Accident Analysis and Prevention 72 (2014) 330–340

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Partial proportional odds model—An alternate choice for analyzing pedestrian crash injury severities

Lekshmi Sasidharan [a,*], Mónica Menéndez [b,1]

[a] Institute for Transport Planning and Systems, Swiss Federal Institute of Technology, ETH Zürich, Stefano-Franscini-Platz 5; HIL F41.1, 8093 Zürich, Switzerland
[b] Institute for Transport Planning and Systems, Swiss Federal Institute of Technology, ETH Zürich, Stefano-Franscini-Platz 5; HIL F37.2, 8093 Zürich, Switzerland

ARTICLE INFO

Article history:
Received 21 March 2014
Received in revised form 21 July 2014
Accepted 21 July 2014
Available online 9 August 2014

Keywords:
Crash severity
Partial proportional odds model
Ordered logit model
Multinomial logit model
Pedestrian
Comparison

ABSTRACT

The conventional methods for crash injury severity analyses include either treating the severity data as ordered (e.g. ordered logit/probit models) or non-ordered (e.g. multinomial models). The ordered models require the data to meet proportional odds assumption, according to which the predictors can only have the same effect on different levels of the dependent variable, which is often not the case with crash injury severities. On the other hand, non-ordered analyses completely ignore the inherent hierarchical nature of crash injury severities. Therefore, treating the crash severity data as either ordered or non-ordered results in violating some of the key principles. To address these concerns, this paper explores the application of a partial proportional odds (PPO) model to bridge the gap between ordered and non-ordered severity modeling frameworks. The PPO model allows the covariates that meet the proportional odds assumption to affect different crash severity levels with the same magnitude; whereas the covariates that do not meet the proportional odds assumption can have different effects on different severity levels. This study is based on a five-year (2008–2012) national pedestrian safety dataset for Switzerland. A comparison between the application of PPO models, ordered logit models, and multinomial logit models for pedestrian injury severity evaluation is also included here. The study shows that PPO models outperform the other models considered based on different evaluation criteria. Hence, it is a viable method for analyzing pedestrian crash injury severities.

1. Introduction

In-depth understanding of the frequency and severity of crashes is very important for developing effective safety management practices, including the selection of appropriate countermeasures, the development and enforcement of adequate policies, and so on. The crash frequency modeling generally focuses on the prediction of expected number of crashes and the identification of unsafe locations. The published literature on crash modeling shows that count models (Poisson and negative binomial regression models (Jones et al., 1991; Shankar et al., 1995; Hadi et al., 1995; Poch and Mannering, 1996; Milton and Mannering, 1998; Abdel-Aty and Radwan, 2000; Savolainen and Tarko, 2005; Lord, 2006; Lord and Park, 2008), zero-inflated models (Shankar et al., 1997; Carson and Mannering, 2001; Lee and Mannering, 2002), random effects and random parameters models (Shankar et al., 1998; Anastasopoulos and Mannering, 2009; Wu et al., 2013; Narayanamoorthy et al., 2013; Bhat et al., 2014), causal models (Karwa et al., 2011; Sasidharan and Donnell, 2013), and Conway–Maxwell–Poisson generalized linear models (Lord and Park, 2008; Francis et al., 2012; Lord and Guikema, 2012)), are commonly used to model crash frequencies. On the other hand, crash severity models focus on the prediction of probability of a crash being fatal, severe injury, minor injury or property damage only (PDO) given the occurrence of the crash. The ordered logit/probit models, generalized ordered response models, multinomial logit models, mixed logit models, and heteroskedastic models are commonly used to analyze pedestrian injury severities (O'Donnell and Connor, 1996; Ulfarsson and Mannering, 2004; Kim et al., 2008; Eluru et al., 2008; Kim et al., 2010; Mooradian et al., 2013; Bhat et al., 2014).

The frequency and severity models used in crash data analysis face several methodological issues: unobserved heterogeneity, crash underreporting, endogeneity, omitted variable bias, risk

* Corresponding author. Tel.: +41 0446333246.
E-mail addresses: lekshmi.sasidharan@ivt.baug.ethz.ch, lechu18181@yahoo.com (L. Sasidharan), monica.menendez@ivt.baug.ethz.ch (M. Menéndez).
1 Tel.: +41/0 44 633 6695.

http://dx.doi.org/10.1016/j.aap.2014.07.025
0001-4575/

compensation, with-in crash correlation, spatial correlation, and so on. Latent class/finite mixture models have recently started to be used in traffic safety studies to address the issue of unobserved heterogeneity (e.g., Park and Lord, 2009; Park et al., 2010; Bhat and Dubey, 2013; Zou et al., 2013). These models use a clustering technique to reduce the effect of bias in parameter estimates when there is a correlation between observed predictors and unobserved predictors. By identifying latent classes for which the parameter estimates are different, the heterogeneity between different groups of individual observations is taken into account (Everitt, 1988). Another noteworthy development in the analysis of discrete crash data is the random parameters (mixed) logit model, which can handle heterogeneous effects and correlated unobserved factors by having an additional error term that is not restricted to follow normal distributions (McFadden and Train, 2000). When the assumption of homoskedastic error variances in crash data is violated, the parameter estimates can be biased, inconsistent, and standard errors incorrect. This issue of heteroskedasticity can be addressed by heterogeneous choice models or location-scale models (Quddus et al., 2010; Keele and Park, 2006; Williams, 2006). Analysis considering injury severity of multiple individuals involved in a crash, needs to account for the unobserved factors that are correlated with each other and that influence the injury severity of different individuals involved in the crash. Eluru et al. (2010) suggested that a coupla-based multivariate approach with stochastic dependence among different severity levels can handle this problem. Another notable problem is the spatial and temporal correlations in the crash data. A crash occurring in one segment, for e.g., might have been caused by limitations in the nearby sections. Castro et al. (2013) reported that a spatial generalized ordered logit model can take into account the spatial dependencies in the injury severity levels of nearby crashes. For a detailed discussion on this refer to Savolainen et al. (2011) and Mannering and Bhat (2014).

