Analytic Methods in Accident Research 3-4 (2014) 11–27

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

A comparison of the mixed logit and latent class methods for crash severity analysis

Donald Mathew Cerwick [a], Konstantina Gkritza [b,*], Mohammad Saad Shaheed [c], Zachary Hans [d]

[a] Highway Department, Burns and McDonnell, 9400 Ward Parkway, Kansas City, MO 64114, USA
[b] Lyles School of Civil Engineering, Purdue University, G167B, West Lafayette, IN 47907, USA
[c] Arizona Department of Transportation, 1615 West Jackson Street, Phoenix, AZ 85002, USA
[d] Center for Transportation Research and Education, Institute for Transportation, Iowa State University, Ames, IA 50011, USA

ARTICLE INFO

Article history:
Received 13 June 2014
Received in revised form
26 September 2014
Accepted 29 September 2014
Available online 4 November 2014

Keywords:
Truck safety
Crash severity
Mixed logit method
Latent class method
Heterogeneity

ABSTRACT

While there have been many studies analyzing crash severity, few studies have accounted for unobserved heterogeneity and compared different crash severity models. The objective of this paper is to investigate the differences between two preferred methods for accommodating individual unobserved heterogeneity, the mixed logit and latent class methods, in exploring the relationship between heavy truck crash severity and its contributing factors. To achieve this, a large sample of crash data on multiple vehicle crashes involving a heavy truck on public roadways in Iowa from 2007 to 2012 was collected. The comparison of the two methods lied on model fit, inferences, and predicted crash severity outcome probabilities. The results suggested a slight superiority of the latent class method in terms of model fit; however, the mixed logit predicted probabilities for all three levels of injury severities were closer (on average) to the observations than the ones predicted by the latent class model. Only a few notable differences in the inferences were found between the two models.

1. Introduction

One of the more frequently used methods of crash investigation in the literature is modeling crash severity using either unordered (multinomial logit, nested logit or mixed logit) or ordered (ordered logit or probit) discrete outcome models. Both ordered and unordered models have their own unique benefits and limitations, and the choice of one method over the other is governed by the availability and characteristics of the data and involves taking tradeoffs into consideration (Savolainen et al., 2011; Ye and Lord, 2014). The mixed logit method has been widely applied to analyze crash severities (Anastasopoulos and Mannering, 2011; Chen and Chen, 2011; Haleem and Gan, 2013; Islam et al., 2014; Kim et al., 2008, 2013; Malyshkina and Mannering, 2008; Manner and Wünsch-Ziegler, 2013; Milton et al., 2008; Moore et al., 2011; Morgan and Mannering, 2011; Shaheed et al., 2013; Ye and Lord, 2014; Weiss et al., 2014). The benefit of using this method is that it can accommodate individual unobserved heterogeneity by allowing parameters to differ across observations, and thus, can provide more reliable parameter estimates. A limitation of this method is that the analyst needs to assume a distribution

* Corresponding author.
E-mail addresses: dcerwick@burnsmcd.com (D.M. Cerwick), nadia@purdue.edu (K. Gkritza), mshaheed@iastate.edu (M.S. Shaheed), zhans@iastate.edu (Z. Hans).

http://dx.doi.org/10.1016/j.amar.2014.09.002
2213-6657/

relating how parameters vary across observations or (group of observations) and/or determine observation groups. This requirement is relaxed in the latent class method, where a discrete distribution, represented by a finite and specified number of mass points, is used to identify homogeneous subgroups of data. Furthermore, the latent class method requires less computation effort than the mixed logit method, but it does not account for the possibility of variation within a class as it assumes homogeneous characteristics of the within-class observations (Mannering and Bhat, 2014).

Random parameter ordered response models can also accommodate unobserved heterogeneity, though the proportional odds assumption can pose a limitation for applying this method to analyze crash injury severity. The limitations of the traditional ordered response formulation have been addressed by the recent development of the mixed generalized ordered logit model (Eluru et al., 2008) that can accommodate unobserved heterogeneity in the crash data and relax the proportion odds assumption. Recent severity studies applying generalized ordered logit model or mixed generalized ordered logit models include: Eluru (2013), Mooradian et al. (2013), Yasmin and Eluru (2013), and Yasmin et al. (2014a).

