Accident Analysis and Prevention 47 (2012) 119–127

Contents lists available at SciVerse ScienceDirect

Accident Analysis and Prevention

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

A latent class modeling approach for identifying vehicle driver injury severity factors at highway-railway crossings

Naveen Eluru[a,*], Morteza Bagheri[b,1], Luis F. Miranda-Moreno[a,2], Liping Fu[c,3]

[a] Department of Civil Engineering and Applied Mechanics, McGill University, Canada
[b] School of Railway Engineering, Iran University of Science and Technology, Iran
[c] Department of Civil & Environmental Engineering, University of Waterloo, Canada

ARTICLE INFO

Article history:
Received 16 January 2011
Received in revised form 17 January 2012
Accepted 18 January 2012

Keywords:
Collision consequence
Highway-railway crossing
Occupant injury severity
Latent segmentation based ordered response model

ABSTRACT

In this paper, we aim to identify the different factors that influence injury severity of highway vehicle occupants, in particular drivers, involved in a vehicle-train collision at highway-railway grade crossings. The commonly used approach to modeling vehicle occupant injury severity is the traditional ordered response model that assumes the effect of various exogenous factors on injury severity to be constant across all accidents. The current research effort attempts to address this issue by applying an innovative latent segmentation based ordered logit model to evaluate the effects of various factors on the injury severity of vehicle drivers. In this model, the highway-railway crossings are assigned probabilistically to different segments based on their attributes with a separate injury severity component for each segment. The validity and strength of the formulated collision consequence model is tested using the US Federal Railroad Administration database which includes inventory data of all the railroad crossings in the US and collision data at these highway railway crossings from 1997 to 2006. The model estimation results clearly highlight the existence of risk segmentation within the affected grade crossing population by the presence of active warning devices, presence of permanent structure near the crossing and roadway type. The key factors influencing injury severity include driver age, time of the accident, presence of snow and/or rain, vehicle role in the crash and motorist action prior to the crash.

1. Introduction

There are more than 250,000 highway-railway crossings in the US catering to a broad spectrum of road and train traffic. In spite of the success of the recent safety initiatives that have substantially reduced the number of highway-railway crossing collisions, the US Federal Railroad Association (FRA) still recorded more than 30,000 collisions during the ten year period from 1997 to 2006. Traffic crashes at highway-railway crossings are often catastrophic and it is of utmost importance for transportation agencies and other stakeholders to identify collision contributing factors and counter measures to reduce traffic collisions and the resulting consequences.

Collisions occurring at these facilities could result in serious consequences including severe injuries to roadway vehicle occupants and train passengers, and substantial property damage to vehicles and trains (e.g. derailment), and delay in railway and highway traffic (Raub, 2009). In collisions involving freight trains carrying hazardous materials the consequences can be further exasperated due to release of hazardous materials into the environment. A number of earlier research studies have focused on identifying the contributing factors that affect the occurrence of collisions at highway-railway crossings (see studies such as Saccomanno et al., 2007; Washington and Oh, 2006; Saccomanno and Lai, 2005). These studies employ different techniques such as factor/cluster analysis, negative binomial regression models, and Bayesian methods. For a literature review, the reader is referred to Lord and Mannering (2010). However, collision frequency is only one element of collision risk at highway-railway crossings. The risk associated with a crossing is typically defined as a function of collision frequency and collision consequence – *total risk* (Miranda-Moreno et al., 2009). To consider just frequency as a measure of risk would ignore crossings with a low expected collision frequency, but high potential for severe consequences. Therefore, it is essential that research efforts in safety literature examine the factors associated with the injury severity (consequence)

* Corresponding author. Tel.: +1 514 398 6823; fax: +1 514 398 7361.
E-mail addresses: naveen.eluru@mcgill.ca (N. Eluru), morteza_bagheri@iust.ac.ir (M. Bagheri), luis.miranda-moreno@mcgill.ca (L.F. Miranda-Moreno), lfu@uwaterloo.ca (L. Fu).
[1] Tel.: +98 21 77240541x3525; fax: +98 21 77451568.
[2] Tel.: +1 514 398 6589; fax: +1 514 398 7361.
[3] Tel.: +1 519 888 4567x33984; fax: +1 519 888 4349.

