Accident Analysis and Prevention 70 (2014) 320–329

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Multivariate random-parameters zero-inflated negative binomial regression model: An application to estimate crash frequencies at intersections

Chunjiao Dong[a,*], David B. Clarke[a], Xuedong Yan[b], Asad Khattak[c], Baoshan Huang[c]

[a] Center for Transportation Research, The University of Tennessee, 600 Henley Street, Knoxville, TN 37996, USA
[b] MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, School of Traffic & Transportation, Beijing Jiaotong University, Beijing 100044, China
[c] Department of Civil & Environmental Engineering, The University of Tennessee, 319 John D. Tickle Building, Knoxville, TN 37996-2321, USA.

ARTICLE INFO

Article history:
Received 24 December 2013
Received in revised form 23 April 2014
Accepted 23 April 2014
Available online 17 May 2014

Keywords:
Crash frequency
Geometric design
MRZINB model
Full Bayesian method

ABSTRACT

Crash data are collected through police reports and integrated with road inventory data for further analysis. Integrated police reports and inventory data yield correlated multivariate data for roadway entities (e.g., segments or intersections). Analysis of such data reveals important relationships that can help focus on high-risk situations and coming up with safety countermeasures. To understand relationships between crash frequencies and associated variables, while taking full advantage of the available data, multivariate random-parameters models are appropriate since they can simultaneously consider the correlation among the specific crash types and account for unobserved heterogeneity. However, a key issue that arises with correlated multivariate data is the number of crash-free samples increases, as crash counts have many categories. In this paper, we describe a multivariate random-parameters zero-inflated negative binomial (MRZINB) regression model for jointly modeling crash counts. The full Bayesian method is employed to estimate the model parameters. Crash frequencies at urban signalized intersections in Tennessee are analyzed. The paper investigates the performance of MZINB and MRZINB regression models in establishing the relationship between crash frequencies, pavement conditions, traffic factors, and geometric design features of roadway intersections. Compared to the MZINB model, the MRZINB model identifies additional statistically significant factors and provides better goodness of fit in developing the relationships. The empirical results show that MRZINB model possesses most of the desirable statistical properties in terms of its ability to accommodate unobserved heterogeneity and excess zero counts in correlated data. Notably, in the random-parameters MZINB model, the estimated parameters vary significantly across intersections for different crash types.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Traffic crashes at urban intersections place a huge burden on society through death, injury, lost productivity, and property damage. Crash frequency at roadway intersections has been increasingly studied in recent years. Several studies (see, for example, Miaou, 1994; Lord et al., 2005; Caliendo et al., 2007; Ye et al., 2009; Anastasopoulos and Mannering, 2009; for a complete review of this literature see Lord and Mannering, 2010) have examined the number of crashes occurring at an intersection as a function of intersection geometric features and traffic factors. The study of crash frequencies has long matured in the field of univariate count models, and many variants of the approaches mentioned above are already used extensively for univariate count data. However, creating general specifications of the univariate count model is a problem for modeling specific types of crash counts (for example, the number of crashes resulting in fatalities, injuries, etc.). In fact, crash frequency data are multivariate in nature and correlated. Univariate count model has not been the case for correlated crash count data, especially for general dependency structures with more than two correlated crash counts.

In such context, one may consider a simple Poisson, negative binomial (NB), or Poisson-lognormal discrete distribution, and develop multivariate versions of these discrete distributions to accommodate correlated counts (Ma and Kockelman, 2006; Park and Lord, 2007; Ma et al., 2008; El-Basyouny and Sayed, 2009a).

* Corresponding author. Tel.: +1 8659743526.
E-mail address: CDONG5@UTK.EDU (C. Dong).

http://dx.doi.org/10.1016/j.aap.2014.04.018
0001-4575/© 2014 Elsevier Ltd. All rights reserved.

These multivariate models are useful in crash-frequency modeling, since they explicitly consider the correlation among the specific crash types of crash counts for each roadway entity (Miaou and Song, 2005; Bijleveld, 2005; Song et al., 2006). However, those multivariate models do have their limitations, most notably their inability to handle excess zero counts, which is a phenomenon when crash counts have many categories because some roadway entities have no crashes reported during the analysis period. They also lack the desirable ability to account for unobserved heterogeneity across roadway entities.

Two recent papers have described analytical method development. Anastasopoulos et al. (2012a) estimated a multivariate tobit model that addresses the possibility of differential censoring across injury-severity levels. Castro et al. (2012) proposed an equivalent latent variable-based generalized ordered response framework for count data models. Their formulations allow handling excess zeros in correlation count data. Comparing those studies, one could conclude that the multivariate model—with its properties of handling extra zeros and accounting for unobserved heterogeneity—has the potential to provide a fuller understanding of the factors affecting crash frequencies.

