Accident Analysis and Prevention 100 (2017) 37–43

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

A Bayesian spatial random parameters Tobit model for analyzing crash rates on roadway segments

Qiang Zeng[a,*], Huiying Wen[a], Helai Huang[b], Mohamed Abdel-Aty[c]

[a] School of Civil Engineering and Transportation, South China University of Technology, Guangzhou, Guangdong 510641, PR China
[b] Urban Transport Research Center, School of Traffic and Transportation Engineering, Central South University, Changsha, Hunan 410075, PR China
[c] Department of Civil, Environmental and Construction Engineering, University of Central Florida, Orlando, FL 32816-2450, United States

ARTICLE INFO

Article history:
Received 2 April 2016
Received in revised form 7 December 2016
Accepted 30 December 2016

Keywords:
Crash rate
Tobit model
Spatial correlation
Random parameters
Bayesian inference

ABSTRACT

This study develops a Bayesian spatial random parameters Tobit model to analyze crash rates on road segments, in which both spatial correlation between adjacent sites and unobserved heterogeneity across observations are accounted for. The crash-rate data for a three-year period on road segments within a road network in Florida, are collected to compare the performance of the proposed model with that of a (fixed parameters) Tobit model and a spatial (fixed parameters) Tobit model in the Bayesian context. Significant spatial effect is found in both spatial models and the results of Deviance Information Criteria (DIC) show that the inclusion of spatial correlation in the Tobit regression considerably improves model fit, which indicates the reasonableness of considering cross-segment spatial correlation. The spatial random parameters Tobit regression has lower DIC value than does the spatial Tobit regression, suggesting that accommodating the unobserved heterogeneity is able to further improve model fit when the spatial correlation has been considered. Moreover, the random parameters Tobit model provides a more comprehensive understanding of the effect of speed limit on crash rates than does its fixed parameters counterpart, which suggests that it could be considered as a good alternative for crash rate analysis.

1. Introduction

Given the enormous importance of highway safety, gaining a better understanding of how the probability of crashes is affected by the relevant risk factors has been an area of research focus for a long time, in the hopes that it will provide useful suggestions for laws, regulations and countermeasures aimed at reducing crash occurrence. In most cases, the detailed driving data such as acceleration, braking and steering information, are not available. As a consequence, the relationship between the risk factors and crash frequency, the number of crashes occurring at certain road entities (e.g. road segments or intersections) over some specified periods (e.g. weeks, months or years), is investigated. Because crash frequencies are non-negative integers, statistical count models have been widely employed. Poisson regression is the basic model which assumes crash occurrence to be a Poisson process while requires the mean and variance of crash frequency to be equal (Jovanis and Chang, 1986). To accommodate certain characteristics of crash data, such as over-dispersion, under-dispersion, excess zero observations, spatiotemporal correlation, multilevel structure and unobserved heterogeneity, several Poisson model's variations have been proposed successively, including Poisson-gamma/negative binomial (Miaou, 1994), Poisson-lognormal (Miaou et al., 2005), gamma (Oh et al., 2006), Conway-Maxwell-Poisson (Lord et al., 2008), zero-inflation (Huang and Chin, 2010), generalized estimating equation (Lord and Persaud, 2000), generalized additive (Xie and Zhang, 2008), multilevel (Huang and Abdel-Aty, 2010; Lee et al., 2015), random effects (Shankar et al., 1998), random parameters (Anastasopoulos and Mannering, 2009), finite mixture (Park and Lord, 2009), Markov switching (Malyshkina et al., 2009), latent class (Peng and Lord, 2011), and generalized ordered-response models (Castro et al., 2012). Besides, some artificial intelligence models, such as the neural network (Chang, 2005; Huang et al., 2016; Zeng et al., 2016a, 2016b), Bayesian neural network (Xie et al., 2007), and support vector machine (Li et al., 2008) have also been developed to predict crash frequencies as they exhibit better approximation performance than traditional count models. More detailed descriptions and assessments of these models can be found in the review papers of Lord and Mannering (2010) and Mannering and Bhat (2014).

* Corresponding author.
E-mail addresses: zengqiang@scut.edu.cn, 641459622@qq.com (Q. Zeng), hywen@scut.edu.cn (H. Wen), huanghelai@csu.edu.cn (H. Huang), m.aty@ucf.edu (M. Abdel-Aty).

http://dx.doi.org/10.1016/j.aap.2016.12.023
0001-4575/

From another perspective, in recent years, more and more efforts have been made to develop methods for crash rate analysis which can be deemed as good alternatives to the traditional crash-frequency approaches (Anastasopoulos et al., 2008). Compared with crash count, crash rate is more appealing because it neutralizes the effect of crash exposure, forms a standardized measure of the risk of collision involvement, and may be a more effective criterion used for identifying hotspots (Ma et al., 2015b). Moreover, crash rates are commonly adopted in accident reporting systems. For example, fatality and injury rates per 100 million vehicle miles traveled are used in the annual crash reports of National Highway Traffic Safety Administration (NHTSA, 2012).