0001-4575/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2012.01.027

sustained in collisions at highway-railway crossings. While many previous studies have focused on predicting the frequency of collisions, there is a lack of substantive research that particularly examines the consequence of collisions at highway-railway crossings.

The current research contributes to our understanding of highway-railway crossing collision related driver injury severity along two dimensions: (1) empirical analysis and (2) modeling framework. In the proposed study we consider the influence of an exhaustive list of exogenous variables on driver injury severity. The study also develops a latent segmentation based ordered logit model to undertake the comprehensive empirical analysis. The rest of the paper is organized as follows: Section 2 presents a discussion of the earlier literature and positions the current research effort. Section 3 provides the details of the econometric model. Section 4 briefly describes the data preparation effort and presents sample characteristics. In Section 5, the intuitive implications of the latent segmentation model, a detailed discussion of the estimation results and elasticity effects of the best fit model are presented. Section 6 concludes the paper.

2. Literature review and current study

In the current section, we discuss safety literature examining collision consequence at highway-railway crossings and highlight the empirical and methodological contributions of our research effort. Specifically, the findings and limitations from earlier studies on highway-railway collision consequence are presented. Subsequently we discuss the prevalent modeling technique to examine driver injury severity in road safety literature and identify how the proposed framework improves the analysis approach.

2.1. Collision consequence literature

The current review focuses solely on studies related to driver injury severity instead of collision frequency. Earlier research on safety at highway-railway crossings has focused predominantly on the influence of grade crossing, geometric and traffic attributes on collision frequency.

There is very little research focusing on collision consequence of train and motor vehicles at highway-railway crossings. Raub (2009) undertakes a descriptive analysis of FRA data from 1998 to 2007. In the study, the author examines the collision consequence through a univariate analysis using gender, age, and type of crash (classified as vehicle struck the train or vice versa). Miranda-Moreno et al. (2009) developed a systematic Bayesian framework to estimate the total risk of a particular highway-railway crossing by considering the total risk as the product of accident frequency and expected consequence. Within this framework, a multinomial logit model was employed to study injury severity of vehicle occupants involved in highway-railway crossing collisions. The proposed approach represents a significant enhancement to earlier research on highway-railway crossing research efforts. However, only train speed and posted speed limit variables were considered in their analysis neglecting many other potential exogenous variables. Hu et al. (2010) represents one of the first research efforts in modeling accident injury severity at highway-railway crossings. The authors formulate a generalized logit model with stepwise variable selection to predict the level of injury severity. The model is estimated using data from traffic accidents at 592 highway railway crossings in Taiwan. From their analysis the authors identify the number of daily trains, number of daily trucks, highway separation, obstacle detection device, and approaching crossing marks as important determinants of injury severity. However, driver demographics are not employed in their analysis of injury severity.

Overall, it is surprising that there are only three studies that have examined vehicle operator injury severity as a consequence of highway-railway crossing collisions. Even those research efforts that examined highway-railway collision consequence have only employed a limited variable database for analysis. In our research study, we examine the influence of a host of exogenous factors on injury severity of vehicle drivers involved in collisions at highway-railway crossings. Specifically, the focus is on examining the influence of two sets of attributes: (a) accident attributes and (b) highway-railway crossing attributes. *Accident attributes* considered include: (1) driver demographics (including gender, age, vehicle occupancy), (2) characteristics of the vehicle involved in the collision (vehicle type), (3) environmental factors (weather, lighting conditions, time of day, etc.), and (4) crash characteristics (role of vehicle in crash, etc.). *Crossing attributes* considered include: (1) crossing characteristics (Annual traffic on the highway, railway traffic, etc.), and (2) crossing safety equipment (presence of gates, traffic signals, watchmen, etc.).

