Analytic Methods in Accident Research 18 (2018) 57–68

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

Analysis of accident injury-severities using a correlated random parameters ordered probit approach with time variant covariates

Grigorios Fountas [a,*], Panagiotis Ch. Anastasopoulos [b], Mohamed Abdel-Aty [c]

[a] *Department of Civil, Structural, and Environmental Engineering, Engineering Statistics and Econometrics Application Research Laboratory, University at Buffalo, The State University of New York, 204B Ketter Hall, Buffalo, NY 14260, United States*

[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*

[c] *Department of Civil, Environmental & Construction Engineering, University of Central Florida, Engr II 211-W, Orlando, FL 32816, United States*

ARTICLE INFO

Article history:
Received 24 October 2017
Received in revised form 11 April 2018
Accepted 11 April 2018

Keywords:
Accident injury-severity
Time-variant parameters
Correlated random parameters
Ordered probit

ABSTRACT

This paper employs a correlated random parameters ordered probit modeling framework to explore time-variant and time-invariant factors affecting injury-severity outcomes in single-vehicle accidents. The proposed approach extends traditional random parameters modeling, by accounting for possible correlations among the random parameters. On the basis of an unrestricted covariance matrix for the random parameters, the proposed framework can capture the combined effect of the unobserved factors – which are captured by the random parameters – on the injury-severity mechanism. The empirical analysis is based on traditional roadway-, traffic- and crash-specific information, and detailed weather and pavement surface disaggregate data, collected in the State of Washington, between 2011 and 2013. The results show that accident injury-severity outcomes are affected by a number of time-variant (ice thickness or water depth on pavement surface, sub-surface temperature) and time-invariant (roadway geometrics, and vehicle-, driver-, and collision-specific characteristics) factors, several of which result in statistically significant parameters – thus they have mixed effects on the injury-severity generation mechanism. The findings also present statistically significant correlation effects among the random parameters, which substantiates the appropriateness of the approach. The comparative assessment between the employed approach and its lower-order counterparts (i.e., fixed parameters, and uncorrelated random parameters ordered probit modeling approaches) shows that accounting for the unobserved heterogeneity interactions results not only in superior statistical performance (in terms of model's fit, and explanatory and prediction performance) but also in less biased and more consistent parameter estimates.

Published by Elsevier Ltd.

1. Introduction

In contemporary transportation research, the emergence of real-time data collection techniques and the subsequent availability of highly disaggregate data have considerably enhanced the methodological potential of the traditional statistical

* Corresponding author.
E-mail addresses: gfountas@buffalo.edu (G. Fountas), panastas@buffalo.edu (P.Ch. Anastasopoulos), M.Aty@ucf.edu (M. Abdel-Aty).

https://doi.org/10.1016/j.amar.2018.04.003
2213-6657/Published by Elsevier Ltd.

and econometric frameworks. In accident analysis, in particular, the consideration of real-time (or near real-time) information allows for a thorough investigation of the effect of time-variant factors, which cannot be sufficiently captured through the use of conventional accident datasets. In addition, the use of highly dimensional data highlights the need for robust methodological extensions that can accommodate the large-scale nature of the former.

In accident injury-severity analysis, past research has identified two broad categories of time-variant factors as significant determinants of the injury-severity outcomes: traffic characteristics and weather conditions. In fact, over the last years, a considerable portion of accident research has explored the separate or combined effect of these characteristics on injury-severity outcomes – Theofilatos and Yannis (2014) provide a detailed review of such studies. Interestingly, the main stream of these studies have used primarily aggregate data to investigate the effect of weather characteristics (Ma and Kockelman, 2004; Hill and Boyle, 2006; Caliendo et al., 2007; Kopelias et al., 2007; Milton et al., 2008; Abdel-Aty et al., 2011; Jung et al., 2011; Peng and Boyle, 2012; El-Basyouny and Kwon, 2012; El-Basyouny et al., 2014; Mohamed et al., 2013; Anderson an Hernandez, 2017); whereas, recent studies have incorporated (near) real-time weather data, to account for the effect of dynamic (time-variant) variations on the injury-severity mechanism (Jung et al., 2010; Xu et al., 2013; Yu and Abdel-Aty, 2014a,b).

