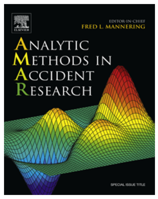
Analytic Methods in Accident Research 1 (2014) 23–38

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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



A latent segmentation based generalized ordered logit model to examine factors influencing driver injury severity

Shamsunnahar Yasmin [a,1], Naveen Eluru [a,*], Chandra R. Bhat [b,2], Richard Tay [c,3]

[a] *Department of Civil Engineering & Applied Mechanics, McGill University, Suite 483, 817 Sherbrooke St. W., Montréal, Québec, Canada H3A 2K6*
[b] *Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, 301 E. Dean Keeton St. Stop C1761, Austin, TX 78712, USA*
[c] *Faculty of Business, Economics and Law, La Trobe University, Melbourne, Victoria 3086, Australia*

ARTICLE INFO

Article history:
Received 5 August 2013
Received in revised form
20 October 2013
Accepted 20 October 2013
Available online 12 November 2013

Keywords:
Latent segmentation
Generalized ordered logit
Driver injury severity
Crash characteristics
Elasticities

ABSTRACT

This paper formulates and estimates an econometric model, referred to as the latent segmentation based generalized ordered logit (LSGOL) model, for examining driver injury severity. The proposed model probabilistically allocates drivers (involved in a crash) into different injury severity segments based on crash characteristics to recognize that the impacts of exogenous variables on driver injury severity level can vary across drivers based on both observed and unobserved crash characteristics. The proposed model is estimated using Victorian Crash Database from Australia for the years 2006 through 2010. The model estimation incorporates the influence of a comprehensive set of exogenous variables grouped into six broad categories: crash characteristics, driver characteristics, vehicle characteristics, roadway design attributes, environmental factors and situational factors. The results clearly highlight the need for segmentation based on crash characteristics. The crash characteristics that affect the allocation of drivers into segments include: collision object, trajectory of vehicle's motion and manner of collision. Further, the key factors resulting in severe driver injury severity are driver age 65 and above, driver ejection, not wearing seat belts and collision in a high speed zone. The factors reducing driver injury severity include the presence of pedestrian control, presence of roundabout, driving a panel van, unpaved road condition and the presence of passengers.

1. Introduction

Road traffic crashes continue to be a leading cause of death burdening the society with heavy economic losses (WHO, 2013). Most developed countries, through co-ordinated multi-sectoral responses to road safety issues, have been able to achieve a reduction in the crash related fatalities. For example, between 1975 and 2008, the annual road fatality rate of Australia declined from 8 deaths per 10,000 registered vehicles to 1 death per 10,000 registered vehicles (Ministry of Infrastructure and Transport, 2010; Australia Transport Council, 2011). In spite of these strides in improving road safety, traffic crashes still lead to substantial economic and emotional losses to the society. While improving road infrastructure

* Corresponding author. Tel.: +1 514 398 6823; fax: +1 514 398 7361.
E-mail addresses: shamsunnahar.yasmin@mail.mcgill.ca (S. Yasmin), naveen.eluru@mcgill.ca (N. Eluru), bhat@mail.utexas.edu (C.R. Bhat), r.tay@latrobe.edu.au (R. Tay).
[1] Tel.: +1 514 398 6823; fax: +1 514 398 7361.
[2] Tel.: +1 512 471 4535; fax: +1 512 475 8744.
[3] Tel.: +61 3 9479 1267; fax: +61 3 9479 3283.

2213-6657/$ - see front matter
http://dx.doi.org/10.1016/j.amar.2013.10.002

design to reduce traffic crash occurrence is essential, it is also important to provide solutions to reduce the consequences in the unfortunate event of a traffic crash. A critical component of identifying and gaining a comprehensive understanding of the factors that contribute to the negative consequences (property damage and injuries) of crash outcomes is the estimation and application of disaggregate level crash severity models.

In traffic crash reporting, injury severity is typically characterized as an ordered variable (such as no injury, minor injury, serious injury, and fatal injury). Thus, it is no surprise that the most commonly employed statistical formulation to model driver injury severity is the ordered response formulation. But the traditional ordered response formulation imposes a restrictive and monotonic impact of the exogenous variables on the injury severity alternatives. More recent research efforts using the ordered response formulation, following Eluru et al. (2008), have addressed the limitation of the traditional ordered response formulation by allowing for the exogenous variable impacts to vary across the alternatives in a generalized ordered logit (GOL) (or proportional odds logit) formulation (see Yasmin and Eluru (2013), Eluru (2013) and Mooradian et al. (2013)).