2.2. Modeling driver injury severity

In road safety literature, a host of studies have examined driver injury severity (in highway crashes) employing the traditional ordered response mechanism to take into account the inherent ordering of the reported driver injury severity (see for example O'Donnell and Connor, 1996; Eluru and Bhat, 2007). These approaches can be easily extended for studying vehicle driver injury severity for highway-railway crossing collisions. The traditional ordered response models may provide inaccurate estimates of the effect of exogenous variables on vehicle driver injury severity because they restrict the impact of accident related exogenous variables to be identical for all highway-railway crossings (Eluru et al., 2008). In reality, the influence of accident attributes on collision severity might vary across the highway-railway crossing population.

To illustrate this, consider the impact of two highway-railway crossing collisions involving male drivers that occurred at two different highway-railway crossings (C1 and C2) with identical accident attributes (i.e. driver demographics, vehicle characteristics environmental factors and crash characteristics are identical). The only highway-railway crossing attribute different between crossing C1 and crossing C2 is the presence of a stop sign. At crossing C1 a stop sign is installed while it is absent at crossing C2. Let us also assume that the drivers are law abiding individuals for the sake of discussion. In the first collision at C1, the driver stopped at the stop sign. So, he must be traveling at a lower speed at the time of collision thus allowing the driver additional time to maneuver the vehicle prior to the collision. This maneuverability will allow the driver to reduce the impact of the collision marginally. In this case, the higher physiological strength of the male driver (compared to a female driver) might result in a less severe injury for male drivers. On the other hand, if the male driver is involved in a collision at crossing C2, the driver would not have stopped and possibly would be traveling at a higher speed at the time of the collision thus reducing the advantage of the additional physiological strength (compared to a female driver) having any effect on injury severity. The additional physiological strength of the male driver can reduce injury severity only in less severe crashes. This differential influence on injury severity will not be apparent for a female driver. This is an example of the "male" attribute exhibiting differential sensitivity based on the crossing attribute – presence of a stop sign. It is plausible that the effect of all accident attributes is moderated by crossing attributes in a similar fashion. If the modeling methodology does not allow for such flexible impacts, the true impact will be lost in the model estimation. Hence, evaluating injury severity employing a traditional ordered response model might possibly lead to incorrect coefficient estimates.

A common approach to address this problem is to relax the homogeneity assumption of the ordered response model by categorizing highway-railway crossings into different segments based on crossing attributes and subsequently model the effect of accident attributes within each segment separately. The challenge, however, is in determining the segmentation. This issue has traditionally been addressed by partitioning the highway-railway crossings into mutually exclusive segments based on key characteristics (such as daily through volume, Average Annual daily traffic (AADT), safety equipment available at the crossing, visibility at the crossing). This approach is appropriate when the focus is on examining segmentation based on one or two variables. However, in reality, we could segment the crossings based on a large set of exogenous variables. For example if we have 4 variables with two attribute levels each, we require 16 crossing segments with one ordered response model per segment. Not only is this approach unwieldy, but also reduces the sample size in each segment substantially resulting in inefficient model estimation.

In this paper, we apply a new modeling framework called latent segmentation approach to segment crossings probabilistically based on a host of crossing attributes (Bhat, 1997). For instance, a crossing with adequate safety equipment available could be classified as "low risk" with a very high probability and "high risk" with a very small probability. On the other hand, crossings that are devoid of safety equipment could be classified as "high risk" with a high probability and "low risk" with low probability. Within each of these segments, the vehicle driver injury severity is determined based on an ordered response model that considers all accident attributes. The newly formulated model will allow us to partition highway-railway crossings into segments based on their attributes and estimate the influence of accident attributes on injury severity separately within each segment. The latent segmentation model developed will enable transportation safety analysts to identify the crossing attributes that contribute to or mitigate the likelihood of severe injuries for vehicle drivers. A conventional ordered response model due to its inflexibility might not be as effective in accurately identifying these factors. Further, the restrictive modeling frameworks employed for the analysis could potentially lead to incorrect and biased model estimation results.