Based on whether the explanatory variables are allowed to vary across the crashes, some researches have used fixed parameters models (e.g., Shankar and Mannering, 1996; Donnell and Mason, 2004; Savolainen and Ghosh, 2008; Ye and Lord, 2011; Sasidharan and Donnell, 2013; Yasmin and Eluru, 2013) and some others have used random parameters or mixed effects models (Milton et al., 2008; Kim et al., 2010; Ye and Lord, 2011; Wu et al., 2013; Yasmin and Eluru, 2013; Ye and Lord, 2014). The advantage of random parameters models is that they allow the explanatory variables to take into account the individual differences among injury severities in different crashes. The main focus of this study is the application of partial proportional odds (PPO) models for analyzing crash injury severities. The study also includes a comparison with two other fixed parameter models—ordered logit and multinomial logit (MNL) models.

This study focuses on crash severity modeling and therefore, four injury severity levels for pedestrian crashes are considered. Those levels are: fatal injury, severe injury, minor injury, and property damage only (PDO) crashes. Notice that the crash severity modeling focuses on the prediction of probability of a crash causing fatality, severe injury, minor injury, or PDO given the occurrence of the crash. As the response variable is choice-based, probabilistic models need to be used to analyze and predict crash severity. One of the commonly employed methods for crash severity modeling is ordered probability models (ordered logit or ordered probit models). Ordered probability models are a type of regression models that can be used when there are three or more categories for the dependent variables, and when the order (rank) of those categories is important. In crash severity modeling, there are different severity levels that are inherently related to one another, and there is a hierarchical order for these levels. This makes ordered probability models very accepted in crash severity modeling to capture the association between different severity levels. Previous safety studies that have used ordered models for severity analysis include Hutchinson (1986), O'Donnell and Connor (1996), Renski et al. (1998), Duncan et al. (1998), Abdel-Aty (2003), Zajac and Ivan (2003), Yamamoto and Shankar (2004), Eluru and Bhat (2007), Eluru (2013), Yasmin and Eluru (2013), Mooradian et al. (2013). Even though ordered models are extensively used in traffic safety research, there is a limitation associated with them. One of the key assumptions of the ordered probability models is that they must meet the proportional odds assumption i.e., the relationship between any two pairs in the dependent variable group are the same. This forces the coefficients for independent variables in the model to remain constant for all levels of the dependent variable (Wang and Abdel-Aty, 2008; Mooradian et al., 2013; Yasmin and Eluru, 2013). However, we know that some covariates might increase the probability of occurrence of some crash severity level(s); whereas they might reduce the probability of occurrence of some other severity level(s). Savolainen and Mannering (2007) and Peterson and Harrell (1990) suggest that ordered logit/probit models cannot account for this. Fu (1998) reported that the violation of proportional odds assumption in ordered logit models may lead to "incorrect, incomplete or misleading results". A recent approach to improve the ordered logit model by allowing moving the thresholds for different injury severity levels is the generalized ordered logit model (Eluru et al., 2008; Quddus et al., 2010; Yasmin et al., 2013a,b; Eluru, 2013; Yasmin and Eluru, 2013). Those studies suggest that these models can also account for the unobserved heterogeneity in the data.

Another conventionally used method to predict crash severity is the multinomial probability model (multinomial logit [MNL] models) (Shankar and Mannering, 1996; Chang and Mannering, 1999; Carson and Mannering, 2001; Lee and Mannering, 2002; Ulfarsson and Mannering, 2004; Khorashadi et al., 2005; Ye and Lord, 2011; Eluru, 2013; Yasmin and Eluru, 2013). MNL models are commonly used to estimate discrete outcomes when the data are not necessarily ordered (Washington et al., 2011). The MNL approach allows all independent variables to affect different levels of the dependent variable distinctively, which can account for non-monotonic effects on dependent variables (Tay et al., 2011). However, MNL models do not account for the ordered levels inherent in crash severity, and could suffer from correlation among outcomes, a problem known as Independence of Irrelevant Alternatives (IIA) which is a fundamental assumption in the derivation of the MNL (that is, the functions used to determine outcome probabilities must not share unobserved characteristics) and, if violated, can result in erroneous coefficient estimates. To resolve possible IIA problems (which may be present in some data) nested logit models have been applied (Shankar et al., 1996; Chang and Mannering, 1999; Hu and Donnell, 2010) to analyze injury severities. In this study, the IIA assumption was tested using Hausman test and Small-Hsiao test. Both results indicate that IIA assumption was not violated and hence, we are using MNL models, and not nested logit models in the present study.

As discussed above, ordered and multinomial probability models have their own limitations in the context of crash severity modeling. Nevertheless, they are still used in traffic safety research for modeling severities and comparing with new methods. Some of the recent studies that have used ordered logit and MNL models include Rifaat et al. (2011), Ye and Lord (2011), Schneider and Savolainen (2011), Eluru (2013), Yasmin and Eluru (2013), Ye and Lord (2014), Jung et al. (2010), Quddus et al. (2010). The ordered nature of the severity data cannot be ignored completely. At the same time, it is not appropriate to force all predictors to have the same effect on different levels of crash severity. A method that can be used under such circumstances is the PPO model proposed by Peterson and Harrell (1990). This model combines the ordered arrangement in ordinal models, and the ability of MNL to allow certain independent variables to affect different levels of dependent