Analytic Methods in Accident Research 15 (2017) 1–16

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

A random thresholds random parameters hierarchical ordered probit analysis of highway accident injury-severities

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

[a] Department of Civil, Structural, and Environmental Engineering, Engineering Statistics and Econometrics Application Research Laboratory, University at Buffalo, The State University of New York, 204 Ketter Hall, Buffalo, NY 14260, United States
[b] 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 13 January 2017
Received in revised form 24 March 2017
Accepted 25 March 2017
Available online 26 April 2017

Keywords:
Accident injury-severities
Hierarchical ordered probit
Random thresholds
Random parameters

ABSTRACT

This study uses highway accident data collected in the State of Washington, between 2011 and 2013, to study the factors that affect accident injury-severities. To account for the fixed thresholds limitation of the traditional ordered probability models – which typically leads to incorrect estimation of outcome probabilities for the intermediate categories – and for the possibility of unobserved factors systematically varying across the observations, a random thresholds hierarchical ordered probit model with random parameters is estimated. This approach simultaneously allows the explanatory parameters to vary across roadway segments, and the thresholds to vary both as a function of explanatory parameters and across the observations, thus accounting for unobserved and threshold heterogeneity, respectively. Using goodness-of-fit measures, likelihood ratio tests and forecasting accuracy measures, the model estimation results are compared with the hierarchical and fixed thresholds ordered probit model counterparts, with fixed and random parameters. The comparative assessment among the ordered probit modeling approaches reveals the relative benefits and the overall statistical superiority of the random thresholds random parameters hierarchical ordered probit model.

Published by Elsevier Ltd.

1. Introduction

Past research has extensively looked at factors affecting accident injury-severities, through the use of a variety of methodological approaches. The latter range from injury-severity frequency and rate studies (Anastasopoulos et al., 2012b; Venkataraman et al., 2013; Wu et al., 2014; Anastasopoulos, 2016; Zeng et al., 2016; Sarwar and Anastasopoulos, 2017), to more traditional accident injury-severity likelihood analyses, which include multinomial or binary logit/probit models (Shankar and Mannering, 1996; Khattak et al., 1998; Ulfarsson and Mannering, 2004; Khorashadi et al., 2005), nested models (Chang and Mannering; 1998; Chang and Mannering, 1999; Lee and Mannering, 2002), mixed logit models (Milton et al., 2008; Anastasopoulos and Mannering, 2011), fixed and random parameters ordered probability models (McCarthy and Madanat, 1994; Khattak, 2001; Abdel-Aty, 2003; Ye and Lord, 2014; Russo et al., 2014; Cerwick et al., 2014; Chen et al., 2016), and latent class ordered probability and multinomial logit models (Yasmin et al., 2014a; Shaheed and Gkritza, 2014; Cerwick et al., 2014; Behnood et al., 2014; Behnood and Mannering, 2016).

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

http://dx.doi.org/10.1016/j.amar.2017.03.002
2213-6657/Published by Elsevier Ltd.

Selection between the discrete outcome models and the ordered probability models is a tedious task, with both approaches sharing benefits and limitations. Several previous studies (Eluru, 2013; Yasmin and Eluru, 2013; Yasmin et al., 2014b) have investigated the statistical potential and appropriateness of these modeling frameworks for analyzing ordinal data. For example, the discrete outcome models allow the estimation of distinguished sets of independent variables for each injury-severity outcome (with the flexibility of possible variable-overlaps among the outcomes). However, these models do not account for the ordinal nature of accident injury-severity data, as they assume that the outcomes are independent of each other.

On the other hand, although ordered probability models account for the ordinal nature of the injury-severity data, they assume that the same set of independent variables affect all injury-severity outcomes. More importantly, the interpretation of the intermediate categories in these models is heavily affected by the thresholds (which are estimable parameters, and are restricted to be fixed across the observations) and the unambiguous effect of the independent variables on the highest and lowest ordered discrete category probabilities, with the effect on the interior category probabilities left unclear (Savolainen et al., 2011; Washington et al., 2011).

