Analytic Methods in Accident Research 11 (2016) 17–32

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

Random parameters multivariate tobit and zero-inflated count data models: Addressing unobserved and zero-state heterogeneity in accident injury-severity rate and frequency analysis

Panagiotis Ch. Anastasopoulos

Department of Civil, Structural and Environmental Engineering, Institute for Sustainable Transportation and Logistics, Engineering Statistics and Econometrics Application Research Lab, University at Buffalo, The State University of New York, 241 Ketter Hall, Buffalo, NY 14260, United States

ARTICLE INFO

Article history:
Received 7 May 2015
Received in revised form
2 June 2016
Accepted 2 June 2016
Available online 20 June 2016

Keywords:
Random parameters
Unobserved heterogeneity
Zero-state heterogeneity
Multivariate tobit
Multivariate zero-inflated negative binomial (ZINB)
Accident injury-severity rate
Accident injury-severity frequency

ABSTRACT

This paper uses data collected over a five-year period between 2005 and 2009 in Indiana to estimate random parameters multivariate tobit and zero-inflated count data models of accident injury-severity rates and frequencies, respectively. The proposed modeling approach accounts for unobserved factors that may vary systematically across segments with and without observed or reported accident injury-severities, thus addressing unobserved, zero-accident state and non-zero-accident state heterogeneity. Moreover, the multivariate setting allows accounting for contemporaneous cross-equation error correlation for modeling accident injury-severity rates and frequencies as systems of seemingly unrelated equations. The tobit and zero-inflated count data modeling approaches address the excessive amount of zeros inherent in the two sets of dependent variables (accident injury-severity rates and frequencies, respectively), which are – in nature – continuous and discrete count data, respectively, that are left-censored with a clustering at zero. The random parameters multivariate tobit and zero-inflated count data models are counter-imposed with their equivalent fixed parameters and lower order models, and the results illustrate the statistical superiority of the presented models. Finally, the relative benefits of random parameters modeling are explored by demonstrating the forecasting accuracy of the random parameters multivariate models with the software-generated mean βs of the random parameters, and with the observation-specific βs of the random parameters.

Published by Elsevier Ltd.

1. Introduction

In modern accident analysis, motor vehicle accidents – frequently viewed within an injury-severity level prism – are observed on roadway segments over a finite period of time. The duration of the observation period typically introduces significant statistical modeling challenges. For example, accidents (or accidents of specific injury-severity level) may not be observed during the study period on specific roadway segments, which in turn may result in a significant number of roadway segments with accident frequencies or rates (or accident injury-severity frequencies or rates) that are left-censored at zero observations. This preponderance of zero-accident observations can be due to either the fact that no accidents were observed or reported over the finite

E-mail address: panastas@buffalo.edu

http://dx.doi.org/10.1016/j.amar.2016.06.001
2213-6657/Published by Elsevier Ltd.

duration of the study period, or simply because the specific zero-accident roadway segments were of a very low accident risk – in terms of no accident occurrence – during the study period. The latter does not imply the certainty of perfectly safe roadway segments, but the possibility that over a period of time accidents may not occur on some roadway segments. In fact, numerous studies (Abdel-Aty and Radwan, 2000; Anastasopoulos et al., 2008; Geedipally and Lord, 2010; Anwaar et al., 2012; Mohammadi et al., 2014; Zou et al., 2014; Barua et al., 2015; for a detailed list, see Lord and Mannering (2010)) have presented a great amount of roadway segments of varying lengths (from a few hundred feet to several miles) with zero accidents over their corresponding finite study periods, ranging from a few months to multiple years. Even though these zero-accident roadway segments are not considered as very low accident risk segments, they form the basis to compare segments with (or with a large amount of) accidents and identify safety countermeasures. This is an indication that very low accident risk segments – determined as zero-accident state segments – are likely to exist (even over a finite period of time), which makes the identification of their characteristics a critical task.

