Accident Analysis and Prevention 57 (2013) 140–149

Contents lists available at SciVerse ScienceDirect

Accident Analysis and Prevention

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

A simultaneous equations model of crash frequency by severity level for freeway sections

Xin Ye[a,*], Ram M. Pendyala[b,1], Venky Shankar[c,2], Karthik C. Konduri[d,3]

[a] Civil Engineering Department, California State Polytechnic University, Room 17-2679, 3801 West Temple Avenue, Pomona, CA 91768, United States
[b] Department of Civil and Environmental Engineering, Arizona State University, Room ECG252, Tempe, AZ 85287-5306, United States
[c] Department of Civil and Environmental Engineering, The Pennsylvania State University, 212 Sackett Building, University Park, PA 16802, United States
[d] Department of Civil and Environmental Engineering, University of Connecticut, 261 Glenbrook Road, Unit 3037, Storrs, CT 06269-3037, United States

ARTICLE INFO

Article history:
Received 19 October 2012
Received in revised form 25 February 2013
Accepted 21 March 2013

Keywords:
Crash severity
Crash frequency
Simultaneous equations model
Freeway safety
Maximum simulated likelihood estimation
Multivariate Poisson regression model

ABSTRACT

This paper presents a simultaneous equations model of crash frequencies by severity level for freeway sections using five-year crash severity frequency data for 275 multilane freeway segments in the State of Washington. Crash severity is a subject of much interest in the context of freeway safety due to higher speeds of travel on freeways and the desire of transportation professionals to implement measures that could potentially reduce crash severity on such facilities. This paper applies a joint Poisson regression model with multivariate normal heterogeneities using the method of Maximum Simulated Likelihood Estimation (MSLE). MSLE serves as a computationally viable alternative to the Bayesian approach that has been adopted in the literature for estimating multivariate simultaneous equations models of crash frequencies. The empirical results presented in this paper suggest the presence of statistically significant error correlations across crash frequencies by severity level. The significant error correlations point to the presence of common unobserved factors related to driver behavior and roadway, traffic and environmental characteristics that influence crash frequencies of different severity levels. It is found that the joint Poisson regression model can improve the efficiency of most model coefficient estimators by reducing their standard deviations. In addition, the empirical results show that observed factors generally do not have the same impact on crash frequencies at different levels of severity.

1. Introduction

Recent work on the simultaneous evaluation of frequency and severity of crashes has indicated that the body of empirical evidence on influential factors is still developing. In particular, the effect of geometrics in the unconditional context (via the simultaneous accommodation of crash frequency and the severity of crashes) is a matter of substantial focus. The seminal papers by Lord and Mannering (2010) and Savolainen et al. (2011) point to the need for more empirical work, since the typology of models is potentially exhaustive. At the very least, the papers seem to suggest that joint estimation of severity and frequency would shed further light into the "efficiency" aspects of statistical parameters associated with such influential factors as geometry and traffic volume. The problem of loss of "efficiency" of parameters when correlations among unobserved factors are ignored in simultaneous equations model systems is well known in the statistical and econometric literature. Yet, the computational aspects of addressing this problem via full-information estimation methods present considerable challenges to the development of empirical models.

This paper is aimed at making a contribution in this area by presenting a model of freeway crash frequency by severity level for 275 freeway sections in the State of Washington. Five year freeway section crash counts (by severity level) are available for the 275 freeway sections and a modeling methodology that can simultaneously account for observed and unobserved factors contributing to crash frequencies by severity level is developed and presented in this paper. The work presented in this paper is motivated by recent studies in the literature along this line of inquiry. Wang et al. (2011) present a model aimed at ranking sites with respect to safety using a two-stage modeling approach. Their objective is to obtain insights into frequency and severity simultaneously by accounting for the effects of spatial correlations in crash occurrence, as well as heterogeneity in severity, on statistical parameters. This particular study is highly insightful in that it points to a ripe area in the domain of

* Corresponding author. Tel.: +1 909 869 3444; fax: +1 909 869 4342.
E-mail addresses: ye@csupomona.edu (X. Ye), ram.pendyala@asu.edu (R.M. Pendyala), shankarv@engr.psu.edu (V. Shankar), kkonduri@engr.uconn.edu (K.C. Konduri).
[1] Tel.: +1 480 727 9164; fax: +1 480 965 0557.
[2] Tel.: +1 814 865 9434; fax: +1 814 863 7304.
[3] Tel.: +1 860 486 2733; fax: +1 860 486 2298.

0001-4575/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.aap.2013.03.025

site ranking – an area that has historically been dominated by single equation methods primarily. Pei et al. (2011) adopted a Bayesian approach to the joint modeling of crash frequency and severity via a Poisson log-normal prediction. Their paper serves as a useful precedent to this study in the sense that it examines crash frequency and severity of signalized intersections as an example application. Quite appropriately, the authors explore the effect of correlations due to the multi-approach effects at signals; however, the insights from this study are limited to signalized intersection contexts. The examination of crash frequency by severity in other empirical contexts still requires attention. The nature of correlations across crash types in the signalized intersection context can be substantially different when compared with other highway contexts, due to the different degrees of exposure to crash types, and therefore, crash severities. El-Basouny and Sayed (2009) make the case for the significance of outcome correlations using a multivariate Poisson log-normal specification albeit in a Bayesian sense. The authors argue that correlations between property damage and injury rates can be significant, an issue that requires further exploration. The authors note that the degree of correlation does have the potential to substantially affect parameter estimates and model inferences, which would then affect the accuracy of hazardous location identification.

