Analytic Methods in Accident Research 3-4 (2014) 1–10

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

A two-stage bivariate logistic-Tobit model for the safety analysis of signalized intersections

Xuecai Xu [a,b,*], S.C. Wong [b], Keechoo Choi [c]

[a] School of Civil Engineering and Mechanics, Huazhong University of Science and Technology, Wuhan, China
[b] Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, China
[c] Department of Transportation Engineering, TOD-based Sustainable Urban Transportation Center, Ajou University, Korea

ARTICLE INFO

Article history:
Received 24 October 2013
Received in revised form
28 August 2014
Accepted 28 August 2014
Available online 31 October 2014

Keywords:
Signalized intersection
Crash severity
Crash risk
Bivariate Logistic–Tobit model

ABSTRACT

Crash frequency and crash severity models have explored the factors that influence intersection safety. However, most of these models address the frequency and severity independently, and miss the correlations between crash frequency models at different crash severity levels. We develop a two-stage bivariate logistic-Tobit model of the crash severity and crash risk at different severity levels. The first stage uses a binary logistic model to determine the overall crash severity level. The second stage develops a bivariate Tobit model to simultaneously evaluate the risk of a crash resulting in a slight injury and the risk of a crash resulting in a kill or serious injury (KSI). The model uses 420 observations from 262 signalized intersections in the Hong Kong metropolitan area, integrated with information on the traffic flow, geometric road design, road environment, traffic control and any crashes that occurred during 2002 and 2003. The results obtained from the first-stage binary logistic model indicate that the overall crash severity level is significantly influenced by the annual average daily traffic and number of pedestrian crossings. The results obtained from the second-stage bivariate Tobit model indicate that the factor that significantly influences the numbers of both slight injury and KSI crashes is the proportion of commercial vehicles. The existence of four or more approaches, the reciprocal of the average turning radius and the presence of a turning pocket increase the likelihood of slight injury crashes. The average lane width and cycle time affect the likelihood of KSI crashes. A comparison with existing approaches suggests that the bivariate logistic-Tobit model provides a good statistical fit and offers an effective alternative method for evaluating the safety performance at signalized intersections.

1. Introduction

A number of different approaches and perspectives have been used in crash prediction modeling. Lord and Mannering (2010) provided a comprehensive review of the different methodological approaches to crash frequency modeling, such as Poisson, negative-binomial, Poisson-lognormal, zero-inflated count, Conway–Maxwell–Poisson, gamma, generalized estimating equations, generalized additive, random-effects, negative-multinomial, random-parameter count, finite-mixture and Markov-switching models, and other intelligent algorithms. Savolainen et al. (2011) described modeling crash injury

* Corresponding author.
E-mail address: xuecai_xu@hust.edu.cn (X. Xu).

http://dx.doi.org/10.1016/j.amar.2014.08.001
2213-6657/

severity using artificial neural networks, Bayesian hierarchical binomial logit, Bayesian-ordered logistic, bivariate binary/ordered logistic, classification and regression tree, generalized ordered logit, Markov-switching multinomial logit, mixed generalized ordered logit, multivariate logit/logistic, nested logit and ordered logit/logistic models.

Models have been developed to assess intersection safety at signalized intersections in terms of either the crash frequency or crash severity. Wong et al. (2007), on whose research this study is based, used Poisson and negative-binomial regression models to quantify the influence of factors contributing to the incidence of slight injury crashes and the incidence of crashes resulting in a kill or serious injury (KSI) in Hong Kong. Liu (2007) generated a back-propagation neural network model using crash records from 62 signalized intersections in Taiwan and the characteristics of those intersections. The results indicated that the effects of the variables on the number of intersection-related crashes varied between intersections, leading to the proposal of a decision-making scheme to prevent erroneous investments. Ye et al. (2009) focused on collision types at rural intersections in Georgia and explored the crash frequency using multivariate Poisson models structured by simultaneous equations. This approach provided new insights into the crash frequency, although the effects of the risk factors that it addressed were found to be modest. Obeng (2011) analyzed the crash severity at signalized intersections using separate ordered logit models for females and males to investigate gender differences. The results indicated that the effects of driving conditions, type of crash, type of vehicle driven and vehicle-safety features on the risk of severe injury varied according to the gender of the driver. Haque et al. (2010) constructed Bayesian hierarchical models to examine motorcycle crashes at four-legged and “T” signalized intersections and found that the significant risk factors differed at the two intersections. In a similar study, Xie et al. (2013) used Bayesian hierarchical negative binomial models to evaluate the safety of signalized intersections in Shanghai at the intersection and corridor levels.

