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 in estimation because of very small sample sizes in some of the segments, and thus analysts tend to fall back to segmenting along 2–3 variable utmost (see Bhat, 1997 for a good discussion of these issues). To address this limitation, more advanced approaches such as clustering techniques that allow to segment based on a multivariate set of factors have been employed (Mohamed et al., 2013; Depaire et al., 2008). However, the approach still requires allocating data records exclusively to a particular segment, and does not consider the possible effects of unobserved factors that may moderate the impact of observed exogenous variables.

A *third approach* to accommodate heterogeneity is to undertake an endogenous (or sometimes also referred to as a latent) segmentation approach (see Bhat, 1997). The approach has been employed recently in the safety literature (Eluru et al., 2012; Xie et al., 2012; Xiong and Mannering, 2013; Yasmin et al., 2013). In this approach, the drivers involved in collisions are allocated probabilistically to different segments, and segment-specific injury severity models are estimated for each segment. At the same time, each segment is identified based on a multivariate set of exogenous variables. Such an endogenous segmentation scheme is appealing in many respects: (a) each segment is allowed to be identified with a multivariate set of exogenous variables, while also limiting the total number of segments to a number that is much lower than what would be implied by a full combinatorial scheme of the multivariate set of exogenous variables, (b) the probabilistic assignment of drivers to segments explicitly acknowledges the role played by unobserved factors in moderating the impact of observed exogenous variables, and (c) there is no need to specify a distributional assumption for the coefficients (Greene and Hensher, 2003). This third approach may be viewed as a combination of the two earlier approaches, in that it considers a multivariate set of exogenous variables in the segmentation and also allows unobserved variable effects to moderate the impact of exogenous variables. In fact, the third approach is equivalent to specifying a (discrete) non-parametric distribution on the coefficients (rather than the continuous parametric distribution assumption of the first approach), while also allowing the non-parametric distribution shape to be a function of a multivariate set of exogenous variables.[4]

In summary, the current study contributes to the literature on driver injury severity in two ways. *First*, it examines the driver injury severity level using a comprehensive set of exogenous variables (*empirical contribution*). *Second*, it formulates and estimates a LSGOL model that accommodates observed and unobserved heterogeneity, and relaxes the constant threshold assumption of the traditional ordered response formulation (*methodological contribution*).

3. Model framework

The analysis in this paper is undertaken at the level of drivers involved in a crash. That is, we focus on driver-level injury severity in a crash. Thus, in the case of a crash involving a single vehicle with an object, there is one driver record with the corresponding injury severity level sustained by the driver. In the case of a crash involving multiple drivers, each driver contributes a record, along with the injury severity level sustained by the driver.

The framework used for modeling driver-level injury severity assumes that drivers can be implicitly sorted into S relatively homogenous (but latent to the analyst) segments based on characteristics of the crash. Within each segment, the effects of exogenous variables are fixed across drivers in the segment. Let s be the index for segments ($s = 1, 2, \ldots S$)), i be the index for drivers ($i = 1, 2, \ldots, N$), and j be the index for driver injury severity levels ($j = 1, 2, \ldots, J$). The crash outcomes are analyzed using a GOL model within each segment. Across segments, the parameters of the GOL model vary. In the GOL response model, conditional on driver i belonging to segment s, the discrete injury severity levels (y_i) are assumed to be a mapping (or partitioning) of an underlying continuous latent variable (y_i^*) as follows:

$$y_i^*|(i \in s) = \boldsymbol{X}_i \boldsymbol{\beta}_s + \varepsilon_{is}, \; y_{is} = j, if \; \tau_{i,j-1,s} < y_i^* < \tau_{i,j,s} \tag{1}$$

where, $\boldsymbol{X}_i$ is a row vector of exogenous variables, $\boldsymbol{\beta}_s$ is a corresponding column vector of unknown parameters specific to segment s

ε_{is} is a segment-specific idiosyncratic random disturbance term assumed to be identically and independently standard logistic

$\tau_{i,j,s}$ ($\tau_{i,0,s} = -\infty$, $\tau_{i,J,s} = +\infty$) represents the segment-specific upper threshold associated with driver i and severity level j, with the following ordering conditions: $(-\infty < \tau_{i1,s} < \tau_{i2,s} < \ldots < \tau_{iJ-1,s} < +\infty) \forall \; s = 1, 2, \ldots S$.

To maintain the ordering conditions and allow the thresholds to vary across drivers within each segment, Eluru et al. (2008) propose the following non-linear parameterization of the thresholds as a function of exogenous variables:

$$\tau_{ij,s} = \tau_{ij-1,s} + exp(\boldsymbol{\delta_{js} Z_{is}}) \tag{2}$$

where $\boldsymbol{\delta}_{js}$ is a segment-specific and injury level-specific row vector of parameters to be estimated and $\boldsymbol{Z_{is}}$ is a corresponding column vector of segment-specific exogenous variables ($\boldsymbol{Z_{is}}$ includes a constant as its first element, with the corresponding coefficient being $\boldsymbol{\delta}_{js,1}$; for identification, we need $\boldsymbol{\delta}_{1s,-1}$ to be a row vector of zero values, where $\boldsymbol{\delta}_{1s,-1}$ is a sub-vector of the

[4] The more recent work of Xiong and Mannering (2013) proposes a latent segmentation model that further specifies unobserved heterogeneity in each segment-level injury severity model using a continuous multivariate normal distribution for the coefficients. This is tantamount to a discrete mixture-of-normals approach. Though, we do not account for unobserved heterogeneity in the segment level models in this paper, we propose to employ a GOL framework that can accommodate the more realistic case of injury reporting in more than two injury severity levels (the study by Xiong and Mannering (2013) on the other hand, was a binary choice model of injury severity).

Table 1
Crash database sample statistics.

Variables	Sample share	
	Frequency	Percentage
Crash characteristics		
Collision object		
Small object	120	2.338
Large object	710	13.835
Collision with animals	207	4.034
Collision with Moving vehicle	4095	79.793
Trajectory of vehicle's motions		
Going straight	2689	52.397
Other movement	2443	47.603
Manners of collision		
Rear-ender	469	9.139
Rear-ended	582	11.341
Near-sideswipe	96	1.871
Near-angular	729	14.205
Short-side angular	1133	22.077
Far-side	783	15.257
Head-on	303	5.904
Other collisions (struck object)	1037	20.207
Driver characteristics		
Driver age		
Age less than 25	1350	26.306
Age 26–65	3282	63.952
Age above 65+	500	9.743
Driver gender		
Female	2420	47.155
Male	2712	52.845
Restraint system use		
Seat belt not used	179	3.488
Seat belt used	4953	96.512
Vehicle characteristics		
Vehicle type		
Sedan	3666	71.434
Station wagon	900	17.537
Utility	447	8.710
Panel van	119	2.319
Vehicle age		
Vehicle age less than 11	2907	56.645
Vehicle age 11 and above	2225	43.355
Roadway design attributes		
Type of road surface		
Paved	4901	95.499
Unpaved	231	4.501
Traffic control device		
No control	3088	60.171
Signal	1107	21.571
Pedestrian control	34	0.663
Roundabout	181	3.527
Stop sign	171	3.332
Yield sign	473	9.217
Other traffic control	78	1.520
Speed zone		
Low speed zone	961	18.726
Medium speed zone	3412	66.485
High speed zone	759	14.790
Environmental factors		
Season		
Summer	1305	25.429
Autumn	1346	26.228
Winter	1265	24.649
Spring	1216	23.694
Time of day		
Morning peak	691	13.465
Off peak	1695	33.028
Evening peak	1250	24.357
Late evening	1160	22.603

Table 1 (*continued*)

Late night	336	6.547
Weather condition		
Clear	4332	84.412
Rain/snow	646	12.588
High wind	85	1.656
Other weather condition	69	1.345
Lighting condition		
Daylight	3558	69.330
Dusk/dawn	361	7.034
Dark-lighted	942	18.355
Dark-unlighted	271	5.281
Situation factors		
Presence of passengers		
No passenger	598	11.652
One passenger	2395	46.668
Two passengers	1136	22.136
More than two passengers	1003	19.544
Driver ejected out of the vehicle		
Ejected out	39	0.800
Did not eject out	5093	99.200

vector $\boldsymbol{\delta}_{1s}$ minus the first element). The traditional ordered logit (OL) model assumes that the thresholds $\tau_{i,j,s}$ remain fixed across drivers ($\tau_{ij,s} = \tau_{j,s}\ \forall\ i$) for each segment; that is, it assumes that $\boldsymbol{\delta}_{js,-1}$ has all zero elements for all j values and all s values.