In contrast to those two existing studies, which address the two challenges discussed above for correlated crash counts, our approach relies on zero-inflated NB (ZINB) model with random parameters. The ZINB model has proved useful for modeling outcomes with numerous zeros. It operates on the principle that the excess zero density that cannot be accommodated by a conventional count structure is accounted by a splitting regime that models a crash-free versus a crash-prone propensity of a roadway entity. The probability of a roadway entity being in zero or non-zero states can be determined by a binary logit or probit model (Lambert, 1992; Lee and Mannering, 2002; Kumara and Chin, 2003). The multivariate ZINB (MZINB) model is potentially an alternate for jointly modeling correlated crash frequency data with excess zeros. However, this model's ability to account for unobserved heterogeneity is limited because it assumes that the parameter estimates are fixed over roadway-entity observations.

Many crash count model application studies assume that parameters are completely fixed, and do not consider unobserved heterogeneity across roadway entities by incorporating random-parameters. However, in the presence of unobserved heterogeneity, such a fixed-parameter approach could result in biased parameter estimations and incorrect inferences (Washington et al., 2011). On the other hand, some researchers use random-effects models (Milton et al., 2008; Shankar et al., 1998; Miaou et al., 2003; Guo et al., 2010; Christoforoua et al., 2010, 2011) to consider unobserved heterogeneity, but in a rather coarse and restrictive form where the common unobserved effects are assumed to be distributed over the spatial/temporal units according to some distribution, and shared unobserved effects are assumed to be uncorrelated with explanatory variables. To account for the potential unobserved heterogeneity issues associated with crash-frequency data, recent research (Anastasopoulos and Mannering, 2009; EI-Basyouny and Sayed, 2009b; Anastasopoulos et al., 2012b) advocates the use of count models with random parameters as an alternate approach. Random-parameters models can be viewed as an extension of random-effects models. However, rather than effectively influencing only the intercept of the model, random-parameters models allow each estimated parameter of the model to vary across individual observation in the dataset. These models attempt to account for the unobserved heterogeneity from one intersection to another. Compared to the conventional crash prediction model, which fits one regression model to the dataset, the random-parameters approach develops different regression models for individual sites. The possibility of accounting for heterogeneity by allowing some or all parameters to vary across roadway entities has considerable potential.

In this paper, we develop a multivariate random-parameters ZINB (MRZINB) regression model to account for unobserved heterogeneity, which the conventional multivariate zero-inflated models cannot address. Since our formulation is based on the MZINB method, it also has the property of accommodating excess zeros in correlated count data. The results of a review of the work performed on the application of zero-inflated model in traffic safety show that the zero-inflated models outperform other models when the zero counts are over 65% in the data. Our data show that the proportion of crash-free sample in Tennessee is 40.50%, 82.85%, and 95.12% for car-only, car–truck, and truck-only crashes, respectively. As seen from the data, although the zero counts are not totally over 65%, they cannot be handled by a normal Poisson or Poisson-gamma process either. Because of this, a MRZINB model needs to be developed for analysis of crash counts across vehicle type. The full Bayesian method is employed to estimate the model parameters. We apply the modeling framework to estimate crash frequencies at urban signalized intersections in Tennessee. For model performance evaluation, the MZINB and multivariate NB (MVNB) models are employed as the comparison models. Two aspects of model performances, including significant factors identifying ability and model goodness of fit have been examined. In addition, the paper investigates the performances of MRZINB and MZINB regression models in establishing the relationship between crashes, pavement conditions, traffic factors, and geometric design of roadway intersections. The primary objective of the model application is to investigate which factors contribute significantly to the crash counts across vehicle types. The secondary objective is to examine if there is any difference in the specific types of crashes for the same factor and how to control the factor to reduce higher crash frequencies for a certain crash type while other types of crashes are not an issue.

2. Model structure and estimation

The following section presents the general forms of MRZINB regression models and provides brief descriptions of its estimation procedures. In the regression setting, the objectives are to identify significant factors influencing the zero-inflation incidence and to determine the extent of the effects of geometric design features, pavement characteristics, and traffic factors on the mean events.

2.1. MZINB distributions

We use a mixture of distributions to construct a MZINB model (Li et al., 1999). These distributions include:

(a) a point mass at 0 (crash-free), the probability mass function (pmf) is

$$P(Y_1=0,\ \ldots,\ Y_m=0)=p_0+p_1\exp(-\lambda_1)+p_2\exp(-\lambda_2)+\cdots+p_m\exp(-\lambda_m)+p_{11}\exp(-\lambda) \tag{1}$$

where Y_i is the number of crashes for crash type i. Note that Y_1, $Y_2, \ldots$, and Y_m can be represented by

$$Y_1=U_1+U_0,\quad Y_2=U_2+U_0,\quad \ldots,\quad Y_m=U_m+U_0 \tag{2}$$

where $p_0+p_1+p_2+\cdots+p_m+p_{11}=1$ and λ_{10}, $\lambda_{20}, \ldots, \lambda_{m0}$, and λ_{00} are the means of $U_1, U_2, \ldots, U_m$, and U_0, respectively. Eq. (2) contains $2(m+1)$ parameters, which increase linearly with m. With the further assumption that $\lambda_1=\lambda_{10}+\lambda_{00}$, $\lambda_2=\lambda_{20}+\lambda_{00}$,

Download English Version:

https://daneshyari.com/en/article/572329

Download Persian Version:

https://daneshyari.com/article/572329

Daneshyari.com