To be sure, the approach proposed in this paper has been employed by researchers in Economics (Greene et al., 2008) and Bio-Statistics (Desantis et al., 2008) recently. In the field of transportation safety a similar approach is attempted by Depaire et al. (2008) and Park and Lord (2009). Depaire et al. (2008) developed a sequential framework for classifying traffic accidents. The authors classify the various accidents into segments and subsequently estimate separate models for each segment. The approach is better than segmentation based on each exogenous variable, but still might result in very small samples for some segments. In their study they generated 7 mutually exclusive segments with sample sizes varying from 3800 to 142. Thus, even this approach is affected by small sample issues. Park and Lord (2009) developed a finite mixture based approach to modeling traffic collision counts. The finite mixture approach is similar to the proposed latent segmentation approach except that the mixture probability is not expressed as a function of exogenous variables i.e. it is not evident how the population is segmented. Consequently, it is not as useful. The proposed latent segmentation approach addresses two concerns: (1) ensures that the parameters are estimated employing the full sample for each segment while employing all data points for model estimation and (2) provides valuable insights on how the exogenous variables affect segmentation. The proposed approach is the first implementation of latent segmentation for an ordered response model in the transportation safety literature.

To summarize, our current study contributes to highway-railway crossing collision consequence literature in two ways: (1) examines the influence of a host of exogenous factors on injury severity of vehicle drivers involved in collisions at highway-railway crossings and (2) formulates and estimates a latent segmentation based ordered logit model that allows us to determine the influence of exogenous variables on driver injury severity accurately.

3. Model framework

The modeling of vehicle driver injury severity is achieved using a latent segmentation based ordered response model. Let us consider S homogenous segments of highway-railway crossings (S is to be determined). The pattern of injury severity within the segment remains identical. However, there are intrinsic differences in the pattern of injury severity across different segments i.e. we have a distinct ordered response model for each segment (1,2, . . ., S).

Within each segment, we formulate the ordered response model in its traditional form. Let q (q = 1, 2, . . ., Q) be an index to represent drivers and let k (k = 1, 2, 3, . . ., K) be an index to represent injury severity. The index k, for example, may take values of "no injury" (k = 1), "injury" (k = 2), and "fatal injury" (k = 3), as in the empirical analysis in the current paper. Eq. (1) represents the latent propensity y_{qs}^* associated with the injury severity sustained by driver q in the accident if s/he were to belong to segment s

$$y_{qs}^* = \alpha_s' x_q + \varepsilon_{qs}, \quad y_q = k \quad \text{if} \quad \psi_{s_{k-1}} < y_q^* < \psi_{s_k} \tag{1}$$

y_{qs}^* is mapped to the actual injury severity level y_q by the ψ thresholds ($\psi_{s_0} = -\infty$ and $\psi_{s_K} = \infty$) in the usual ordered-response fashion. x_q is an $(L \times 1)$ column vector of attributes (not including a constant) that influences the propensity associated with injury severity. α is a corresponding $(L \times 1)$-column vector of coefficients and ε_q is an idiosyncratic random error term assumed to be identically and independently standard logistic distributed across individuals q.