Given the above set-up, the probability that driver i suffers an injury severity outcome j, conditional on driver i belonging to segment s, may be written as:

$$P_i(j)|s = \Lambda(\tau_{ij-1,s} + exp(\boldsymbol{\delta_{js}Z_{is}}) - \boldsymbol{X}_i\boldsymbol{\beta}_s) - \Lambda(\tau_{ij-2,s} + exp(\boldsymbol{\delta_{j-1,s}Z_{is}}) - \boldsymbol{X}_i\boldsymbol{\beta}_s) \tag{3}$$

where $\Lambda(.)$ represents the standard logistic cumulative distribution function.

Of course, the analyst does not observe the segment to which driver i belongs. So, the analyst specifies this segment assignment to be a function of a column vector of observed crash factors $\boldsymbol{\eta_i}$. To also acknowledge the presence of unobserved factors that may influence this assignment, the analyst develops an expression for the probability of driver i belonging to segment s. While many parametric expressions may be used for this probability expression (the only requirement is that the probabilities sum to one across the segments for each driver i), the most commonly used form corresponds to the multinomial logit structure (see Bhat, 1997; Greene and Hensher, 2003; Eluru et al., 2012):

$$P_{is} = \frac{\exp[\boldsymbol{\alpha_s\eta_i}]}{\sum_s \exp[\boldsymbol{\alpha_s\eta_i}]} \tag{4}$$

where $\boldsymbol{\alpha_s}$ is a row vector of parameters to be estimated. Then, the unconditional probability of driver i leading up to injury severity level j can be written as

$$P_i(j) = \sum_{s=1}^{S} (P_i(j)|s) \times (P_{is}) \tag{5}$$

The log-likelihood function for the entire dataset can be written as

$$L = \sum_{i=1}^{N} \log\left[\sum_{s=1}^{S} (P_i(j)|s) \times (P_{is})\right] \tag{6}$$

The parameters to be estimated in the LSGOL model are the segment parameters ($\boldsymbol{\beta}_s$ &$\boldsymbol{\delta}_{js}$), the class probability parameters ($\boldsymbol{\eta_i}$) for each s, and the appropriate number of segments S. For identification reasons, we need to restrict one of the $\boldsymbol{\delta_{js}}$ vectors to zero. It is worthwhile to mention here that the estimation of latent segmentation based models using quasi-Newton routines can be computationally unstable (see Bhat, 1997 for a discussion). The estimation of such models requires employing good starting values for the estimation procedure. Hence, for our analysis, the log-likelihood function and its corresponding gradient function were coded in the Gauss Matrix programming language. The coding of the gradient function ensures the reduction in instability associated with such an estimation process.

4. Data

4.1. Data source

Data for our empirical analysis is sourced from the Victoria crash database of Australia for the years 2006 through 2010. The data includes information reported by Victorian police officers for crashes involving at least one motor vehicle traveling on a roadway and resulting in property damage, injury or death, which are then compiled by VicRoads (a statutory body responsible for road transport in the state of Victoria). For the five years, the crash database has a record of 67,809 crashes involving 118,842 motor vehicles and 166,040 individuals, resulting in 1550 fatalities and 87,855 injuries to the crash victims. A four point ordinal scale is used in the database to represent the injury severity of individuals involved in these crashes: (1) no injury, (2) minor injury, (3) serious injury and (4) fatal injury.

4.2. Sample formation and description

This study is focused on the injury severity outcome of drivers, who are involved in either a single or a two passenger vehicle collisions. The crashes that involve more than two passenger vehicles are excluded from the analysis (about 9.9% of the sample). The crashes that involve commercial vehicles are also excluded to avoid the potential systematic differences between the crashes involving commercial and non-commercial driver groups. The final dataset, after removing records with missing information for essential attributes, consisted of 42,812 driver records. The final sample had a very small percentage of records that involved a fatally injured driver (about 1% of total crashes). Therefore, both the fatal and serious injury category levels are merged together in the current analysis.

From the dataset of 42,812 driver records, a sample of 5132 records is randomly drawn for the purpose of estimating models and 37,680 records are set aside for the purpose of validation. In the final estimation sample, the distribution of driver injury severity levels is as follows: no injury 41.8%, minor injury 36.9% and serious/fatal injury 21.3%. Table 1 offers a summary of the sample characteristics of the exogenous factors in the estimation dataset. From the descriptive analysis, we observe that a large portion of crashes involve short-side angular collisions (22.1%), and at locations with no traffic control (60.1%), in a medium speed zone location (66.5%), during the off peak period (33%), in clear weather (84.4%), in daylight (69.3%) and in the presence of at least one passenger in the vehicle (88.4%). The majority of drivers are adult (63.9%), use seat-belts (96.5%) and drive a sedan (71.4%). The drivers are somewhat more likely to be male than female (male 52.8% versus female 47.1%). It is also quite interesting to note that the share of vehicles that are more than 10 years old is quite large (43.4%).

5. Empirical analysis

5.1. Variables considered

The collision attributes considered in the empirical study can be grouped into six broad categories: crash characteristics, driver characteristics, vehicle characteristics, roadway design attributes, environmental factors and situational factors.

The *crash characteristics* examined were collision object (small object, large object, animal, and moving vehicle), trajectory of vehicle's motion (going straight or other movement), and manner of collision. The database compiles the manner of collision at a high level of disaggregation, and as a combination of collision type (rear-end, sideswipe, angular, and head-on) and the initial point of contact. A schematic diagram of the initial point of impact relative to the driver's seat position is shown in Fig. 1 (the collision type and the initial point of impact are computed relative to the position of driver in the vehicle). Based on the collision type and the point of impact, we identified seven categories for the "manner of collision ": Rear-ender (the rear vehicle that is involved in a rear-end collision), Rear-ended (the front vehicle that is involved in the rear-end collision), Near-sideswipe (sideswipe/near-side), Near-angular (angular/near-side), Short-side angular (angular/

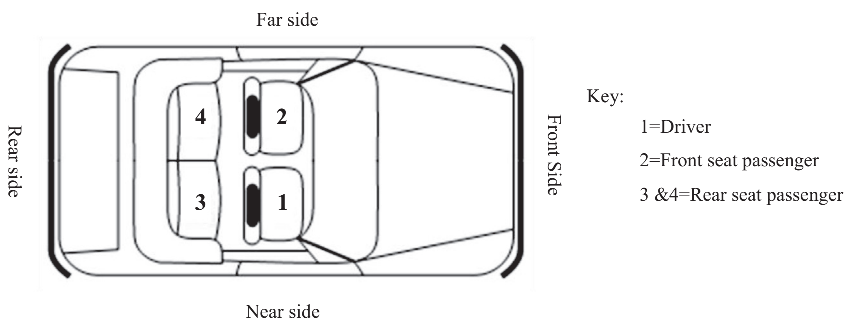

Fig. 1. Schematic diagram of initial point of impact relative to the driver's seat position.

front and rear side), Far-side (angular and sideswipe/far-side) and Head-on (head-on/front side). The reader would note that, in a two vehicle crash it is possible that the individual drivers might have different effects in the manner of collision variable for a same type of crash. For example, in a rear-end collision (collision type), one of the vehicles will be classified as rear-ended and the other will be classified as a rear-ender.

The *driver characteristics* included are driver gender, age and seat belt use information. *Vehicle characteristics* considered are vehicle type (characterized as sedan, station wagon, utility and panel van) and vehicle age. The *roadway design attributes* considered in the analysis are road surface type, presence of traffic control device, and presence of a speed zone (speed zone is a length or an area of road along which a signposted regulatory speed limit applies). The *environmental factors* included are season, time of day, weather condition, and lighting condition. Finally, the *situational factors* included in the model are the number of passengers and whether or not the driver was ejected. The final specification of the model development was based on combining the variables when their effects were not statistically different and by removing the statistically insignificant variables in a systematic process based on statistical significance (90% significance level). For continuous variables, linear, polynomial and spline forms were tested.

