Analytic Methods in Accident Research 11 (2016) 48–61

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

Modeling the equivalent property damage only crash rate for road segments using the hurdle regression framework

Lu Ma, Xuedong Yan *, Chong Wei, Jiangfeng Wang

MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, PR China

ARTICLE INFO

Article history:
Received 27 August 2015
Received in revised form
4 July 2016
Accepted 4 July 2016
Available online 12 August 2016

Keywords:
Equivalent-property-damage-only rate
Hurdle model
Tobit model
Mixed distribution
Accident loss

ABSTRACT

The understanding of the distributional characteristics of the equivalent property damage only (EPDO) crash rate is limited in the existing literature. Models without a proper distribution of EPDO rate could result in biased estimations and misinterpretations of factors. The importance of prediction accuracy and modeling performance for the EPDO rate should be acknowledged since they directly affect the allocation of limited public funds to safety management for road networks. The general objective of this study is to investigate the distributional characteristics of the EPDO rate and accordingly develop proper econometric models for connecting the EPDO rate to explanatory variables. A hurdle framework was proposed in order to accommodate the zero-positive mixed domain of the EPDO rate. For the positive part of the EPDO rate, three representative distributions (lognormal, gamma and normal) were tested and then the three hurdle models were compared against the Tobit model and the random-parameters Tobit model. The empirical results illustrate the lognormal hurdle model's superior modeling performance in comparison to the other four models, and more importantly that conclusion also holds for several different definitions of the EPDO rate under different combinations of property damage only (PDO) equivalency factors.

1. Introduction

The frequency and rate of traffic accidents have been identified as important measures for the safety management of transportation facilities. Many studies have contributed to this field by developing statistical models connecting crash counts to influencing factors including geometric design features and traffic operation characteristics of road segments, and socioeconomic and demographic attributes of traffic analysis zones. The crash frequency and the crash rate have been studied using various types of count models (see studies by Lord and Mannering (2010), Mannering and Bhat (2014), Anastasopoulos et al. (2008, 2012a, 2012b), Gregoriades and Mouskos (2013), Lee et al. (2015), Ma et al. (2015) for more detailed discussions).

Nevertheless, as recently emphasized (Ma et al., 2014; Washington et al., 2014), neither the crash frequency nor the crash rate can accurately reflect the magnitude of traffic risk, because accidents range widely in terms of their severity levels, which should be an important consideration in the conducting of safety evaluations. The single value of the crash frequency

* Corresponding author.
E-mail addresses: lma@bjtu.edu.cn (L. Ma), xdyan@bjtu.edu.cn (X. Yan), chwei@bjtu.edu.cn (C. Wei), wangjiangfeng@bjtu.edu.cn (J. Wang).

http://dx.doi.org/10.1016/j.amar.2016.07.001
2213-6657/

or crash rate measures the chance of accident occurrence in a specific transportation facility but lacks information on the severity and losses caused by each accident. Therefore, it is necessary to use a criterion that is capable of comprehensively integrating the probability of accident occurrence and the likely severity.

An important approach for combining the information on crash severity with that on crash count is based on equivalent property damage only (EPDO) crashes, in which property damage only crashes – also called "no injury" crashes – are treated as one unit and other types of crashes are weighted with respect to their severity levels. EPDO crashes can also be normalized by certain exposure variables to form the EPDO crash rate, which more clearly reflects the risk of harm due to an accident. EPDO and EPDO rates have been applied in practice and in studies for a long time (e.g. Harkey, 1999; Hunter et al., 2001; PIARC et al., 2004; Campbell and Knapp, 2005; HRPDC, 2006; Felsburg Holt and Ullevig, 2008; Rifaat et al., 2010; Oh et al., 2010; Montella, 2010; UMassSafe, 2011; Boudreau, 2014; Washington et al., 2014). Among these works, a few have focused on statistical characteristics of EPDO-related quantities. For the EPDO crash count, Oh et al. (2010) adopted a negative binomial regression model and Washington et al. (2014) used a nonparametric quartile regression. However, some of the distributional assumptions might still be questionable (Oh et al., 2010). For the EPDO rate, there is still a lack of both an understanding of its distributional characteristics, and analytic models of the statistical analyses. To this end, this study was motivated by the desire to contribute by investigating the distributional characteristics of the EPDO rate and then developing appropriate statistical models.

Generally, the hurdle model is an appropriate approach if the response variable has two different states whose statistical behaviors can be treated hierarchically. For count data modeling, the zero counts can be treated separately from all other positive counts due to the excessive observations of zeros. Similarly, for modeling the EPDO rate, zero observations require special treatment. Thus, the hurdle model is a potential approach for these situations. As censored data, the EPDO rate can also be modeled using the Tobit model. Although, the hurdle model's feature allowing the two states to have different parameters gives it an advantage by enhancing its flexibility, the more essential motivation for this study is that the hurdle model provides a framework allowing almost any choice of distributional assumption for the positive part of the EPDO rate. However, such a property is not available in the Tobit model. Therefore, the hurdle model enables examinations of the performance of different distributions for modeling the positive EPDO rate, whereas the Tobit model can only use the truncated normal distribution.

According to the definition of risks in decision theory (Lehmann and Casella, 1998), the expected EPDO rate is the risk of losses caused by crashes occurring in certain transportation facilities. The PDO equivalency factors are values reflecting the average magnitude of losses due to different accidents, and they might vary according to the types of losses selected. According to the existing literature, the PDO equivalency factors can be societal crash costs (e.g. Oh et al., 2010; Washington et al., 2014), economic losses (e.g. Montella, 2010) or other types (HRPDC, 2006; Felsburg Holt and Ullevig, 2008; UMassSafe, 2011). Therefore, choices of PDO equivalency factors can differ dramatically. Therefore, it would be interesting to understand the influence of different PDO equivalency factors on the performance of different models, something that has generally been missing from previous studies. Hence, another objective of this study is to inspect the impact of different combinations of PDO equivalency factors on the selection of preferred econometric models.

In the following sections, the distributional characteristics of the EPDO rate are discussed first. Then, several competitive econometric structures are introduced and applied to an observed accident dataset using the proposed hurdle-type models (lognormal, gamma and normal distributions for the positive domain) and two Tobit-type models (Tobit and random-parameters Tobit). Next, the models are applied to several different definitions of EPDO rates using PDO equivalency factors from previous studies and practical applications.

2. Method

2.1. An overview of the distribution of the equivalent-property-damage-only rate

A general definition of the EPDO rate is presented in Eq. (1), where N_i is the number of accidents of the ith type of severity, and L_i is the corresponding PDO equivalency factor. Normalizing the weighted sum by certain exposure quantities gives the EPDO rate:

$$EPDO\ rate = \frac{EPDO}{exposure} = \frac{L_1N_1 + L_2N_2 + L_3N_3 \ldots}{exposure} \tag{1}$$

In order to identify an appropriate regression model, it is necessary to understand the distributional characteristics of the EPDO rate. Here, the number of accidents of each severity type could follow some distribution, e.g. Poisson or negative binomial, but the distributional form of the EPDO rate is intractable due to the complicated correlative relationships between the crash counts for different severity levels and the exposure variables. One practical way to deal with such a problem is to try to find approximate parametric distributions.

According to the definition, the EPDO rate is continuous and nonnegative and equals zero if no accident is observed during the study period. Therefore, a certain amount of probability is concentrated on zero EPDO rates, which might be considerable if excessive zero counts of accidents are present. The positive part of the EPDO rate is continuously distributed,