The probability that driver q sustained injury severity k is given by:

$$P_q(k)|s = \Lambda(\psi_{s_k} - \alpha_s' x_q) - \Lambda(\psi_{s_{k-1}} - \alpha_s' x_q) \tag{2}$$

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

Now we need to determine how to assign the crossings that the drivers had accidents probabilistically to the segments. The random utility based multinomial logit structure is employed for the segmentation model. The utility for assigning a driver q's crossing to segment s is defined as:

$$U_{qs}^* = \beta_s' y_q + \xi_{qs} \tag{3}$$

y_q is an $(M \times 1)$ column vector of attributes (not including a constant) that influences the propensity of belonging to segment s. β is a corresponding $(M \times 1)$-column vector of coefficients and ξ_{qs} is an idiosyncratic random error term assumed to be identically and independently Type 1 Extreme Value distributed across individuals q and segment s. Then the probability that driver q's crossing belongs to segment s is given as:

$$P_{qs} = \frac{\exp(\beta_s' y_q)}{\sum_s \exp(\beta_s' y_q)} \tag{4}$$

Based on the above discussion, the unconditional probability of individual sustaining injury severity k is given as:

$$P_q(k) = \sum_{s=1}^{S} (P_q(k)|s)(P_{qs}) \tag{5}$$

The parameters to be estimated in the model are β_s and α_s for each s and the number of segments S. The log-likelihood function for the entire dataset is provided below:

$$L = \sum_{q=1}^{Q} \log(P_q(k)) \tag{6}$$

The model estimation approach begins with a model considering two segments. The final number of segments is determined by adding one segment at a time until further addition does not enhance intuitive interpretation and data fit. It is important to note that the estimation of latent class 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. For our analysis, the log-likelihood function and its corresponding gradient function were coded in Gauss Matrix programming language. The coding of the gradient function ensures we reduce the instability associated with the estimation process.

4. Data

The Federal Railroad Administration (FRA) crossing database provides information on the type, causes, consequences, and mitigating circumstances of train collisions experienced annually nation-wide in the US for the period 1975–2010. These data are readily available for downloading from the FRA, Office of Safety Analysis Web Site (http://safetydata.fra.dot.gov/OfficeofSafety/). The US FRA website contains two databases related to HRC: (1) collision records (called "Highway-Rail Grade Crossing Accident/Incident Form F 6180.57") and (2) inventory database. In this analysis records for the 10-year period from 1997 to 2006 were employed. The collision database contains information such as driver demographics, vehicle characteristics, the driver actions during collision, and crossing safety infrastructure deployed; the inventory database contains detailed information on railway traffic flow, list of crossing safety infrastructure deployed, roadway type classification, highway Annual Average Daily Traffic (AADT), presence and type of advance warning signs etc. corresponding to all the crossings in the U.S. The data sets contain a unique identifier to merge the crossing infrastructure information with the actual collision record. The collision database was merged with appropriate crossing information using this unique identifier.

The raw database consists of about 30,000 records. In this research, the analysis is confined to collisions occurring at public grade crossings on the main railway line, excluding those occurring at yards, sidings and industrial locations. Furthermore, we focus on the injury severity of the passenger motor vehicle drivers only; that is, collisions involving commercial vehicles were removed. The data assembly process involved removing records with missing and inconsistent information on variables such as driver injury, gender and age. The final sample compiled, after checking thoroughly for consistency, contains 14,532 observations. The injury severity of each individual involved in a crash is coded on a three-point ordinal scale: (1) no injury, (2) injury, and (3) fatal injury. The driver injury severity distribution in the final data sample is as follows: no injury (62.0%), injury (27.6%) and fatality (10.4%).

Table 1 offers a summary of the characteristics of the sample used in this empirical study. From the descriptive analysis, we can observe that the majority of the drivers are male (66.4%), under the age of 40 years (62.7%), and are primarily driving a sedan vehicle (72.7%). Further, a large portion of collisions occurs during the 3 PM–7 PM time period. The majority of collisions occur during fair weather (68.2%) and temperature (50.1%) conditions. With regard to presence of safety equipment at highway-railway crossings, cross bucks are most commonly employed safety device (69%). Other commonly employed safety equipment includes standard flashing lights, gates, and audible signals. Only a very small percentage of highway-railway crossings where collisions have occurred do not have any safety equipment (0.4%).

Table 1
Crash database sample statistics.