5.2. Variable considered for segmentation of crashes

The proposed modeling approach theoretically can accommodate classification of segments based on the universal set of variables. However, in our analysis, we consider segmentation based only on traffic crash characteristics for two reasons. First, while it is plausible to consider all attribute sets in the latent segmentation consideration, the estimation of latent segmentation models with the entire attribute set is likely to result in convergence challenges as well as difficulty in interpreting the results (see Sobhani et al. (forthcoming) and Eluru et al. (2012) for discussions on challenges associated with latent segmentation models). Second, in the safety literature, there has been substantial interest in exploring the impact of crash characteristics on injury severity. In fact, many previous injury severity studies have focused only on a specific type of crash, which is tantamount to specifying separate injury severity models for each crash type.[5] While these research attempts are very useful, the approach results in models where injury severity records are exclusively allocated to about various segments (defined by crash type) and analyzed through separate severity models for each segment. However, doing so implies that the model estimation is undertaken on a relatively small sample of the accident records for at least some crash types. In our paper, we offer an alternate approach by examining segmentation on the basis of crash characteristics (collision object, the trajectory of vehicle's motion, and manner of collision), and analyze driver-level injury severity within each segment using other crash attributes. The approach allows us to retain a smaller number of segments while assigning individuals probabilistically. In this manner, we ensure that the entire sample is utilized in model estimation for each segment. Thus, the latent segmentation based model provides an elegant and effective approach to study the influence of crash characteristics through segmentation, while acknowledging the need for separate injury severity models for each segment.

5.3. Model specification and overall measures of fit

The empirical analysis involves the estimation of four models: (1) the ordered logit (OL) model, (2) the generalized ordered logit (GOL) model, (3) the latent segmentation based ordered logit (LSOL) model, and (4) the latent segmentation based generalized ordered logit (LSGOL) model. Prior to discussing the estimation results, we compare the performance of these models in this section. The model comparisons are undertaken in two stages. First, we determine the appropriate latent segmentation scheme for the OL and GOL models. Second, we compare the traditional (unsegmented) OL and GOL models with the more general latent models (LSOL and LSGOL) obtained from the first step.

5.3.1. Determining the appropriate latent segmentation model

The estimation of the latent segmentation based model involves the probabilistic assignment of the drivers involved in collisions into a given number of segments (S) based on the available exogenous variables. In the application of these models, determining the appropriate number of segments is a critical issue with respect to interpretation and inferences. Therefore, we estimate these models with increasing numbers of segments ($S=2, 3, 4, \ldots$) until an addition of a segment does not add value to the model in terms of data fit. Many of the earlier studies suggest that the Bayesian Information Criterion (BIC) is the most consistent information criterion (IC) among all other traditionally used ICs (AIC, AICc, adjusted BIC) for segment analysis (Nylund et al., 2007; Bhat, 1997; Collins et al., 1993). The advantage of using the BIC is that it imposes substantially higher penalty than other ICs on over-fitting. Thus, in the current study context, the most appropriate number of segments in the LSOL and LSGOL models is determined based on the BIC measure.

[5] For instance, Head-on collision: (Gårder, 2006; Conroy et al., 2008; Zuxuan et al.. 2006; Zhang and Ivan, 2005; Rear-end collision: Khattak, 2001; Yan et al., 2005; Das and Abdel-Aty, 2011; Abdel-Aty and Abdelwahab, 2003); Ran-off road/Hit-fixed object collision: (Ray, 1998; McGinnis et al., 2001; Holdridge et al., 2005); Angular collision: (Jin et al., 2010; Chipman, 2004).

Table 2
Segment characteristics and mean values of segmentation variables for LSGOL model.

Components	Segments	
	Segment 1	Segment 2
Segmentation characteristics		
Driver population share	0.418	0.582
Injury severity		
Property damage only	0.102	0.640
Minor injury	0.578	0.239
Serious injury and fatal injury	0.320	0.121

We estimated the LSOL and LSGOL models with $S=2$ (LSOL II and LSGOL II models) and 3 (LSOL III and LSGOL III models) segments and computed the BIC values for each of these models. The BIC for a given empirical model is equal to

$$\text{BIC} = -2LL + K\ \ln(Q) \tag{7}$$

where LL is the log-likelihood value at convergence, K is the number of parameters, and Q is the number of observations. The model with the *lower* BIC is the preferred model. For the LSOL model, the computed BIC values with 2 and 3 segments are 10049.72 (37 parameters) and 10257.93 (34 parameters), respectively. The BIC values for the LSGOL model with 2 and 3 segments are 10024.01 (41 parameters) and 10385.26 (31 parameters), respectively. Thus, we selected two segments as the appropriate number of segments for both the LSOL and LSGOL models.

5.3.2. Comparison across all models – non-nested test

To evaluate the performance of the estimated OL, GOL, LSOL and LSGOL models, the BIC values are computed as shown in Eq. (7). Also, the AICc values are computed for each of the four models as

$$\text{AIC}_c = 2K - 2\ \ln(LL) + \frac{2\ K(K+1)}{(Q\ - K - 1)} \tag{8}$$

Model with *lower* BIC and AICc values are preferred to models with higher values for these ICs. The BIC (AICc) values for the final specifications of the OL, GOL, LSOL and LSGOL models are 10086.40 (9929.59), 10048.06 (9769.44), 10049.72 (9808.17) and 10024.01 (9756.42), respectively. The comparison exercise clearly highlights the superiority of the LSGOL model in terms of data fit compared to all the other models.

5.4. Estimation Results

In presenting the effects of exogenous variables in the model specification, we will restrict ourselves to the discussion of the LSGOL model. Table 3 presents the estimation results. Following Bhat (1997), we first present some descriptive characteristics of the two segments in the LSGOL model, before proceeding to a discussion of the variables that impact segmentation and the injury severity levels of drivers within each segment.

5.4.1. Descriptive Characteristics of the segments in the LSGOL model

To delve into the characteristics that delineate the segments and to understand the characteristics of each segment, the model estimates are used to generate information on: (1) the population share of each of the two segments and (2) the overall injury severity level shares within each segment. These estimates are presented in Table 2. The population share or the size of each segment is computed as:

$$G_s = \frac{\sum_i P_{is}}{N} \tag{9}$$

where N is the total number of drivers in the estimation sample. From the first row of Table 2 labeled “Driver population share”, it is evident that a driver is more likely to be assigned to segment 2 than to segment 1. Further, the driver injury severity outcome probabilities, conditional on assignment to a segment, are obtained using Eq. (3). The segment-specific injury outcome shares are then computed by taking the average (across all drivers) of the driver-specific probabilities associated with each injury outcome level. The results are presented in the second row panel of Table 2. It is clear that a driver, if allocated to segment 1, is likely to be involved in a more severe crash than if allocated to segment 2. Thus, we may label segment 1 as the “*high risk segment*” and segment 2 as the “*low risk segment*”.

5.4.2. Latent segmentation component

The latent segmentation component determines the relative prevalence of each class, as well as the probability of a driver being assigned to one of the two latent segments based on the crash characteristics. In our empirical analysis, the crash characteristics that affect the allocation of drivers to segments include collision object, trajectory of vehicle's motion, and manner of collision. The results in Table 3 provide the effects of these crash characteristics, using the high risk segment

Table 3
LSGOL estimates.

