Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

Analysis of accident injury-severity outcomes: The zero-inflated hierarchical ordered probit model with correlated disturbances

Grigorios Fountas[a,*], Panagiotis Ch. Anastasopoulos[b]

[a] Transport Research Institute, School of Engineering and the Built Environment, Edinburgh Napier University, 10 Colinton Road, Edinburgh EH10 5DT, UK

[b] Department of Civil, Structural and Environmental Engineering, Stephen Still Institute for Sustainable Transportation and Logistics, University at Buffalo, The State University of New York, 241 Ketter Hall, Buffalo, NY 14260, United States

HIGHLIGHTS

- Two underlying states of accident injury-severity outcomes are considered.
- A zero-inflated hierarchical ordered probit model is estimated.
- The proposed approach consists of a binary probit and an ordered probit component.
- Correlation of the disturbance terms among the model components is unrestricted.
- The proposed approach provides statistical and forecasting accuracy improvements.

ARTICLE INFO

Keywords:
Accident injury-severities
Zero-inflated hierarchical ordered probit
Correlated disturbances
Injury-severity states
Threshold decomposition

ABSTRACT

In accident injury-severity analysis, an inherent limitation of the traditional ordered probit approach arises from the *a priori* consideration of a homogeneous source for the accidents that result in a no-injury (or zero-injury) outcome. Conceptually, no-injury accidents may be subject to the effect of two underlying injury-severity states, which are more likely to be observed in accident datasets with excessive amounts of no-injury accident observations. To account for this possibility along with the possibility of heterogeneity stemming from the fixed nature of the ordered probability thresholds, a zero-inflated hierarchical ordered probit approach with correlated disturbances is employed, for the first time – to the authors' knowledge – in accident research. The latter consists of a binary probit and an ordered probit component that are simultaneously modeled in order to identify the influential factors for each underlying injury-severity state. At the same time, the model formulation accounts for possible correlation between the disturbance terms of the two model components, and allows for the ordered thresholds to vary as a function of threshold-specific explanatory variables. Using injury-severity data from single-vehicle accidents that occurred in the State of Washington, from 2011 to 2013, the implementation potential of the proposed approach is demonstrated. The comparative assessment between the zero-inflated hierarchical ordered probit approach with correlated disturbances and its lower-order counterparts highlights the potential of the proposed approach to account for the effect of underlying states on injury-severity outcome probabilities and to explain more with the same amount of information.

* Corresponding author.
E-mail addresses: G.Fountas@napier.ac.uk (G. Fountas), panastas@buffalo.edu (P.C. Anastasopoulos).

https://doi.org/10.1016/j.amar.2018.09.002
Received 3 April 2018; Received in revised form 7 September 2018; Accepted 8 September 2018
2213-6657/

1. Introduction

In contemporary accident research, the investigation of the determinants of accident injury-severity outcomes is largely based on high-scale datasets consisting of police-reported accidents that occurred throughout a specific period of time. A structural characteristic of such datasets originates from the preponderance of accident observations that are associated with a no-injury outcome (Jiang et al., 2013). Even though the high percentage of no-injury accidents does not imply traffic safety improvements in the roadway network, such excessive amount of observations is typically anticipated, especially when information about urban settings or highly congested networks is collected. In this context, accounting for such a common pattern in the injury-severity data constitutes an imminent statistical challenge with possible effect on parameters' efficiency and inferences' accuracy.

Recent advances on statistical and econometric approaches have addressed several issues arising from the restrictive formulations of the traditional ordered and discrete outcome frameworks, which are typically used in the analysis of accident injury-severity data (for a detailed review, see: Lord and Mannering, 2010; Savolainen et al., 2011; Mannering and Bhat, 2014, Mannering et al., 2016; Mannering, 2018). Such issues include various patterns of unobserved heterogeneity (Russo et al., 2014; Yasmin et al., 2014; Behnood et al., 2014; Eluru and Yasmin, 2015; Bogue et al., 2017; Seraneeprakarn et al., 2017; Behnood and Mannering, 2016; Fountas and Anastasopoulos, 2017; Osman et al., 2018; Fountas et al., 2018a), endogeneity (Rana et al., 2010; Abay et al., 2013; Sarwar and Anastasopoulos, 2017), temporal and spatial correlation (Castro et al., 2012; Chiou and Fu, 2015; Behnood and Mannering, 2015; Bhat et al., 2017; Zeng et al., 2018; Osama and Sayed, 2017; Paleti et al., 2017; Mannering, 2018), and under-reporting of accidents (Savolainen et al., 2011; Mannering and Bhat, 2014). Despite the development of more flexible estimation structures (such as random parameters and latent class techniques, multivariate formulations, and so on), the theoretical causality for the preponderance of no-injury accidents and its possible implications on model estimation are not effectively accounted for in such structures.

