Analytic Methods in Accident Research 19 (2018) 16–32

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

A joint econometric approach for modeling crash counts by collision type

Tanmoy Bhowmik, Shamsunnahar Yasmin *, Naveen Eluru

Department of Civil, Environmental & Construction Engineering, University of Central Florida, United States

ARTICLE INFO

Article history:
Received 15 January 2018
Received in revised form 13 June 2018
Accepted 13 June 2018

Keywords:
Joint system
Multinomial fractional split model
Crash types
Equivalent log-likelihood
Comparison exercise
Unobserved effect

ABSTRACT

In recent years, there is growing recognition that common unobserved factors that influence crash frequency by one attribute level are also likely to influence crash frequency by other attribute levels. The most common approach employed to address the potential unobserved heterogeneity in safety literature is the development of multivariate crash frequency models. The current study proposes an alternative joint econometric framework to accommodate for the presence of unobserved heterogeneity – referred to as joint negative binomial-multinomial logit fractional split (NB-MNLFS) model. Furthermore, the study undertakes a first of its kind comparison exercise between the most commonly used multivariate model (multivariate random parameter negative binomial model) and the proposed joint approach by generating an equivalent log-likelihood measure. The empirical analysis is based on the zonal level crash count data for different collision types from the state of Florida for the year 2015. The model results highlight the presence of common unobserved effects affecting the two components of the joint model as well as the presence of parameter heterogeneity. The equivalent log-likelihood and goodness of fit measures clearly highlight the comparable performance offered by the proposed joint model relative to the commonly used multivariate approach. Overall, the model interpretations and fit measures clearly highlight the potential complementary role of the proposed approach for crash frequency analyses.

1. Introduction

Road traffic crashes are responsible for nearly 1.25 million fatalities every year and are a leading cause of death among people aged between 15 and 29 years old (World Health Organization, 2015). The extent of societal, emotional and economic impacts of these unfortunate events has warranted coordinated multi-sectoral responses from the fields of transportation, public health, and medicine. A major analytical tool employed for examining the critical factors influencing crash occurrence include the econometric crash prediction/crash frequency models. These models examine crashes at the micro-level (such as an intersection or roadway segment) or at the macro-level (such as a county or Traffic Analysis Zone (TAZ)). The various crash frequency dimensions frequently explored in existing literature include total crashes, crashes by severity, crashes by collision type and crashes by vehicle type for a spatial unit over a given time period.

A majority of the existing studies in safety literature developed crash frequency models for a single dependent variable; the methods are referred to as univariate modeling approaches (see Lord and Mannering, 2010; Yasmin and Eluru, 2018 for a

* Corresponding author.
E-mail addresses: tanmoy78@knights.ucf.edu (T. Bhowmik), shamsunnahar.yasmin@ucf.edu (S. Yasmin), naveen.eluru@ucf.edu (N. Eluru).

https://doi.org/10.1016/j.amar.2018.06.001
2213-6657/

detailed review of these studies). In recent years, there is growing recognition that univariate approaches, while adequate for analyzing a single dependent variable, fall short in modeling multiple crash frequency variables for a single observational unit. For example, the total number of crashes in a TAZ are a sum of crashes by different collision types (or severity levels) *i.e.* as opposed to analyzing a single total crash variable it is possible to examine crash frequency by different attribute categories. In this case, an extension of univariate approach would be to develop multiple univariate models with frequency by attribute levels considered as multiple dependent variables. Through this approach, the exogenous variables affecting crash counts can exhibit distinct impacts on different attribute levels allowing for a flexible specification. The separate models for crash frequency by attribute level allows us to capture realistic estimates of exogenous variables. Yet, the approach only accommodates for observed factors and inherently neglects the information that the multiple crash frequency variables for a TAZ are potentially correlated. For example, for zonal level crash frequency analysis, it is possible that several characteristics specific to the zone such as driver behavior, geometric design and build quality (possibly of higher or lower quality relative to the other zones) and traffic signal design objectives might influence different crash counts by collision type (such as head-on, rear-end). These factors that influence crash frequency by one attribute level are also likely to influence crash frequency by other attribute levels. Such detailed characteristics are rarely available to analysts for consideration in model development. Ignoring for the presence of such unobserved heterogeneity in model development will result in inaccurate and biased model estimates (see Mannering et al., 2016 for an extensive discussion).