Variables	**Segment 1**				**Segment 2**			
					Estimate		***t*-Stat**	
Segmentation components								
Constant	–		–		1.435		5.679	
Crash characteristics								
Collision object (Base: other moving vehicle and animals)								
Small object	–		–		−3.531		−4.552	
Large object	–		–		−4.035		−8.342	
Trajectory of vehicle's motions (Base: Other movement)								
Going straight	–		–		−0.480		−3.423	
Manner of collision (Base: rear-ender and short side-angular)								
Rear-ended	–		–		−1.574		−7.172	
Near-angular	–		–		−0.826		−4.427	
Head-on	–		–		−1.055		−4.402	
Far-side	–		–		0.530		2.197	
Variables	**Latent propensity**		**Threshold**		**Latent propensity**		**Threshold**	
	Estimates	***t*-Stat**	**Estimate**	***t*-Stat**	**Estimates**	***t*-Stat**	**Estimate**	***t*-Stat**
Constant	3.859	7.962	1.468	13.571	−1.746	−6.237	0.307	2.650
Injury severity components								
Driver characteristics								
Driver age (Base: age 25–64)								
Age less than 25	−0.734	−2.478	−0.137	−1.643	–	–	–	–
Age above 65+	0.885	3.779	–	–	0.432	2.629	−0.498	−2.533
Driver gender (Base: male)								
Female	–	–	–	–	1.189	8.003	0.316	2.740
Restraint system use (Base: seat belt used)								
Seat belt not used	–	–	−0.197	−2.161	0.717	2.472	–	–
Vehicle characteristics								
Vehicle type (Base: sedan)								
Station wagon	–	–	–	–	−0.744	−4.452	–	–
Utility	–	–	–	–	−0.939	−3.313	–	–
Panel van	−1.109	−2.495	–	–	–	–	–	–
Vehicle age (Base: vehicle age less than 10)								
Vehicle age 11 and above	–	–	–	–	0.349	3.309	–	–
Roadway design attributes								
Type of road surface (Base: paved)								
Unpaved	−1.338	−1.901	–	–	–	–	–	–
Traffic control device (Base: none traffic control and other control device)								
Pedestrian control	−2.135	−1.974	–	–	–	–	–	–
Roundabout	−1.292	−3.016	–	–	–	–	–	–
Stop sign	0.900	2.579	–	–	−1.326	−3.273	–	–
Yield sign	–	–	–	–	−0.471	−2.527	–	–
Speed zone (Base: low speed ≤ 50 km/h)								
Medium speed (60–90 km/h)	–	–	–	–	0.356	2.469	–	–
High speed (≥ 100 km/h)	–	–	−0.189	−4.468	1.244	5.934	–	–
Environmental factors								
Season (Base: spring, summer, fall)								
Winter	–	–	–	–	0.332	2.886	–	–
Time of day (Base: morning peak, off peak and late evening)								
Evening peak	−0.751	−4.61	–	–	–	–	–	–
Late night	0.586	3.146	–	–	–	–	–	–
Weather condition (Base: clear)								
Rain/snow	–	–	–	–	0.314	2.069	–	–
High wind	–	–	–	–	0.736	2.174	–	–
Lighting condition (Base: Daylight)								
Dusk/dawn	–	–	–	–	0.488	2.711	–	–
Dark-lighted	−0.897	−2.932	−0.408	−4.124	–	–	–	–
Situational factors								
Presence of passengers (Base: no passenger)								
One passenger	−0.532	−4.040	–	–	–	–	–	–
Two passengers	−2.201	−5.877	−0.426	−3.947	–	–	–	–
Driver ejected out of the vehicle	–	–	−0.929	−2.624	–	–	–	–

(segment one) as the base segment. Thus, a positive (negative) sign for a variable indicates that crashes with the variable characteristic are more (less) likely to be assigned to the low risk segment relative to the high risk segment, compared to crashes that correspond to the characteristic represented by the base category for the variable. The positive sign on the constant term does not have any substantive interpretation, and simply reflects the larger size of the low risk segment compared to the high risk segment.

The results for the "collision object variables" indicate an increased likelihood of drivers being assigned to the high risk segment in case of a collision with stationary objects (small or large object) compared to a collision with another moving vehicle. In terms of the trajectory of the vehicle's motion, the driver of a vehicle traveling straight through just prior to a crash is at a higher risk of severe injury relative to drivers making other turning movements. This result is to be expected because straight-through drivers are likely to be traveling at higher speeds.

Consistent with several previous studies (Chiou et al., 2013; Khattak, 2001), our analysis also shows that being the driver of the rear-ended vehicle in a rear-end collision increases the probability of a high risk crash. The driver of the vehicle is likely to be pushed backward into the seat when struck by the following vehicle, which results in higher probability of whiplash or neck injury due to the continuous movement of the neck at a different speed than the head and the rest of the body (Khattak, 2001; Krafft et al., 2000; Nordhoff, 2005). Thus, the biomechanics of this type of collision explains the increased probability of a high risk crash.

The result associated with a head-on collision also reflects an increased likelihood of assigning the drivers involved in the crash to the high risk segment. Head-on collisions are often caused by drivers violating traffic rules, crossing the centerline by mistake and losing control of their vehicles (Zhang and Ivan, 2005). The pre-impact speed vectors of motor vehicles are directed in opposing directions during a head-on collision, resulting in greater dissipation of kinetic energy and heavier deformation of motor vehicle bodies (Prentkovskis et al., 2010), resulting in higher risk of injury (Tay and Rifaat, 2007; Gårder, 2006). The drivers who are involved in a near-angular collision also are likely to be assigned to the high risk segment. These crashes impose more risk on the driver due to the angle of impact (Jin et al., 2010) and the greater force of impact (Tay and Rifaat, 2007). Moreover, there is less collapsible structure between the striking force and the drivers, which might result in significant passenger compartment intrusion and the direct loading of impact resulting in serious chest and abdominal injury (Mackay et al., 1993; McLellan et al., 1996).

For the far-side manner of collision, the result indicates that this kind of collision reduces the propensity of drivers being in the high risk segment. The significant gap between the collision impact point and driver position might lessen the direct impact of force as a large amount of kinetic energy is absorbed by the vehicle (Sobhani et al., 2011), thereby reducing the risk of high injury severity.

5.4.3. Injury severity component: high risk segment (segment 1)

The injury severity component within the high risk segment (segment 1) is discussed in this section. The two columns of the corresponding segment in Table 3 represent the latent injury risk propensity and the threshold demarcating the minor injury level from the serious/fatal injury level, respectively.

Driver Characteristics: The age of drivers involved in the collision has a significant influence on crash severity. The estimation results indicate a reduction in the risk propensity for young drivers (age less than 25). But the impact of driver age on the threshold demarcating the minor injury and serious/fatal injury levels indicates that the distance between these thresholds get contracted for young drivers relative to other adult drivers (age 25 to 64). The net implication is that young drivers in this first segment have a higher probability of sustaining no injury, and a lower overall probability of some kind of an injury (minor injury or serious/fatal injury). But the contraction of the distance between the thresholds implies that the effect of age on the minor injury and serious/fatal injury categories is crash and driver-specific; for some contexts, the minor injury probability can increase with a concomitant decrease in the serious/fatal injury probability, while for other contexts the reverse can hold. This highlights the advantage of a GOL framework that allows for flexible exogenous variable impacts. The lower probability of injury among young adults may reflect the higher physiological strength of young drivers in withstanding crash impacts (Xie et al., 2012; O'Donnell and Connor, 1996; Castro et al., 2013), while the higher probability of serious/fatal injuries in some crashes may represent the lack of driving experience of young drivers because of which they do not take evasive maneuvers to reduce the impact of a crash in the making. Of course, other explanations are also possible. The parameter characterizing the effect of old age (age ≥ 65) on driver injury severity suggests a higher injury risk propensity for this group of drivers relative to other adult individuals. As indicated in earlier studies (Bédard et al., 2002; Kim et al., 2013; Zhang et al., 2000; Williams and Shabanova, 2003), older drivers tend to be slow in reacting to hazardous situations, may not be able to withstand crash impact forces well, and may suffer cognitive impairment and other medical conditions; all or some of these factors might contribute to their higher injury severity risk. It is interesting to note here that driver gender has no significant influence on crash severity outcome for segment 1. A plausible reason for this effect may be the additional physiological strength of male drivers (compared to female drivers) is less likely to lessen the effect of a more severe crash. Finally, in the category of driver characteristics, seat belt use significantly influences driver injury severity. The negative effect of this variable on the threshold separating the minor injury and serious/fatal injury levels indicates an increased likelihood of serious/fatal injuries for the drivers not wearing seat belts. The result can be explained by the reduction in restraint as well as possible high-risk driving behavior of those not using seatbelts (Obeng, 2008; Yau, 2004; Yasmin et al., 2012; Eluru and Bhat, 2007).