Driver gender	
Male	66.4
Female	33.6
Driver age	
≤ 25 years	32.2
25–40 years	30.5
41–64 years	25.5
41–64 years	11.8
Driver vehicle type	
Sedan	72.7
Minivan	21.8
Pickup	5.7
Collision time period of day	
12 AM–6 AM	14.0
6 AM–9 AM	11.9
9 AM–12 PM	15.6
12 PM–3 PM	16.8
3 PM–7 PM	23.6
7 PM–12 AM	18.1
Temperature conditions	
≤32F	13.0
32F–60F	36.9
>60F	50.1
Weather conditions	
Clear	68.2
Cloudy	20.0
Rain	7.1
Fog	1.9
Sleet	0.3
Snow	2.5
Type of warning device present	
Gates	31.1
Cantilever fls	15.3
Standard fls	40.1
Wigwags	1.6
Highway traffic signals	3.8
Audible signals	30.9
Cross bucks	69.0
Stop signs	13.8
Watchman	0.1
Flagged by crew	0.9
Other	13.8
None	0.4
Sample size	14532

5. Model estimation results

This section presents the model estimation results in detail. In Section 5.1 the overall model specification considerations and model performance in terms of goodness-of-fit are discussed. An intuitive discussion of the best fit latent segmentation model is provided in Section 5.2. Section 5.3 presents a detailed discussion of the impact of exogenous factors on latent segmentation and injury severity components. The magnitude of the impact of exogenous factors on injury severity is examined through elasticity effects in Section 5.4.

5.1. Variables considered

In our research, we examine latent segmentation based on the crossing attributes while the injury severity component models are estimated only using the accident attributes. The *crossing characteristics* considered include the highway roadway classification (interstate, arterial, major collector, minor collector), train traffic (daily, nightly), AADT, and maximum train speed. The *crossing safety equipment* considered includes presence of safety devices (such as gates, cross bucks, wigwags, highway traffic signals and flashing lights) and pavement markings. The *driver demographics* considered in the analysis include age, gender and vehicle occupancy. The *vehicle characteristics* available include the vehicle type of the car (sedan, pickup and van), number of locomotives on the train, number of cars on the train, direction of travel for the vehicle and the train (North, East, West, South). *Environmental factors* included in the model are time of day, temperature, weather conditions (clear, cloudy, rain, snow and/or fog) and visibility. The *crash characteristics* examined are role of the vehicle in the crash (defined as vehicle struck the train or vice versa), motorist action at the event of a crash, estimated train speed, the railway equipment involved in the crash (such as train unit pulling/pushing, train standing, car/s standing or moving, light locomotive/s standing or moving) and train car position.

The final specification was based on a systematic process of removing statistically insignificant variables and combining variables when their effects were not significantly different. The specification process was also guided by prior research, intuitiveness and parsimony considerations. We should also note here that, for the continuous variables in the data (such as age and time of day), we tested alternative functional forms including linear and spline (or piece-wise linear), and dummy variables for different ranges.

5.2. Model specification and overall performance

In this research effort, we considered three different model specifications including: (1) traditional ordered logit (OL) model, (2) latent segmentation based ordered logit model with two segments (LSOL II) and (3) latent segmentation based ordered logit model with three segments (LSOL III). Prior to discussing the model results we compare the performance of the OL, LSOL II and LSOL III models. These models are not nested within each other. Hence, we employ two goodness of fit measures that are suited to comparing non-nested models: (1) Bayesian Information Criterion (BIC)[4] and (2) Ben-Akiva and Lerman's adjusted likelihood ratio (BL) test.

The BIC for a given empirical model is equal to $-2\ln(L)+K\ln(Q)$, where $\ln(L)$ 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 value is the preferred model. The BIC values for the final specifications of the OL, LSOLII and LSOL III models are 22,964, 22,948 and 23,013 respectively.