The most common approach employed to address the potential unobserved heterogeneity in safety literature is the development of multivariate crash frequency models. In this approach, the impact of exogenous variables is quantified through the propensity component of count models. The main interaction across different count variables is sought through unobserved effects *i.e.* there is no interaction of observed effects across the multiple count models. These approaches, in general, partition the error components of the dependent variables to accommodate for a common term and an independent term across dependent variables (see Mannering et al., 2016 for a detailed discussion of various methodologies). In our current study, we develop an alternative approach to accommodate for the presence of observed and unobserved heterogeneity. The approach employs a joint crash frequency and multinomial fractional split model to provide an alternative to the multivariate count models in extant literature. The approach builds on recent work by Yasmin and colleagues in multiple studies (Lee et al., 2018; Yasmin et al., 2016; Yasmin and Eluru, 2018). Furthermore, the current study undertakes a first of its kind comparison exercise between the most commonly used multivariate count model (multivariate random parameter negative binomial model) and the proposed approach of the current study. The reader would note that the log-likelihood functions across these models are not directly comparable. Hence, to facilitate a comparison, an equivalent log-likelihood measure is also generated for the proposed joint crash frequency and fractional split model. Finally, an in-sample prediction exercise comparing the two systems is conducted. The models are estimated by using data from Florida at the Statewide Traffic Analysis Zone (STAZ) level for the year 2015.

The rest of the paper is organized as follows: Section 2 provides a brief review of previous relevant research. Section 3 and 4 provide a description of the modeling approach and data, respectively. Model estimation results and comparison results are presented in Section 5. Finally, a summary of model findings and conclusions are presented in Section 6.

2. Background and current study in context

2.1. Earlier research

Earlier research efforts in safety literature have focused on univariate model systems for crash frequency analysis. Majority of these studies focus on crash frequency by vehicle involvement (Ivan et al., 2000; Persaud and Mucsi, 1995; Qin et al., 2004; Zhou and Sisiopiku, 1997) or crash type (Chai and Wong, 2014; Li et al., 2016; Wang and Abdel-Aty, 2008a, 2008b, 2006; Yan et al., 2005). It is beyond the scope of our paper to review the vast literature of univariate models (please see Lord and Mannering, 2010; Yasmin and Eluru, 2018 for a literature review).

Recently, research in safety literature has shifted toward modeling multiple dependent variables for each observation unit. The most common approach for modeling multiple dependent variables such as crash frequency by severity or collision type is based on using a multivariate crash frequency model. In these models, every crash frequency variable is associated with its corresponding propensity equation (similar to univariate system). Thus, we allow for the impact of exogenous variables to vary across crash frequency variables. For example, consider the exogenous variable - presence of left guardrail on the roadway. In the presence of a left guardrail, vehicles are prevented from entering the opposite direction thus reducing head-on crashes. On the other hand, vehicles on hitting the guardrail might collide with other vehicles travelling in the same direction. Thus, the overall impact of the guardrail might be an increase in total crashes with distinct effects on head-on and sideswipe crashes. So, considering the guardrail variable in the total crash would yield a positive sign. However, considering the same variable in separate univariate models for head-on collisions and sideswipe collisions offer different results. This is an example of how observed variables exhibit contrasting effects on crash occurrence by collision[1] type. Thus, developing separate models for frequency by collision type allows us to capture realistic estimates of exogenous variables.

[1] We use crash and collision synonymously in the current study context.