Vehicle Characteristics: The only vehicle characteristic influencing driver injury severity for the high risk segment is vehicle type. Table 3 shows that drivers in panel vans are associated with a lower injury risk propensity than drivers in other vehicle types, presumably because panel vans are larger and may offer more protection (Kockelman and Kweon, 2002; Xie et al., 2009; Eluru et al., 2010; Wang and Kockelman, 2005; Fredette et al., 2008).

Roadway Design Attributes: The roadway design attributes indicate a lower injury risk propensity for crashes occurring (a) on unpaved roads (perhaps because of very low speeds on such roads), (b) at intersections with some form of control for pedestrian movement and at roundabouts (relative to other types of intersections). The last result regarding roundabouts may be the consequence of moderated vehicle speeds and the angular movements at these locations, which can result in safer impact angles at the time of collision (Retting et al., 2001; Persaud et al., 2001; Chipman, 2004). On the other hand, crashes at stop-sign controlled intersections seem to increase injury severity risk relative to crashes at other intersections, attributable perhaps to non-compliance to stop signs and judgment problems (Chipman, 2004; Retting et al., 2003). Also, crashes occurring on very high speed roads, not surprisingly, lead to a high probability of serious/fatal injuries.

Environmental Factors: Time-of-day and lighting conditions are two of the environmental factors that significantly influence driver injury severity for the high risk segment. Injury risk reduces during the evening, but increases during the late night. The former effect may be a result of traffic congestion and slow driving speeds, because of which, when a crash does happen, the injury sustained tends to be rather mild. The latter result associated with late night crashes is well established in the literature; attributable to reduced visibility, fatigue, higher incidence of alcohol use, longer emergency response times, higher driver reaction time, and increased traffic speed (Plainis et al., 2006; Arnedt et al., 2001; Helai et al., 2008; Hu and Donnell, 2010; Kockelman and Kweon, 2002; de Lapparent, 2008). The lighting condition effect show a higher probabilty of no injury crashes during dark-lighted conditions, perhaps due to more cautious driving relative to broad daylight. As with the young driver effect, the impact of this variable on the other two injury severity categories is context-dependent.

Situational Factors: The presence of one or more passengers increases the probability of no injury, relative to the case of driving alone. This may be associated with public self-consciousness, where individuals behave and drive more responsibly with others around (Eluru et al., 2010). As expected, drivers who are ejected out of their vehicle during a crash have a high probability of sustaining serious/fatal injuries for the high risk segment.

5.4.4. Injury severity component: low risk segment (segment 2)

The injury severity component within the low risk segment (segment 2) is discussed in this section. The impact of exogenous variables within the low risk segment is different (for some variables) in magnitude as well as in sign from the impact of exogenous variables within the high risk segment. Also, the number of variables moderating the effect is different across the two segments.

Driver Characteristics: For the low risk segment, the influence of driver age on crash severity is along expected lines. We find that older drivers are associated with higher likelihood of severe crashes compared to other adult drivers as also seen in the other segment. Unlike the high risk segment, driver gender has a significant influence on driver injury severity outcome for low risk segment. The coefficient corresponding to driver gender of passenger vehicle reflects higher injury risk propensity for female drivers compared to male drivers perhaps because females are less capable of bearing physical and mental trauma compared to males (Evans, 2004; Sivak et al., 2010; Xie et al., 2009; Chen and Chen, 2011). As expected, our analysis showed an unequivocal benefit from employing seat belts. It is interesting to note that the seat belt variable affects the driver injury severity in different ways for the two segments.

Vehicle Characteristics: In the low risk segment, the results for the vehicle type reveal that the drivers of both station wagon and utility vehicles have a lower injury risk propensity, perhaps due to the larger weight of these vehicles. The vehicle age estimate demonstrates that drivers in older vehicles (Vehicle age 11 and above) have a higher risk propensity compared to the drivers in newer vehicles (vehicle age $\leq$ 10 years). The higher injury risk of drivers from older vehicles might be attributed to the absence of safety features, presence of mechanical defects, and/or the involvement of suspended and unlicensed drivers in these vehicles (Lécuyer and Chouinard, 2006, Kim et al., 2013; Islam and Mannering, 2006).

Roadway Design Attributes: The presence of traffic control devices significantly affect the severity of crashes. For both stop and yield sign variables, the corresponding latent propensity coefficients are negative indicating a lower injury risk; reduced traveling speed of drivers might be a plausible reason for such result. However, the effect of stop sign is strikingly different in the low risk segment compared to the impact of stop sign in the high risk segment. The different impacts in the two segments for stop sign highlight how the same variable can have distinct influence on injury severity based on the segment to which the driver is allocated.

The results for speed zone indicate that the injury propensity is higher for crashes occurring in zones with medium and higher speed limits relative to crashes occurring in lower speed limit zones. As is expected, within the two speed categories considered the higher speed category has a larger impact relative to the medium speed category. Such rapid increase in severity with progressive increase in speed limit has also been documented empirically by many earlier studies (Eluru et al., 2010; Chen et al., 2012; Tay and Rifaat, 2007).

Environmental Factors: The findings of the low risk segment indicate that if collisions occur in the winter season, the consequence is likely to be more injurious as compared to the accident in non-winter seasons (spring, summer and autumn). The prevalent adverse and damp weather conditions in winter might pose such risk on Victorian drivers. With respect to weather condition, the results presented in Table 3 indicate that the rainy/snowy weather condition results in more severe crashes compared to the clear weather, which may be attributed to the unfavorable driving conditions resulting from reduced visibility and reduced friction of the road surface. The results also reveal that injury propensity is higher for drivers in the presence of high wind compared to crashes occuring during clear weather. It is possible that under high wind conditions drivers suddenly lose vehicle control and sideswipe or run-off from their designated routes (Jung et al., 2011; Young and Liesman, 2007; Khattak and Knapp, 2001). With respect to lighting condition, the likelihood of driver injury risk propensity is found to be higher during dawn/dusk compared to other lighting conditions. This may be associated with sunglare during dawn/dusk period (Jurado-Piña et al., 2010; Gray and Regan,2007).

Situational factors: With respect to the situational factors, none of the variables are found to affect injury severity in the low risk segment.

6. Elasticity effects

The parameter estimates of Table 3 do not directly provide the impact of exogenous variables on injury severity categories. On the other hand, the aggregate-level elasticity effects quantify the effects of these variables on driver injury severity outcomes. For this purpose, we compute the aggregate level "elasticity effects" for all independent variables (see Eluru and Bhat (2007) for a discussion on the methodology for computing elasticities) and present the computed elasticities in Table 4. The effects are presented by injury severity categories for both the LSOL and LSGOL models for comparison purpose. The results in the table can be interpreted as the percentage change (increase for positive sign and decrease for negative sign) in the probability of the crash severity categories due to the change in that specific exogenous variable.

Table 4
Elasticity effects.