The generalized ordered logit model (Williams, 2006; Wang and Abdel-Aty, 2008; Eluru et al., 2008; Quddus et al., 2010) and its extension, the mixed generalized ordered logit model (Eluru et al., 2008; Eluru, 2013; Yasmin and Eluru, 2013; Yasmin et al., 2014b, 2015; Eluru and Yasmin, 2015), can relax the first restriction – i.e., the assumption that the same set of independent variables affect all injury-severity outcomes; however, the generalized ordered logit model can occasionally predict negative probabilities (Greene and Hensher, 2010a).[1] The hierarchical ordered probit (HOPIT) model accounts for the negative probability limitation; the thresholds are always positive and ordered, and are a function of a set of unique explanatory parameters that do not necessarily affect directly the ordered probability outcomes, hence partially addressing the second limitation – i.e., the threshold heterogeneity.[2]

This paper seeks to address – to a larger extent – the aforementioned threshold limitation, by allowing the thresholds to be a function of unique explanatory parameters and to vary across the observations, while simultaneously allowing the effect of the explanatory parameters that determine the ordered probability outcomes to also vary across the observations. To that end, a random thresholds random parameters hierarchical ordered probit model is estimated. This modeling approach seeks to address threshold heterogeneity and unobserved heterogeneity that, if left unaccounted for, can lead to inconsistent, biased, and inefficient predictors, and incorrect estimation of outcome probabilities for all – and in particular the intermediate – categories (Greene and Hensher, 2010b). To evaluate the statistical benefits of the proposed approach, the results of the random thresholds random parameters hierarchical ordered probit model are compared with the traditional hierarchical ordered probit and the fixed- and random-parameters ordered probit model counterparts.

2. Methodology

To study accident injury-severity probabilities in an ordered probability setting, the ordered probit model is defined as (Washington et al., 2011):

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

where, y is an integer corresponding to ordering of injury-severity outcomes, $\boldsymbol{\beta}$ are vectors of estimable parameters, $\mathbf{X}$ are vectors of explanatory variables, μ are threshold parameters that define y and are estimated with $\boldsymbol{\beta}$, j are the integer ordered injury-severity levels, and ε are random error terms that are assumed to be normally distributed with zero mean and variance equal to one.

Under the hierarchical ordered probit modeling scheme, the thresholds can vary as a function of a set of explanatory parameters as (Greene and Hensher, 2010b),

$$\mu_{i,j} = \mu_{i,j-1} + \exp(t_j + \mathbf{d}_j\mathbf{S}_i) \tag{2}$$

where, t is the intercept for each threshold, $\mathbf{S}$ are vectors of variables affecting the thresholds, and $\mathbf{d}$ are vectors of estimable parameters for $\mathbf{S}$. And to allow the thresholds to concurrently vary across the observations, Eq. (2) can be re-written as (Greene and Hensher, 2010b),

$$\mu_{i,j} = \mu_{i,j-1} + \exp(t_j + \gamma_j u_{ij} + \mathbf{d}_j\mathbf{S}_i) \tag{3}$$

where, u_{ij} is a normally distributed term with mean zero and standard deviation one, while t_j and γ_j are the mean and standard deviation of the threshold intercept term, respectively.

[1] The mixed generalized ordered response logit model allows the effect of the explanatory parameters to vary across the observations, and the thresholds to be determined as a function of the explanatory parameters (Eluru et al., 2008; Yasmin and Eluru, 2013; Yasmin et al., 2014b, 2015; Eluru and Yasmin, 2015). To address threshold heterogeneity, the model structure of the mixed generalized ordered logit model allows for the effect of the threshold-specific parameters to vary across observations (Eluru et al., 2008).

[2] According to Greene and Hensher (2010b) and Greene (2012), the hierarchical ordered model constitutes an alternate term for the generalized ordered model. This term does not imply the presence of any hierarchical data structure, but indicates the property of the model structure to allow for systematic decomposition of the threshold heterogeneity through the inclusion of exogenous variables in the corresponding threshold parametric forms (Green and Hensher, 2010a,b).