At the same time, the effect of the factors affecting the zero-accident state (as well as the non-zero-accident state) segments are likely to vary across the observations, due to unobserved factors that may vary systematically across the zero-accident state (or the non-zero-accident state, as well) segments. This can result in unobserved heterogeneity, an important misspecification issue that typically results in biased parameter estimates, reduced model fit, and forecasting inaccuracy (for example see, Benhood and Mannering (2015), Coruh et al. (2015), Barua et al. (2016), Mannering et al. (2016), and Mothafer et al. (2016)). Its origins lie – within the context of accident injury-severity frequency and rate analysis – in information that is possibly available only after an accident has occurred (such as, vehicle-specific information, socio-economic and behavioral driver characteristics, residual environmental effects, and so on), and thus cannot be used to predict accident occurrences.

From a methodological viewpoint, past research has used a variety of statistical modeling methods to study accident and accident injury-severity frequencies and rates, and account for the possibility of zero-accident state segments. Mannering and Bhat (2014) provide a detailed list of such approaches, with the zero-inflated count data models being the most popular in accident frequency analysis (discrete count data), and the tobit models in accident rate analysis (continuous data). The two approaches are similar, in the sense that they consider the dependent variables to be left-censored with a clustering at zero, thus accounting for an underlying zero-accident state. In their traditional framework, the overall accident frequencies and rates, respectively, are modeled irrespective of their injury-severity level, with the assumption that the equivalent effect of the censoring sources is the same for all injury-severity levels. Evidently, this can be unrealistic, if low injury-severity accidents are less likely to be observed or reported. To that end, Anastasopoulos et al. (2012b) and Dong et al. (2014) proposed modeling accident frequencies and rates, respectively, by injury-severity level in a multivariate scheme (e.g., as a system of seemingly unrelated equations), which accounts for the zero-accident state, the censoring masking, and the contemporaneous error correlation of the latent variables due to commonly shared unobserved characteristics. Dong et al. (2014) also addressed the issue of unobserved heterogeneity across roadway segments with accidents (non-zero-accident state segments).

The intent of this paper is to extend the methodological work in accident injury-severity frequency and rate analysis, by explicitly accounting for unobserved heterogeneity in both the zero-accident and non-zero-accident injury-severity states, through the use of random parameters modeling. To that end, accident injury-severity rates are modeled with a random parameters multivariate tobit model, and accident injury-severity frequencies with a random parameters zero-inflated negative binomial model (with random parameters on both the zero-accident and the non-zero-accident injury-severity states). The results are counter-imposed against their corresponding fixed parameter and lower order models.

2. Method and approach

With the injury-severity levels typically defined by the degree of injury of the most severely injured occupant in the accident, the accident injury-severity rate and frequency data can be left censored with a clustering at zero. This is possible when accidents of a specific injury-severity level may not be observed throughout the study period. The zero-accident state may in turn be due to either data constraints including under-reporting (e.g., low severity accidents that do not involve injury, typically are only reported if the value of the property damage exceeds a pre-specified threshold value), or simply because there may be a class of roadway segments where accidents of a specific injury-severity level rarely occur. In both accident frequency and rate analyses, the equivalent effect is assumed to be the same, left censored with a clustering at zero observations.

2.1. Multivariate tobit model

For the statistical analysis and modeling of accident injury-severity rates, multivariate tobit regression (for the traditional tobit model, see Tobin, 1958) has been found to be an appropriate approach (Anastasopoulos et al., 2008, 2012a, 2012b; Chen et al., 2014; Xu et al., 2014). The multivariate tobit model (Huang et al., 1987; Huang, 1999; Trivedi and Zimmer, 2005) accounts for contemporaneous (cross-equation) error correlation among the accident injury-severity rates (i.e., property damage only crashes per 100-million vehicle miles traveled, injury crashes per 100-million vehicle miles traveled, and fatal crashes per 100-million vehicle miles traveled) on specific roadway segments. The multivariate tobit model with three left-censored at zero dependent variables can be expressed as (with y_{ij}^* being a latent variable observed only when positive):

Download English Version:

https://daneshyari.com/en/article/1104542

Download Persian Version:

https://daneshyari.com/article/1104542

Daneshyari.com