Modeling crash frequencies by type of crash (angle, head-on, rear-end, etc.), number of vehicles involved (single-vehicle, two-vehicle, multi-vehicle, etc.), and severity level (property damage only, possible injury, incapacitating injury, etc.) has been the subject of much research in the transportation safety arena. Researchers have employed a variety of single equation count data models including Poisson models, Negative Binomial models, and zero-inflated versions of these count models (to account for the presence of zero crash frequency counts in the data set that may be due to facilities being truly safe or simply due to the limited window of observation for which crash frequency data is collected) to model crash frequency. These single equation models have often been employed to identify factors contributing to total crash frequencies on highway facilities. These methods can also be employed to model crash frequencies by severity level (e.g., modeling the number of fatal crashes as a function of roadway and traffic characteristics, environmental conditions, etc.). Although modeling crash frequency by severity level using single-equation methods (sets of independent equations) can offer valuable insights into factors affecting crash frequencies, the fact that such model systems ignore the simultaneity that may be prevalent in the safety phenomenon under investigation is an issue that merits being addressed. A model system in this paper refers to a series of equations to model a number of dependent variables that are mutually interrelated. There may be a host of unobserved factors related to driver characteristics, vehicular characteristics, roadway and traffic characteristics that contribute to crash frequencies at various severity levels. In a single equation method, the random error component (and the constant term) may be viewed as capturing the effects of these unobserved factors. However, simultaneity may arise in crash frequency modeling due to the possibility that unobserved factors affecting crash frequency at one severity level may be correlated significantly with unobserved factors affecting crash frequency at another severity level. This possibility calls for the deployment of simultaneous equations modeling methodologies to effectively model crash frequencies at multiple severity levels. This paper aims to make a contribution in this area by applying a multivariate count data model that is capable of accounting for correlated unobserved factors across equations representing crash frequencies at different severity levels. The correlation is accommodated by allowing for the presence of error covariances across equations, thus contributing to the simultaneity in the phenomenon under investigation.

The need for modeling crash frequencies by severity level in a simultaneous equations framework has been recognized; however, the analytical and computational complexity associated with formulating and estimating such systems has hindered the development of these model systems, particularly in the count model (data) context. This paper applies the modeling estimation technique proposed in Ye et al. (2009), where an n-dimensional multivariate count data model (Poisson regression) is formulated and presented to account for error correlations through the incorporation of normally distributed heterogeneity terms. Model estimation is achieved through the use of maximum simulated likelihood estimation (MSLE) methods that provide consistent parameter estimates and valid statistics for hypotheses tests. This paper makes a contribution to the understanding of crash frequencies at various severity levels for freeway sections. By applying a simultaneous equations model system for freeway crash severities, the paper provides key insights into the factors that impact crash frequencies at various severity levels while accounting for the presence of error covariances (common unobserved factors).

This paper has two major objectives. The first objective is to explore the use of the normal distribution to represent the heterogeneity in Poisson regression models of traffic crash frequency, as opposed to the log-gamma distribution which has been widely used in the Negative Binomial model. The paper aims to show that researchers can take advantage of the multivariate normal distribution to accommodate correlations of heterogeneities among multiple interrelated traffic crash frequencies at different severity levels in order to improve efficiency of coefficient estimators. The second major objective of the paper is to offer insights into the effects of various roadway, geometric, and traffic volume factors on crash frequencies by severity for freeway sections, while explicitly accounting for correlations across unobserved attributes.

Following a brief review of the literature, the paper presents the modeling methodology adopted in this paper. This is followed by a description of the dataset. Model estimation results and key conclusions are presented in the final two sections of the paper.

2. Modeling crash frequency and severity

Previous research in crash severity analysis such as that undertaken by Shankar et al. (1996) and Shankar and Mannering (1996) has mostly involved the development of univariate models of severity, with specific focus on total severity of the crash. In such studies, the most severe outcome of the crash is modeled. While the most severe outcome approach is useful in terms of a methodology for identifying model specifications, it does not provide for a comprehensive analysis of the severity of a crash. For example, the specific injury levels of occupants, and the variation in property damage among vehicles in multiple vehicle crashes are not adequately modeled. When one starts to consider these aspects, the notion of multivariate severity modeling is certainly appealing, but makes the modeling task more complex. In addition to these issues, the consideration of multiple modes tends to complicate injury modeling. For example, when pedestrians are involved, it is very rare that property damage outcomes are observed. In this case, the severity distribution is near-truncated at the possible injury level on the lower bound, which then leads to additional modeling complexities when accounting for multiple modal characterizations in multivariate or univariate models.

There is undoubtedly a vast body of literature devoted to modeling crash severity outcomes as a function of crash type, driver characteristics, roadway and traffic characteristics, and environmental conditions. These papers have used a variety of discrete choice modeling approaches, most notably the ordered probit and multinomial logit modeling approaches, to model crash severity outcomes. Examples of ordered probit models of injury severity include Quddus et al. (2002), Kockelman and Kweon (2002),