Several studies have investigated the heterogeneity of signalized intersections. For instance, Karlaftis and Tarko (1998) used negative binomial models and cluster analysis to explore the relationships between the crash frequency and possible influencing factors. However, the approach to accidents was quite general and the specific case of signalized intersections was not addressed. Chin and Quddus (2003) used the random-effects negative binomial model to identify the elements affecting intersection safety in Singapore, attempting to address the heterogeneity problem. Wang and Abdel-Aty (2006) and Wang et al. (2007) investigated the rear-end crash frequency at signalized intersections using the generalized estimating equations approach to account for temporal or spatial correlations within the dataset, which required that the same correlation matrix be used for different corridors. Guo et al. (2010) integrated the Poisson and negative binomial models with a Bayesian approach to evaluate the intersection safety with reference to corridor-level spatial correlations between 170 signalized intersections in Florida. The results indicated that the Poisson spatial model provided the best model fit and that its performance was related to the proximity function. However, the performance of alternative functions, such as the exponential function, should still be investigated. More recently, Castro et al. (2012) reformulated count modeling as a special case of generalized ordered-response modeling to address intersections. They presented a flexible count model, one accommodating temporal effects and the other accommodating both temporal and spatial effects. These models addressed temporal and spatial correlations and provided a fairly generalized method of crash analysis, which can be developed to accommodate more specific cases in future research.

However, these studies concentrated on either the crash frequency or crash severity at signalized intersections, and the possible correlations between their model estimates at different crash severity levels were not considered, which may have led to bias in the estimates (Lord and Mannering, 2010). Other studies have dealt simultaneously with the crash frequency and severity, using such methods as multi-level hierarchical structures (Kim et al., 2007), simultaneous equations (Kim and Washington, 2006) and multivariate analysis (Ma and Kockelman, 2006). These approaches either integrated crash frequency and crash severity models or involved a two-stage model.

For instance, Abdel-Aty and Keller (2005) explored overall crash severity levels using an ordered logistic model and specific crash severity levels using a hierarchical tree-based regression model. Their results showed that the aggregation of crash types was a less effective method than the development of separate models for each level of collision, an insight that has informed the design of this study. However, it should be noted that the two models presented were kept relatively separate, with no interaction permitted. Pei et al. (2011, 2012) developed a joint-probability model to integrate crash occurrence prediction and crash severity prediction within a single framework and used the Markov-chain Monte Carlo approach to establish a full Bayesian estimate of the effects of the explanatory factors. The results indicated that the proposed model was appropriate for signalized intersections and roadway safety, but only the binary approach to crash severity was provided as an illustrative example. El-Basyouny and Sayed (2011) used a multivariate Poisson-lognormal intervention model for the analysis of crash counts by severity level, and extended the model to incorporate random parameters to account for the correlation between sites. Chiou and Fu (2013) addressed the crash frequency and severity simultaneously in an integrated model with a multinomial generalized Poisson structure. The proposed covariance structure was shown to enhance the model's performance.

Wang et al. (2011) used the less-common two-stage model approach to model the crash frequency at different severity levels. They proposed a two-stage mixed multivariate model and showed how disaggregated data at the level of individual accident could be used to predict a certain type of low-frequency accident. Bhat et al. (2014) formulated a count outcome model with multinomial probit selection that accommodates unobserved heterogeneity and endogeneity issues at intersections. Their results showed that the model can be used for intersection crash analysis.

Crashes do not occur at every roadway segment and intersection during a particular observation period. Crash or crash-rate data can therefore be considered left-censored at zero, in accordance with the requirements of the Tobit model, which has been used by previous scholars to address the issue of safety. For example, Obeng and Burkey (2006) used a Tobit model