than the traditional fixed parameters model and can explicitly account for heterogeneity across observations that is due to unobserved road geometrics, traffic characteristics, environmental factors, driver behavior and other confounding factors. Moreover, the random parameters model allows the parameters to vary across observations, which can also capture the variable effect that a parameter may have on the dependent variable. Despite the fact that the random parameters model outperformed traditional fixed parameters models, limited research has used this approach in safety research as random parameters models (i) are very complex to estimate, (ii) are less convenient for engineering purposes, (iii) lack an estimation tool for large samples (Chen and Tarko, 2014) and (iv) lack transferability to other datasets (Lord and Mannering, 2010; Shugan, 2006; Washington et al., 2010).

Most studies in the literature used random parameters in a univariate modeling framework. Regardless of the fact that collision data is multivariate in nature and that it is necessary to account for the likely correlation between collision counts at different levels of classification (Aguero-Valverde and Jovanis, 2009; Bijleveld, 2005; El-Basyouny and Sayed, 2009; El-Basyouny et al., 2014a, 2014b; Maher, 1990; Ma and Kockelman, 2006; Ma et al., 2008; Park and Lord, 2007), multivariate random parameters have rarely been explored in the literature. An earlier study by El-Basyouny and Sayed (2013) used time-varying coefficients (random parameters) in multivariate collision models to identify and prioritize hotspots. Similarly, another study by El-Basyouny et al. (2014a) employed time-varying coefficients in multivariate collision type models to assess the effects of weather elements on seven collision types. A recent study by Dong et al. (2014) demonstrated the use of a multivariate random parameters zero-inflated negative binomial regression (MRZINB) model for jointly modeling collision counts. The authors found that the MRZINB model outperformed the fixed parameters zero-inflated negative binomial regression model and possessed more desirable statistical properties in terms of its ability to accommodate unobserved heterogeneity and excess zero counts in correlated data.

Of the very few multivariate random parameters safety models in the literature, almost all have used heterogeneous effects in addition to random parameters to account for unobserved or unmeasured heterogeneity. The foremost motivations of those studies were to reduce bias and inconsistent estimation, improve the precision of the estimates, and thereby increase the model's predictability. However, when collision data are collected with reference to location (Quddus, 2008), a spatial correlation exists between observations. Ignoring the effect of spatial correlation in collision models may lead to biased estimation of model parameters, as some of the unobserved factors are likely to be correlated with space and there might be some possible correlation among neighboring sites (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; Quddus, 2008; Mitra, 2009; Noland and Quddus, 2004). It might be argued that the significance of spatial correlation is simply an artifact of omitting important variables or inefficient determination of homogeneous road segments. Thus, with appropriate definition and selection of road segments, along with proper selection of pertinent covariates, the spatial correlation would be reduced. In addition, the random parameters and heterogeneous effects most likely can capture the site variation and unobserved or unmeasured heterogeneity, and thereby reduce the effects of spatial correlation. While this point may be partially valid, it will be difficult to find an exhaustive selection of explanatory variables to describe the variability in collision occurrence. Thus,