However, few studies have used a latent class approach for analyzing crash (or driver injury) severities (Chu, 2014; Eluru et al., 2012; Shaheed and Gkritza, 2014; Xie et al., 2012; Xiong and Mannering, 2013; Yasmin et al., 2014a, 2014b). Dataset-specific evidence in the literature suggest stronger statistical support for latent class logit models for exploring crash severity compared to conventional multinomial logit models (Xie et al., 2012), ordered logit models (Chu, 2014), and generalized ordered logit models (Yasmin et al., 2014a, 2014b). Direct comparisons of the latent class and mixed logit methods for crash severity analysis are not available. Similar comparisons by transport mode choice data showed that the latent class method performed statistically better than the mixed logit method (Greene and Hensher, 2003; Shen, 2009), though the evidence could not support that one model specification is better than the other.

The objective of this paper is to investigate the differences between the two methods (mixed logit and latent class) for exploring the relationship between heavy truck crash severity and its contributing factors. Three crash severity outcomes are considered: fatal and major injury, minor and possible/unknown injury, and no injury crashes. The comparison of the two methods lies on model fit, inferences, and predicted crash severity outcome probabilities, and can assist transportation safety analysts in their selection of a model specification. The estimated results and inferences can also be used by law enforcement personnel, members of the truck industry, as well as roadway and automotive engineers for improving heavy truck safety.

2. Literature review

2.1. Heavy truck crash severity models

There have been many studies that examined contributing factors to crash injury severity outcomes with an emphasis on passenger vehicle crashes. Research efforts on the application of severity models on heavy truck crashes have been surprisingly scant in the literature. Duncan et al. (1999) examined truck–passenger car collisions using an ordered probit model. Chang and Mannering (1999) estimated heavy truck crash severity outcomes using a nested logit model, nesting three severity outcomes: fatality or injury, possible injury, and no injury under nests for different levels of vehicle occupancy. Khorashadi et al. (2005) estimated a multinomial logit model to examine the relationship between heavy truck crash severity and environmental factors for crashes in urban and rural areas. Lemp et al. (2011) estimated both an ordered probit model and a heteroskedastic ordered probit model to study the impact of vehicle, occupant, driver, and environmental characteristics on the injury severity outcome of large truck crashes in the United States. In general, both models produced consistent results, however it was determined that the more flexible heteroskedastic ordered probit model performed significantly better. Qin et al. (2013) estimated three logistic regression models: multinomial, partial proportional odds, and mixed logistic models to identify the contributing factors of the severity of crashes involving large trucks in Wisconsin. The authors concluded that the partial proportional odds model was the preferred model among the three. Other, less common, discrete outcome crash severity methods employed in recent heavy track safety studies include random-parameter logit models (Chen and Chen, 2011; Islam and Hernandez, 2013; Islam et al., 2014), and non-parametric classification and regression tree models (Chang and Chien, 2013). These models were found promising for analyzing truck driver injury severity, although they were not directly compared to other model specifications in terms of prediction accuracy and inferences.

2.2. Factors affecting heavy truck crash severity

2.2.1. Crash specific characteristics

Past studies (Chang and Mannering, 1999; Lemp et al., 2011) found that the number of vehicles involved in a crash was positively associated to an increase in crash severity. Collisions between a heavy truck and another heavy truck were more likely to result in a severe crash outcome (Duncan et al., 1999; Khorashadi et al., 2005). In addition, crash contributing factors and their effect (for example, overturn/rollover, collision with a fixed object and others) on heavy truck severity have been found to differ between single and multiple vehicle crashes, leading researchers to estimate separate models based on the number of vehicles involved in the crash (Chen and Chen, 2011; Islam et al., 2014; Martensen and Dupont, 2013). Qin et al. (2013) found that the crash severity outcome of an overturned heavy truck was more severe compared to a jackknifed