Other from crash frequencies (which are discrete integers), crash rates are continuous and non-negative numbers. Zero crash rates may be observed at some sites over finite time duration,[1] resulting in data left-censored at zero. To deal with the censoring problem, Anastasopoulos et al. (2008) first introduced the Tobit model to analyze crash rates. Later on, Anastasopoulos et al. (2012a, b) proposed a random parameters Tobit model to account for unobserved heterogeneity across observations and a multivariate Tobit model for modeling the crash-injury-severity rates simultaneously. Furthermore, a multivariate random-parameters Tobit model was proposed for jointly modeling crash rate by severity (Zeng et al., 2017). A correlated random parameters Tobit model was developed to monitor the interactions between independent variables (Yu et al., 2015), and a random parameters Tobit model with refined-scale panel data was developed to accommodate serial correlation across observations (Ma et al., 2015a). Caliendo et al. (2015) compared the random parameters Tobit regression with the random parameters negative binomial model, and found that the significance of some explanatory variables is not consistent in the two models. In addition, Ma et al. (2015b) advocated a lognormal hurdle model with flexible scale parameter for the purpose of approximating the distribution of crash rates more accurately.

Most of the proposed methods aimed at analyzing crash rates are based on the Tobit regression. However, none of them has accounted for spatial correlation between neighboring sites. In highway safety analysis, spatial correlation is an important issue to be considered, because observation units that are in close proximity may share confounding factors. Recently, significant spatial effects have been found in crash prediction models for road entities (Abdel-Aty and Wang, 2006; Aguero-Valverde and Jovanis, 2008, 2010; Barua et al., 2014, 2016; El-Basyouny and Sayed, 2009; Mitra, 2009; Wang et al., 2017), road network (Zeng and Huang, 2014), regional units (e.g., wards, neighborhoods, counties, traffic analysis zones) (Aguero-Valverde, 2013; Dong et al., 2014, 2015; Noland and Quddus, 2004; Quddus, 2008; Xu et al., 2014; Xu and Huang, 2015; Xu et al., 2017) and injury severity (Castro et al.,, 2013). Congdon (2006) has pointed out that ignoring spatial dependence may lead to underestimation of variability. Moreover, Aguero-Valverde and Jovanis (2008) concluded the advantages of the inclusion of spatial correlation: (1) using spatial correlation, site estimates pool strength from adjacent sites, thereby improving model estimation; (2) spatial dependence can be a surrogate for unknown and related covariates, thus reducing model misspecification; and (3) spatial dependence is able to provide information for grouping sites in corridors for further analysis.

Methodologically, a variety of approaches, ranging from generalized estimation equations (Abdel-Aty and Wang, 2006), simultaneous auto-regressive (Quddus, 2008), conditional auto-regressive (CAR) (Aguero-Valverde and Jovanis, 2008, 2010; Ahmed et al., 2011; Dong et al., 2014, 2015; Mitra, 2009; Quddus, 2008; Siddiqui et al., 2012; Xu et al., 2014) and spatial error model (Quddus, 2008), to multiple membership (El-Basyouny and Sayed, 2009), extended multiple membership (El-Basyouny and Sayed, 2009), geographic weighted regression (Hadayeghi et al., 2003), and geographic weighted Poisson regression (Hadayeghi et al., 2010; Xu and Huang, 2015; Xu et al., 2017), have been proposed to assess spatial effects in crash-frequency data. Among these approaches, CAR prior is the most prevalent for modeling spatial correlation. Moreover, as noted by Quddus (2008), CAR model under the Bayesian framework can lead to more appropriate estimation results than classic spatial models.

In this study, the main objective is to develop a spatial model to analyze crash rates on roadway segments, which can be formulated by incorporating the CAR prior into a Tobit model. To accommodate the unobserved heterogeneity across observations as well, the coefficients of covariates can be further set as random parameters. In order to demonstrate the proposed models, a (fixed parameters) Tobit, a spatial (fixed parameters) Tobit and a spatial random parameters Tobit model are compared in the Bayesian context.

2. Methodology

In this section, firstly, the formulations of the three candidate models for crash rate analysis, Tobit, spatial Tobit and spatial random parameters Tobit regressions, are specified explicitly under the Bayesian framework. Then, a criterion in the context of Bayesian inference, the Deviance Information Criteria (DIC), is introduced for the purpose of model comparison.

2.1. Model specification

2.1.1. Tobit model

Owing to James Tobin (1958), the Tobit model is a regression for modeling the continuous dependent variable which is censored at either a lower threshold (left-censored), an upper threshold (right-censored), or both. Generally, crash rates are left-censored at zero, because crashes may not be reported at some sites during the study period (Anastasopoulos et al., 2008). The Tobit regression for modeling crash rates is expressed as follows:

$$Y_{it}^{*} = \beta_0 + \sum_{m=1}^{M} \beta_m x_{it}^{m} + \varepsilon_{it}, \tag{1}$$

$$Y_{it} = \begin{cases} Y_{it}^{*}, & if Y_{it}^{*} > 0 \\ 0, & if Y_{it}^{*} \le 0 \end{cases}, \quad i = 1, 2, \cdots, N, t = 1, 2, \cdots, T. \tag{2}$$

In the above equations, Y_{it} and x_{it}^{m} are the observed values of crash rate and the mth covariate at site i during period t, respectively. M, N and T are the number of covariates, observed sites and periods respectively. β_0 is the constant, while β_m is the estimable coefficient of the mth covariate. Y_{it}^{*} is a latent variable observed only when positive, and ε_{it} denotes the unstructured error which is assumed to follow independently a normal distribution with zero mean and standard deviation $\sigma (\sigma > 0)$, that is,

$$\varepsilon_{it} \sim normal(0, \sigma^2). \tag{3}$$

[1] This phenomenon may be caused by several reasons. One is simply that there is no crash occurrence at the sites over the observation period. Another is that no injury crashes are not reported when the property damage is not beyond a specific value. Anastasopoulos et al. (2012a, 2012b) illustrated this phenomenon in more detail.