Even though the studies using disaggregate data can shed more light on the time-varying impact of weather characteristics, the effect of unobserved characteristics – which are likely to vary systematically across observations – on the resulting injury-severity outcomes remains an important modeling limitation. Such unobserved characteristics may originate from non-observable driving behavior traits, accident-specific attributes, time-invariant roadway- or driver-specific characteristics, and time-variant weather elements, which cannot all be captured with real-time collection techniques. To account for the effect of such unobserved heterogeneity, random parameters modeling techniques have been employed in the recent safety research (Anastasopoulos and Mannering, 2009; Kim et al., 2010; Anastasopoulos and Mannering, 2011; Russo et al., 2014; Anastasopoulos and Mannering, 2016; Sarwar et al., 2017a,c).

Under the traditional random parameters modeling approach, the sources of heterogeneity are assumed to be independent (Mannering et al., 2016). In fact, there is strong possibility for the sources of heterogeneity to be correlated, due to possible interactions among the unobserved characteristics. Even though allowing to capture such unobserved characteristics, the traditional random parameters modeling approach cannot provide parameter estimates that account for the possible correlation effects among the distributions of the random parameters, and, in turn, among the various sets of unobserved factors (Mannering et al., 2016). Such correlation effects may be more evident when consideration is given to the determinants of the injury-severity outcomes, due to the presence of numerous sources of heterogeneity that are anticipated to have interactive effects on the accident injury-severity mechanism.

This paper simultaneously addresses unobserved heterogeneity and unobserved heterogeneity interactions, by employing a correlated random parameters modeling framework. Specifically, possible correlation effects among the latter are explicitly captured, by applying the unrestricted covariance matrix of the random parameters (Yu et al., 2015; Fountas et al., 2018a). In this context, a correlated random parameters ordered probit model is estimated, to study the factors affecting injury-severity outcomes. The employed approach has the potential to account for the effect of time-variant characteristics, by using highly disaggregate time-variant information coupled with the traditional roadway- and accident-specific information. To evaluate the statistical, explanatory, and forecasting potential of the employed approach, the fixed and uncorrelated random parameters ordered probit counterparts are estimated and compared against the suggested approach.

2. Methodology

To account for the ordinal discrete nature of accident injury-severity data, a considerable portion of previous research has utilized an ordered probability framework to model injury-severity outcomes (Abdel-Aty, 2003; Haleem and Abdel-Aty, 2010; Rana et al., 2010; Yasmin and Eluru, 2013; Russo et al., 2014; Yasmin et al., 2015a,b; Eluru and Yasmin, 2015; Bogue et al., 2017). In the context of an ordered probit formulation, the traditional ordered probability model is defined on the basis of a latent continuous variable, z_i, as (Washington et al., 2011):

$$z_i = \boldsymbol{\beta}\mathbf{X}_i + \varepsilon_i,\; y_i = j,\; if\, \mu_{j-1} < y_i < \mu_j, \quad j = 1, 2, \ldots J \tag{1}$$

where $\boldsymbol{\beta}$ represents a vector of estimable parameters, $\mathbf{X}_i$ are vectors of observable characteristics, y is an integer that reflects the observed injury-severity outcome, j are integers corresponding to the various injury-severity levels, μ denote the threshold parameters that determine y and are also ordered in nature, and ε_i is a random error component assumed to be normally distributed.

To account for the effect of unobserved factors that can potentially vary systematically across accident observations, the random parameters modeling approach is employed (Milton et al., 2008; Anastasopoulos and Mannering, 2016; Anastasopoulos, 2016; Sarwar et al., 2017a; Nahidi et al., 2017; Seraneeprakarn et al., 2017; Bhat et al., 2017; Alarifi et al., 2017; Chen et al., 2017). The latter allows the estimation of accident-specific parameter vectors ($\boldsymbol{\beta}$s) for (some or all of) the explanatory variables, in an effort to capture underlying variations in the effect of observable characteristics. To that end, each accident-specific parameter vector ($\boldsymbol{\beta}_i$) is specified as:

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \Gamma\omega_i \tag{2}$$