The current research effort contributes to the safety literature methodologically and empirically by building on the GOL formulation. In terms of methodology, we formulate and estimate a latent segmentation based generalized ordered logit (LSGOL) model. The LSGOL model relaxes the traditional GOL formulation assumption that the effects of exogenous variables on the injury risk propensity, and on the thresholds that map the risk propensity to injury severity outcomes, are fixed across all drivers involved in collisions. Empirically, the LSGOL model is estimated using driver injury severity data from the state of Victoria, Australia, employing a comprehensive set of exogenous variables.

The rest of the paper is organized as follows. A discussion of earlier research on crash injury severity is presented in Section 2, while also positioning the current study. Section 3 provides details of the econometric model framework used in the analysis. In Section 4, the data source and sample formation procedures are described. The model comparison results, elasticity effects and validation measures are presented in Section 5, 6 and 7, respectively. Section 8 concludes the paper and presents directions for future research.

2. Earlier research and current study in context

Road safety researchers have employed several statistical formulations for analyzing the relationship between injury severity and crash related factors. Savolainen et al. (2011) provide a detailed review of the different modeling formulations employed in crash injury severity analysis. But, as indicated earlier, the most prevalent formulation to study injury severity is the ordered response formulation (for example see Yasmin and Eluru (2013)). The traditional ordered response formulation imposes a restrictive monotonic assumption regarding the impact of exogenous variables on the injury severity levels (Eluru et al., 2008). To address this limitation, researchers have employed the unordered response formulation that allows the impact of exogenous variables to vary across injury severity levels. The most common model used under the unordered response formulation is the multinomial logit model (Khorashadi et al., 2005; Islam and Mannering, 2006; Awadzi et al., 2008; Schneider et al., 2009; Ulfarsson and Mannering, 2004). However, the unordered response model does not recognize the inherent ordering of the crash severity outcome and, therefore, neglects vital information present in the data. To recognize the ordinality of the injury severity levels, as well as provide as much flexibility as the unordered response formulation, Eluru et al. (2008) proposed the generalized ordered response formulation that bridges the divide between the traditional ordered-response and the traditional unordered-response formulations (Eluru, 2013; Yasmin and Eluru, 2013).

The widely employed discrete outcome formulations (ordered, generalized ordered, or unordered) typically restrict the impact of exogenous variables to be the same across the entire population of crashes (Eluru et al., 2012; Xie et al., 2012; Yasmin et al., forthcoming). *One approach* to extend these formulations to allow heterogeneity effects (variations in the effects of variables across the driver population) is to specify random coefficients (rather than impose fixed coefficients) (for example, see Eluru and Bhat, 2007; Paleti et al., 2010; Srinivasan, 2002; Morgan and Mannering, 2011; Kim et al., 2013). But, while the mean of the random coefficients can be allowed to vary across drivers based on observed crash-specific variables, the random coefficients approach usually restricts the variance and the distributional form of a random coefficient to be the same across all drivers. Thus, in a crash context, the impact of a rear-end crash (relative to an angular crash) may lead to a certain distribution of injury risk propensity due to unobserved factors. This distribution may be tight for low speed crashes (that is, the injury risk may be negative in the mean and tightly distributed about this mean), but more variant for high speed crashes (that is, the injury risk may be quite volatile in high-speed situations, with rear-end collisions leading to high injury severity in some cases and low injury severity in some other cases). This is a case of the distribution on the rear-end crash variable being dependent on another variable (low speed or high speed crashes). Such possibilities cannot be easily accommodated in random coefficients models. Besides, an *a priori* distribution form has to be imposed on the random coefficients, and the normal distribution assumption is usually imposed even though there is no reason why other distribution forms may not be more appropriate.

A *second approach* to allow heterogeneity effects is to consider segmenting the population based on exogenous variables (such as collision type, initial impact point of collision, speed, and location of impact) and estimate separate models for each segment (see Aziz et al., 2013 for segmentation based on location; Islam and Mannering, 2006 for segmentation based on driver demographics). However, because there may be many variables to consider in the segmentation scheme, the number of segments (formed by the combination of the potential segmentation variables) can explode rapidly. This causes problems