Variables	LSOL			LSGOL		
	No injury	Minor injury	Serious/fatal injury	No injury	Minor injury	Serious/fatal injury
Crash characteristics						
Rear – ended	–42.536	30.199	30.639	–38.906	28.750	26.138
Near-sideswipe	–18.587	13.139	13.488	–	–	–
Near-angular	–24.690	17.491	17.850	–19.507	14.382	13.162
Head-on	–31.219	22.105	22.591	–25.717	18.958	17.357
Far-side	–	–	–	11.630	–8.488	–7.998
Small object	–68.129	48.024	49.677	–68.894	50.634	46.768
Large object	–82.773	56.013	64.419	–81.995	57.905	59.764
Going Straight	–12.860	9.054	9.396	–10.607	7.776	7.234
Driver characteristics						
Age less than 25	3.170	4.548	–14.126	12.339	–7.676	–10.800
Age above 65+	–20.643	–16.127	68.500	–17.328	–22.938	73.855
Female	–32.773	21.423	26.821	–33.961	27.437	18.742
Seat belt not used	–19.413	–4.348	45.571	–20.767	–4.876	49.151
Vehicle characteristics						
Station wagon	18.506	–12.480	–14.478	19.257	–12.312	–16.277
Utility	23.853	–16.822	–17.378	23.021	–15.277	–18.486
Panel van	13.191	6.629	–37.366	11.199	8.029	–35.904
Vehicle age 11 and above	–8.629	5.284	7.684	–9.667	5.566	9.241
Roadway design attributes						
Unpaved	15.147	5.989	–40.081	14.465	7.321	–41.066
Pedestrian control	22.987	2.511	–49.365	27.860	–0.111	–54.360
Roundabout	16.863	6.004	–43.463	13.554	8.069	–40.585
Stop sign	24.148	–35.585	14.733	24.694	–36.275	14.788
Yield sign	12.034	–7.910	–9.772	12.339	–7.676	–10.800
Medium speed zone	–8.934	5.619	7.697	–9.555	5.667	8.846
High speed zone	–39.029	4.574	68.420	–36.141	0.485	69.923
Environmental factors						
Winter	–7.986	4.794	7.279	–9.299	5.214	9.133
Evening peak	6.299	7.697	–25.740	6.238	8.905	–27.715
Late night	–4.434	–10.787	27.471	–3.696	–9.685	24.095
Rain/snow	–7.323	4.335	6.780	–8.852	4.861	8.873
High wind	–21.818	11.163	23.256	–21.333	10.121	24.155
Dusk/dawn	–15.040	8.381	14.835	–13.939	7.298	14.592
Dark-lighted	–2.359	–4.079	11.724	7.797	–17.138	14.563
Situation factors						
One passenger	–1.491	10.792	–15.884	4.042	7.282	–20.590
Two passenger	10.557	9.688	–37.540	24.627	–8.417	–33.573
Driver ejected out of the vehicle	–1.010	–32.988	16.292	0.000	–47.556	82.777

The following observations can be made based on the elasticity effects of the variables presented in Table 4. *First*, the most significant variables in terms of increase in serious/fatal injury (from both models) for drivers are driver age above 65, driver ejection, not wearing seat belts, and collision in high speed zone. In terms of serious/fatal injury reduction, the important factors are presence of pedestrian control, presence of roundabout, driving a panel van, unpaved road condition and presence of passengers. *Second*, the segmentation variables exhibit significant influence on injury severity profile with struck object collisions having the most significant contribution to increase in serious/fatal injury. *Third*, there are substantial differences in the elasticity effects of LSOL and LSGOL models. For instance, the LSOL model predicts an increase in minor injury for young driver while LSGOL model predicts a decrease in the same category. Such differences can also be observed for other variables – collision in dark-lighted condition, in the presence of one passenger and for pedestrian control.

7. Validation analysis

We also carried out a validation experiment in order to ensure that the statistical results obtained above are not a manifestation of over fitting to data. 100 different data samples of about 2500 records were generated randomly from the hold out validation sample consisting of 37,680 records to test the predictive performance of the estimated models. We evaluate both the aggregate and disaggregate measure of predicted fit by using these 100 different validation samples and present the average measure from the comparison, and also the confidence interval (C.I.), of the fit measures at 90% level.

At the disaggregate level we compute predictive log-likelihood (computed by calculating the log-likelihood for the predicted probabilities of the sample), predictive adjusted likelihood ratio index and probability of correct prediction (defined as an indicator if the observed outcome has the highest predicted probability). The results for disaggregate measures are presented in Table 5 (top row panel). At the aggregate level, root mean square error (RMSE) and mean absolute percentage error (MAPE) are computed by comparing the predicted and actual (observed) shares of injuries in each injury severity level for each set of full validation sample. The results for aggregate measure computation are also presented in Table 5 (bottom row panel). The comparison of LSOL and LSGOL model at the disaggregate level is far from conclusive with a slight edge to the LSGOL model. This is not surprising because the difference in BIC values for the two models is relatively small. The LSGOL model represents a clearly superior performance compared to that of the LSOL model at aggregate level.

8. Conclusions

This paper formulates and estimates an econometric model for examining driver injury severity that accommodates systematic heterogeneity based on crash characteristics and relaxes the constant threshold assumption of traditional ordered logit model. The model is referred to as the latent segmentation based generalized ordered logit model. In traffic crash reporting, injury severity is typically characterized as an ordered variable resulting in application of the ordered response model for identifying the impact of exogenous variables. However, ordered systems impose a uni-directional impact of exogenous variables on injury severity alternatives. On the contrary, the generalized ordered logit model relaxes the restriction by allowing for the estimation of individual level thresholds as a function of exogenous variables. The widely employed ordered outcome model also restricts the impact of exogenous variables to be same across the entire population – homogeneity assumption. An alternative

Table 5
Measures of fit in validation sample.

Summary statistic		LSOL		LSGOL	
		Predictions	C.I.	Predictions	C.I.
Disaggregate level (100 validation samples)					
Number of observations		2547.540	–	2547.540	–
Number of parameters		37.000	–	41.000	–
Log-likelihood at zero		−2798.759	−2807.853 , −2789.665	−2798.759	−2807.853 , −2789.665
Log-likelihood at sample shares		−2696.838	−2706.066 , −2687.609	−2696.838	−2706.066 , −2687.609
Predictive Log-likelihood		−2438.553	−2447.559 , −2429.546	−2433.249	−2442.324 , −2424.174
Predictive adjusted likelihood ratio index		0.082	0.081 , 0.083	0.083	0.082 , 0.084
Average probability of correct prediction		0.531	0.529 , 0.533	0.531	0.529 , 0.532
Injury categories/measures of fit	**Actual shares**	**LSOL**		**LSGOL**	
		Predictions	**C.I.**	**Predictions**	**C.I.**
Aggregate level (100 validation samples)					
No injury	42.522799	42.309	42.247 , 42.371	42.364	42.300 , 42.428
Non-incapacitating injury	36.518347	36.595	36.559 , 36.629	36.492	36.452 , 36.531
Incapacitating/fatal injury	20.958854	21.096	21.049 , 21.142	21.144	21.099 , 21.189
RMSE	–	0.821	0.753 , 0.888	0.817	0.753 , 0.882
MAPE	–	2.352	2.351 , 2.354	2.343	2.341 , 2.345

approach, referred to as latent segmentation approach, accommodates systematic heterogeneity by allocating the drivers probabilistically to different segments and by estimating segment specific models for each segment. The current research effort contributes to safety literature empirically by building on the GOL model – by formulating and estimating a latent segmentation based generalized ordered logit (LSGOL) model.

The empirical analysis was conducted using the Victoria crash database from Australia for the years 2006 through 2010. The model was estimated using a comprehensive set of exogenous variables – crash characteristics, driver characteristics, vehicle characteristics, roadway design attributes, environmental factors and situational factors. The empirical analysis involved the estimation of models using six different statistical frameworks: (1) OL, (2) GOL, (3) LSOL with two segments, (4) LSOL with three segments, (5) LSGOL with two segments and (6) LSGOL with three segments. The comparison exercise, based on information criterion metrics, highlighted the superiority of the LSGOL model with two segments on the estimation sample in terms of data fit compared to the other ordered outcome models. In the LSGOL approach, drivers were assigned probabilistically to two segments – high risk segment and low risk segment – based on a host of crash characteristics. In our empirical analysis, the crash characteristics that affected the allocation of drivers into segments include: collision object, trajectory of vehicle's motion and manner of collision. According to our results, the impact of exogenous variables in the low risk segment was different (for some variables) in magnitude as well as in sign from the impact of exogenous variables in the high risk segment.