The latter is important, especially when consideration is given to the sources of no-injury observations (or zero-injury observations, according to the typical numbering of the injury-severity outcomes in an ordinal scale). Specifically, zero-injury observations may not arise from a uniform source, because the underlying accident generation mechanisms may vary. For example, a portion of the zero-injury observations may be associated with very minor accidents, whose accident-specific conditions and contributing factors are unlikely to lead to a more severe, injury-involved outcome. Acknowledging that in the majority of injury-severity analyses the property-damage-only and possible injury outcomes are aggregated within the no-injury outcome (Savolainen et al., 2011), the aforementioned group of accidents is more likely to result in property damage only. Some indicative accident types relating to the aforementioned group of zero-injury observations include (but are not limited to): parking-related accidents, low-speed accidents, accidents at bottlenecks, or accidents involving low impact collision. Similarly, the remaining group of zero-injury observations may be associated with accidents that under different traffic-, weather-, roadway-, driver-, or vehicle-specific circumstances could naturally result in a more severe, injury-involved outcome. In the context of single-vehicle accidents, possible accident types associated with the latter group of zero-injury observations involve run-off-road accidents, collision with roadway structures, animal-involved accidents, and so on. Despite their observed no-injury outcome, the underlying mechanism corresponding to this group of zero-injury observations may share considerable (observed or unobserved) similarities with the mechanism of the injury-involved accidents.

From a theoretical perspective, the possible presence of alternate injury-severity mechanisms can lead to the consideration of two distinct injury-severity states in the analysis of accident data: (i) the zero-injury state; and (ii) the ordered injury-severity state. In line with the previous distinction of the zero-injury observations, the group of zero-injury observations that do not have the potential to result in more severe outcomes, form the basis of the zero-injury state. On the contrary, the second group of no-injury accidents (i.e., those that can potentially result in more severe outcomes) as well as all the injury-involved accidents, form the basis of the ordered injury-severity state. Note that the consideration of the ordered injury-severity state has a two-fold function: (i) to illustrate the generation mechanism of the non-zero-injury state; and (ii) to account for the inherent generation differences among the injury-severity outcomes.

From a modeling perspective, several statistical and econometric approaches have been developed to accommodate the possibility of excessive amount of zero observations in accident datasets. Such approaches include the zero-inflated count data models, such as the zero-inflated Poisson (Shankar et al., 1997; Boucher et al., 2009; Aguero-Valverde, 2013; Dong et al., 2014a,) and the zero-inflated negative binomial models (Jang et al., 2010; Usman et al., 2010; Dong et al., 2014b; Anastasopoulos, 2016; Cai et al., 2016; Liu et al., 2018) as well as the (univariate or multivariate) tobit models (Anastasopoulos et al., 2012a, 2012b; Anastasopoulos, 2016; Zeng et al., 2017). In accident research, these two streams of statistical techniques primarily account for the presence of zero-accident and non-zero-accident states in roadway segment based accident frequency and rate analysis, respectively (Anastasopoulos, 2016). However, the possibility of two distinct states at a more disaggregate level, and particularly at the level of accident observations, has been left under-explored, especially within the context of accident injury-severity research. Accounting for the presence of underlying states in the level of accident observations can result not only in more reliable parameter estimates, but also in the efficient use of after-crash information (i.e., accident-, vehicle-, driver-, time-varying weather-, and pavement-specific information) that cannot be used for the identification of underlying accident states at the roadway segment level.

In the context of injury-severity studies, Jiang et al. (2013) identified two possible underlying classes for the zero-injury accidents: (i) the injury-free accidents; and (ii) the injury-prone accidents. To statistically account for this possibility within an ordered probability setting, Jiang et al. (2013) employed a zero-inflated ordered probit approach; the latter can address the excessive amount of zero observations by identifying distinct regimes on the basis of binary probit and traditional ordered probit processes.

This study aims to extend the methodological potential of the zero-inflated ordered probit model, by additionally accounting for