Available online at www.sciencedirect.com

ScienceDirect

Procedia - Social and Behavioral Sciences 138 (2014) 452 – 459

Procedia
Social and Behavioral Sciences

The 9th International Conference on Traffic & Transportation Studies (ICTTS'2014)

Modeling Zonal Traffic Accident Counts with the Regression Under Zero-adjusted Inverse Gaussian Distribution

Lu Ma*, Xuedong Yan

MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, No.3 Shangyuancun, Haidian District, Beijing 100044, P.R. China

Abstract

Statistical analyses on crash counts/rates are important steps for transportation safety planning and evaluations of certain transportation facilities. Even there are various types of regression models in connecting between the response variable and characteristics of transportation facilities in relevance with considerations of different aspects of the frequency distribution of crashes, due to the complicated data generating process, past models still lack sufficient precision for accounting for several issues of crash counts especially the widely mentioned problem of excessive zero counts of accidents. Towards to this end, this study seeks to provide an alternative approach for predicating and inferring the response variable through a set of explanatory variables. As the crash rates fundamentally follow a mixed distribution such that non-zero probability will be associated with zero point in the domain, the zero-adjusted inverse Gaussian regression is introduced to overcome these issues as well as provide new thoughts in this field. Based on an observed dataset of accidents aggregated for the level of traffic analysis zone (TAZ), the regression model is developed and estimated using the principle of maximum likelihood estimation. The results discover that several attributes related to the zonal social economic, average traffic condition as well as roadway geometric design characteristics are statistically significant. Under such empirical analysis, the zero-adjusted inverse Gaussian distribution is appropriate to reflect the distributional characteristics of crash rates and provide a new direction of approaches in the field of crash data analyses.

Peer-review under responsibility of Beijing Jiaotong University(BJU), Systems Engineering Society of China (SESC).

Keywords: Crash data; Safety evaluation; zero-adjusted inverse Gaussian; Regression model; Traffic analysis zone

* Corresponding author. Tel.: +86-(0)10-5168-4599; Fax: +86-(0)10-5184-0080.
E-mail address: lma@bjtu.edu.cn .

Peer-review under responsibility of Beijing Jiaotong University(BJU), Systems Engineering Society of China (SESC).
doi:10.1016/j.sbspro.2014.07.224

1. Introduction

During the process of safety planning as well as the risk evaluation of transportation facilities, it is necessary to understand the relationships between the chance of traffic accident occurrence and characteristics related to the facilities. Toward to this end, statistical regression models are usually adopted to connect variables to the crash counts or crash rates associated to certain transportation facilities, including road segments, intersections (Vogt and Bared, 1998) and even recently emphasize traffic analysis zones (TAZs). And these important explanatory variables could be in the aspects of traffic flow conditions, climates, social economic and demographic characteristics, pavement conditions and roadway geometric designs. Past studies have thoroughly modeling such connections between crash counts/rates and explanatory variables using count models in which the distribution of accident frequency are treated in terms of counting process and correspondingly Poisson distribution is considered in earlier literatures (Nicholson and Wong, 1993).

Empirical analyses also indicate two important aspects of traffic accident counts, namely the usually observed excessive zero counts (Lord et al., 2005) and the over-dispersed distributional properties (Miaou, 1994). In accommodating to capture the over-dispersed crash counts, negative binominal distribution allows an additional dispersion parameter to provide more flexibility to extract information of counting dispersion and therefore negative binominal regression models (Miaou, 1994) are widely used to modeling the crash counts. With the consideration of previous mentioned issue of excessive zero counts, the zero inflated models (Lord et al., 2005) can be adopted to measure the chance of non-crash observations by allowing two sources of probability of zero into the distribution. Therefore, zero-inflated negative binomial models are major techniques in the field of crash data analysis. Furthermore, advanced statistical models are continuously developed to enhance the understanding of the distributional characteristics of crash counts. For example, multivariate response models (Ma and Kockelman, 2006; Anastasopoulos et al., 2012; Chiou and Fu, 2013) are raised to account for the correlative relationships among different severity-types of accidents and nonparametric analyses (Xie and Zhang, 2008) are also performed in order to provide flexible forms for explanatory variables with continuously domain of distributions.

Other than the considerations of zero inflated negative binomial distributions, many other distributions (Lord and Mannering, 2010 for detailed information) are recently proposed to improve the modeling precision as well as the reliability of inference. Several recent studies (Lord and Geedipally, 2011; Geedipally et al., 2012) also indicated that the traditional zero-inflated types of regression models still have drawbacks in accounting for the nature of excessive zero counts of accidents due to the essentially complicated data generating process and issues related to the collection of the data. Therefore, it is necessary to test alternative advanced approaches for measuring crash count distributions with more accurate specifications as well as comprehensive interpretations.

This study as a result proposed a zero-adjusted inverse Gaussian distribution to modeling crash counts. First of all, zero-adjusted inverse Gaussian is continuously distributed from zero to infinity with non-zero probability on the domain of zero. Secondly, the zero-adjusted inverse Gaussian is capable and appropriate to reflect especially crash rate as it is continuously distributed. Further, the proposed distribution have two parameters, namely the mean parameter and shape parameter and therefore it is flexible and possibly to provide a structure for mining more information from the data. In sum, the study proposed an alternate choice of the distributional structure for modeling crash counts/rates through regression models with zero-adjusted inverse Gaussian formulations. The following analysis will also use an observed crashed dataset to perform empirical analyses in additional to current understanding of the relationship between crash occurrence and attributes related to transportation facilities.

2. About the Data

This research adopts data which was collected in Pikes Peak Area, Colorado, and three sources of information (the TAZ dataset, the traffic and roadway dataset, and the accident datasets) are integrated to construct the finally used cross-sectional data format. The TAZ dataset provides information on the geographic boundaries for each traffic analysis zone as well as zonal attributes including population, number of household and number of workers.

In the following modeling analysis, traffic and roadway related attributes including number of lanes, roadway length, and traffic characteristics are extracted from traffic and roadway datasets. Finally all these attributes are