In our research, to further understand the impact of various exogenous factors, elasticity effects were estimated for both the LSOL and LSGOL models for comparison purpose. The elasticity effects indicated that the most significant variables in terms of increase in serious/fatal injury (from both models) for drivers were driver age above 65, driver ejection, not wearing seat belts, and collision in high speed zone. In terms of serious/fatal injury reduction, the important factors were presence of pedestrian control, presence of roundabout, driving a panel van, unpaved road condition and presence of passengers. Further, the performance evaluation of these models on a validation sample revealed that the LSGOL model represents a clearly superior performance compared to that of the LSOL model at an aggregate level. But, the comparison of LSOL and LSGOL model at a disaggregate level was far from conclusive with a slight edge to the LSGOL model. In summary, the comparison exercise supports the hypothesis that LSGOL is a promising ordered response framework for accommodating population heterogeneity and for relaxing the fixed threshold assumption in the context of driver injury severity.

The study is not without limitations. In our research effort, we employed data from the Victorian crash database of Australia for the years 2006 through 2010. However, in our analysis, we have not considered differences across the five years. It will be an interesting exercise to model the impact of temporal effects on segment specific injury severity models.

Acknowledgments

The corresponding author would like to acknowledge financial support from Fonds de recherche du Québec – Nature et technologies (FQRNT) Team Grant Program for undertaking the research. The authors would like to acknowledge useful feedback from the editor and anonymous reviewers.

References

Abdel-Aty, M.A., Abdelwahab, H.T., 2003. Configuration analysis of two-vehicle rear-end crashes. Transportation Research Record Board 1840, 140–147.
Arnedt, J., Wilde, G., Munt, P., MacLean, A., 2001. How do prolonged wakefulness and alcohol compare in the decrements they produce on a simulated driving task? Accident Analysis and Prevention 33 (3), 337–344.
Australia Transport Council, 2011. National Road Safety Strategy, 2011–2020.
Awadzi, K.D., Classen, S., Hall, A., Duncan, R.P., Garvan, C.W., 2008. Predictors of injury among younger and older adults in fatal motor vehicle crashes. Accident Analysis and Prevention 40 (6), 1804–1810.
Aziz, H.M.A., Ukkusuri, V.S., Hasan, S., 2013. Exploring the determinants of pedestrian–vehicle crash severity in New York city. Accident Analysis and Prevention 50, 1298–1309.
Bédard, M., Guyatt, G.H., Stones, M.J., Hirdes, J.P., 2002. The independent contribution of driver, crash, and vehicle characteristics to driver fatalities. Accident Analysis and Prevention 34 (6), 717–727.
Bhat, C.R., 1997. Endogenous segmentation mode choice model with an application to intercity travel. Transportation Science. 31 (1), 34–48.
Castro, M., Paleti, R., Bhat, C.R., 2013. A spatial generalized ordered response model to examine highway crash injury severity. Accident Analysis and Prevention 52, 188–203.
Chen, F., Chen, S., 2011. Injury severities of truck drivers in single- and multi-vehicle accidents on rural highways. Accident Analysis and Prevention 43 (5), 1677–1688.
Chen, H., Cao, L., Logan, D.B., 2012. Analysis of risk factors affecting the severity of intersection crashes by logistic regression. Traffic Injury Prevention 13 (3), 300–307.
Chiou, Y.C., Hwang, C.C., Chang, C.C., Fu., C., 2013. Modeling two-vehicle crash severity by a bivariate generalized ordered probit approach. Accident Analysis and Prevention 51, 175–184.
Chipman, M.L., 2004. Side impact crashes - factors affecting incidence and severity: review of the literature. Traffic Injury Prevention 5 (1), 67–75.
Collins, L.M., Fidler, P.L., Wugalter, S.E., Long, J.D., 1993. Goodness-of-fit testing for latent class models. Multivarlate Behavioral Research 28 (3), 375–389.
Conroy, C., Tominaga, G.T., Erwin, S., Pacyna, S., Velky, T., Kennedy, F., Sise, M., Coimbra, R., 2008. The influence of vehicle damage on injury severity of drivers in head-on motor vehicle crashes. Accident Analysis and Prevention 40 (4), 1589–1594.
Das, A., Abdel-Aty, M.A., 2011. A combined frequency-severity approach for the analysis of rear-end crashes on urban arterials. Safety Science 49 (8–9), 1156–1163.
de Lapparent, M., 2008. Willingness to use safety belt and levels of injury in car accidents. Accident Analysis and Prevention 40 (3), 1023–1032.
Depaire, B., Wets, G., Vanhoof, K., 2008. Traffic accident segmentation by means of latent class clustering. Accident Analysis and Prevention 40 (4), 1257–1266.

Eluru, N., 2013. Evaluating alternate discrete choice frameworks for modeling ordinal discrete variables. Accident Analysis and Prevention 55 (0), 1–11.
Eluru, N., Bagheri, M., Miranda-Moreno, L.F., Fu, L., 2012. A latent class modeling approach for identifying vehicle driver injury severity factors at highway-railway crossings. Accident Analysis and Prevention 47, 119–127.
Eluru, N., Bhat, C.R., 2007. A joint econometric analysis of seat belt use and crash-related injury severity. Accident Analysis and Prevention 39 (5), 1037–1049.
Eluru, N., Bhat, C.R., Hensher, D.A., 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accident Analysis and Prevention 40 (3), 1033–1054.
Eluru, N., Paleti, R., Pendyala, R.M., Bhat, C.R., 2010. Modeling injury severity of multiple occupants of vehicles: copula-based multivariate approach. Transportation Research Record 2165, 1–11.
Evans, L., 2004. Traffic Safety. Science Serving Society, Bloomfield, Michigan.
Fredette, M., Mambu, L.S., Chouinard, A., Bellavance, F., 2008. Safety impacts due to the incompatibility of suvs, minivans, and pickup trucks in two-vehicle collisions. Accident Analysis and Prevention 40 (6), 1987–1995.
Gårder, P., 2006. Segment characteristics and severity of head-on crashes on two-lane rural highways in Maine. Accident Analysis and Prevention 38 (4), 652–661.
Gray, R., Regan, D., 2007. Glare susceptibility test results correlate with temporal safety margin when executing turns across approaching vehicles in simulated low-sun conditions. Ophthalmic Physiological Optics 27 (5), 440–450.
Greene, W.H., Hensher, D.A., 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. Transportation Research B 37 (8), 681–698.
Helai, H., Chin, H., Haque, M., 2008. Severity of driver injury and vehicle damage in traffic crashes at intersections: a Bayesian hierarchical analysis. Accident Analysis and Prevention 40 (1), 45–54.
Holdridge, J.M., Shankar, V.N., Ulfarsson, G.F., 2005. The crash severity impacts of fixed roadside objects. Journal of Safety Research 36 (2), 139–147.
Hu, W., Donnell., E., 2010. Median barrier crash severity: some new insights. Accident Analysis and Prevention 42 (6), 1697–1704.
Islam, S., Mannering, F., 2006. Driver aging and its effect on male and female single-vehicle accident injuries: some additional evidence. Journal of Safety Research 37 (3), 267–276.
Jin, Y., Wang, X., Chen, X., 2010. Right-angle crash injury severity analysis using ordered probability models. In: Proceedings of 2010 International Conference on Intelligent Computation Technology and Automation (ICICTA 2010), pp. 206–209.
Jung, S., Qin, X., Noyce, D.A., 2011. Injury severity of multivehicle crash in rainy weather. Journal of Transportation Engineering 138 (1), 50–59.
Jurado-Piña, R., Pardillo-Mayora, J.M., Jiménez, R., 2010. Methodology to analyze Sun glare related safety problems at highway tunnel exits. Journal of Transportation Engineering 136 (6), 545–553.
Khattak, A.J., 2001. Injury severity in multivehicle rear-end crashes. Transportation Research Record 1746, 59–68.
Khattak, A., Knapp, K., 2001. Interstate highway crash injuries during winter snow and nonsnow events. Transportation Research Record 1746, 30–36.
Khorashadi, A., Niemeier, D., Shankar, V., Mannering, F., 2005. Differences in rural and urban driver-injury severities in accidents involving large-trucks: an exploratory analysis. Accident Analysis and Prevention 37 (5), 910–921.
Kim, J.-K., Ulfarsson, G.F., Kim, S., Shankar, V.N., 2013. Driver-injury severity in single-vehicle crashes in california: a mixed logit analysis of heterogeneity due to age and gender. Accident Analysis and Prevention 50, 1073–1081.
Kockelman, K.M., Kweon, Y.J., 2002. Driver injury severity: an application of ordered probit models. Accident Analysis and Prevention 34 (3), 313–321.
Krafft, M., Kullgren, A., Tingvall, C., Boström, O., Fredriksson, R., 2000. How crash severity in rear impacts influences short- and long-term consequences to the neck. Accident Analysis and Prevention 32 (2), 187–195.
Lécuyer, J.F., Chouinard, A., 2006. Study on the effect of vehicle age and the importation of vehicles 15 years and older on the number of fatalities, serious injuries and collisions in Canada. In: Proceedings of the Canadian Multidisciplinary Road Safety Conference XVI.
Mackay, G., Hill, J., Parkin, S., Munns, J., 1993. Restrained occupants on the nonstruck side in lateral collisions. Accident Analysis and Prevention 25 (2), 147–152.
McGinnis, R.G., Davis, M.J., Hathaway, E.A., 2001. Longitudinal analysis of fatal run-off-road crashes, 1975–1997. Transportation Research Record 1746, 47–58.
McLellan, B., Rizoli, S., Brenneman, F., Boulanger, B., Sharkey, P., Szalai, J., 1996. Injury pattern and severity in lateral motor vehicle collisions: a Canadian experience. Journal of Trauma—Injury, Infectection and Critical Care 41 (4), 708–713.
Ministry of Infrastructure and Transport, 2010. Road deaths in Australia 1925–2008, Department of Infrastructure, Transport, Regional Development and Local Government. Australia Government.
Mohamed, M.G., Saunier, N., Miranda-Moreno, L.F., Ukkusuri, S.V., 2013. A clustering regression approach: a comprehensive injury severity analysis of pedestrian-vehicle crashes in New York, US and Montreal, Canada. Safety Science 54, 27–37.
Mooradian, J., Ivan, J.N., Ravishanker, N., Hu, S., 2013. Analysis of driver and passenger crash injury severity using partial proportional odds models. Accident Analysis and Prevention 58, 53–58.
Morgan, A., Mannering, F.L., 2011. The effects of road-surface conditions, age, and gender on driver-injury severities. Accident Analysis and Prevention 43 (5), 1852–1863.
Nordhoff, L.S., 2005. Motor vehicle collision injuries: biomechanics, diagnosis, and management, Jones and Bartlett Learning.
Nylund, K.L., Asparouhov, T., Muthén, B.O., 2007. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. Structural Equation Modeling 14 (4), 535–569.
Obeng, K., 2008. Injury severity, vehicle safety features, and intersection crashes. Traffic Injury Prevention 9 (3), 268–276.
O'Donnell, C.J., Connor, D.H., 1996. Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice. Accident Analysis and Prevention 28 (6), 739–753.
Paleti, R., Eluru, N., Bhat, C.R., 2010. Examining the influence of aggressive driving behavior on driver injury severity in traffic crashes. Accident Analysis and Prevention 42 (6), 1839–1854.
Persaud, B.N., Retting, R.A., Garder, P.E., Lord, D., 2001. Safety effect of roundabout conversions in the united states: empirical Bayes observational before-after study. Transportation Research Record 1751, 1–8.
Plainis, S., Murray, I., Pallikaris, I., 2006. Road traffic casualties: understanding the night-time death toll. Injury Prevention 12 (2), 125–128.
Prentkovskis, O., Sokolovskij, E., Bartulis, V., 2010. Investigating traffic accidents: a collision of two motor vehicles. Transport 25 (2), 105–115. (+Ia+IIa).
Ray, M.H., 1998. Impact conditions in side-impact collisions with fixed roadside objects. Accident Analysis and Prevention 31 (1-2), 21–30.
Retting, R.A., Persaud, B.N., Garder, P.E., Lord, D., 2001. Crash and injury reduction following installation of roundabouts in the United States. American Journal of Public Health 91 (4), 628–631.
Retting, R.A., Weinstein, H.B., Solomon, M.G., 2003. Analysis of motor-vehicle crashes at stop signs in four U.S. cities. Journal of Safety Research 34 (5), 485–489.
Savolainen, P.T., Mannering, F.L., Lord, D., Quddus, M.A., 2011. The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. Accident Analysis and Prevention 43 (5), 1666–1676.
Schneider, W., Savolainen, P., Zimmerman, K., 2009. Driver injury severity resulting from single-vehicle crashes along horizontal curves on rural two-lane highways. Transportation Research Record 2102, 85–92.
Sivak, M., Schoettle, B., Rupp, J., 2010. Survival in fatal road crashes: body mass index, gender, and safety belt use. Traffic Injury Prevention 11 (1), 66–68.
Sobhani A., Eluru N., Faghih-Imani, A., 2013. A latent segmentation based multiple discrete continuous extreme value model. Transportation Research http://dx.doi.org/10.1016/j.trb.2013.07.009.
Sobhani, A., Young, W., Logan, D., Bahrololoom, S., 2011. A kinetic energy model of two-vehicle crash injury severity. Accident Analysis and Prevention 43 (3), 741–754.

Srinivasan, K.K., 2002. Injury severity analysis with variable and correlated thresholds: ordered mixed logit formulation. Transportation Research Record 1784, 132–142.
Tay, R., Rifaat, S.M., 2007. Factors contributing to the severity of intersection crashes. Journal of Advanced Transportation 41 (3), 245–265.
Ulfarsson, G.F., Mannering, F.L., 2004. Differences in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger car accidents. Accident Analysis and Prevention 36 (2), 135–147.
Wang, X., Kockelman, K.M., 2005. Use of heteroscedastic ordered logit model to study severity of occupant injury distinguishing effects of vehicle weight and type. Transportation Research Record 1908, 195–204.
WHO, 2013. Global Status Report on Road Safety. World Health OrganizationWorld Health Organization.
Williams, A.F., Shabanova, V.I., 2003. Responsibility of drivers, by age and gender, for motor-vehicle crash deaths. Journal of Safety Research 34 (5), 527–531.
Xie, Y., Zhang, Y., Liang, F., 2009. Crash injury severity analysis using bayesian ordered probit models. Journal of Transportation Engineering 135 (1), 18–25.
Xie, Y., Zhao, K., Huynh, N., 2012. Analysis of driver injury severity in rural single-vehicle crashes. Accident Analysis and Prevention 47, 36–44.
Xiong, Y., Mannering, F.L., 2013. The heterogeneous effects of guardian supervision on adolescent driver-injury severities: a finite-mixture random-parameters approach. Transportation Research B 49, 39–54.
Yan, X., Radwan, E., Abdel-Aty, M., 2005. Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model. Accident Analysis and Prevention 37 (6), 983–995.
Yasmin, S., Anowar, S., Tay, R., 2012. Effects of drivers' actions on severity of emergency vehicle collisions. Transportation Research Record 2318, 90–97.
Yasmin, S., Eluru, N., 2013. Evaluating alternate discrete outcome frameworks for modeling crash injury severity. Accident Analysis and Prevention 59, 506–521.
Yasmin S., Eluru, N., Ukkusuri, S.V., 2013. Alternative ordered response frameworks for examining pedestrian injury severity in New York City. Journal of Transportation Safety and Security http://dx.doi.org/10.1080/19439962.2013.839590.
Yau, K.K.W., 2004. Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong. Accident Analysis and Prevention 36 (3), 333–340.
Young, R.K., Liesman, J., 2007. Estimating the relationship between measured wind speed and overturning truck crashes using a binary logit model. Accident Analysis and Prevention 39 (3), 574–580.
Zhang, C., Ivan, J.N., 2005. Effects of geometric characteristics on head-on crash incidence on two-lane roads in Connecticut. Transportation Research Record 1908, 159–164.
Zhang, J., Lindsay, J., Clarke, K., Robbins, G., Mao, Y., 2000. Factors affecting the severity of motor vehicle traffic crashes involving elderly drivers in Ontario. Accident Analysis and Prevention 32 (1), 117–125.
Zuxuan, D., Ivan, J.N., Gårder, P., 2006. Analysis of factors affecting the severity of head-on crashes two-lane rural highways in Connecticut. Transportation Research Record 1953, 137–146.