The BL test statistic (Ben-Akiva and Lerman, 1985) is computed as: $\tau = \Phi\{-[-2(\bar{\rho}_2^2 - \bar{\rho}_1^2)L(C) + \sqrt{(M_2 - M_1)}\}$ where $\bar{\rho}^2$ represents the McFadden's adjusted rho-square value for the model. It is defined as $\bar{\rho}_2^2 = 1 - ((L_i(\beta) - M_i)/L(C))$ where $L_i(\beta)$ represents log-likelihood at convergence for the ith model, $L(C)$ represents log-likelihood at sample shares and M_i is the number of parameters in the model (Windmeijer, 1995). The $\Phi()$ represents the cumulative standard normal distribution function. The BL test compares two models by computing the probability (τ) that we could have obtained the higher $\bar{\rho}^2$ value for the "best" model even though this is not the case. The $\bar{\rho}^2$ values thus computed for the OL, LSOL II and LSOL III models are 0.119, 0.123and 0.119. The resulting τ value for the comparison of OL and LSOL II models and LSOL III and LSOL II is 0, 0 respectively, clearly indicating that LSOL II offers superior fit compared to OL and LSOL III models.

[4] The reader will note that we chose to employ BIC because it imposes substantially higher penalty on over-fitting with excess parameters compared to the penalty imposed by Alkaike Information Criterion (AIC). AIC is defined as $-2\ln(L)+2K$.

Table 2
LSOL II model estimates.

Segment	Driver population share	Injury severity within each segment		
		No injury	Severe injury	Fatal injury
1 (High risk)	0.21	0.16	0.59	0.25
2 (Low risk)	0.79	0.74	0.19	0.07

In our case study, the BIC and the BL test statistics clearly confirm that the LSOLII model offers substantially superior data fit compared to the OL and LSOL III models. The results clearly provide credence to our hypothesis that driver injury severity can be better examined through segmentation of highway-railway crossings. In the following presentation of empirical results we will confine ourselves to a discussion of LSOLII model results for the sake of brevity.

5.3. Intuitive interpretation of the LSOL II model

Prior to discussing the impact of various coefficients on segmentation and injury severity, it is important to discuss the overall segmentation characteristics. The model estimations can be used to generate information regarding: (1) percentage population share across the two segments and (2) overall injury severity shares within each segment. These estimates are provided in Table 2. Clearly, the likelihood of drivers being assigned to segment 2 is substantially higher than the likelihood of being assigned to segment 1. Further, the injury severity probabilities for drivers conditional on their belonging to a particular segment offer very distinct results indicating that the two segments exhibit distinct injury severity profiles. The drivers allocated to segment 1 are less likely to escape injury (only 16%) whereas the drivers assigned to segment 2 are less likely to sustain severe or fatal injuries (only 26%). In effect, it is clear that individuals involved in highway-railway crossing collisions that are assigned to segment 1 are likely to sustain severe injuries compared to those individuals involved in collisions assigned to segment 2. To facilitate the discussion from here on, we label segment 1 as the "high risk" segment and segment 2 as the "low risk" segment. These results clearly highlight the pitfalls of modeling using a traditional OR model where the variables are restricted to have the same injury profile for all individuals.

5.4. Estimation results

The LSOLII model estimation results, for the segmentation component and the injury severity components for low risk and high risk segments, are presented in Table 3.

5.4.1. Latent segmentation component

The latent segmentation component determines the probability that a driver is assigned to one of the two latent segments based on the highway crossing attributes. In our empirical analysis, the high risk segment is chosen to be the base and the coefficients presented in the table correspond to the propensity for being a part of low risk segment (see Eq. (3)). The results provide interesting insights on the likelihood of assigning individuals to different segments based on the exogenous variables.

The constant term clearly indicates a larger likelihood for drivers being part of segment two. Other crossing characteristics that affect the assignment of drivers include: daily total number of trains through the crossing, roadway classification, pavement markings, presence of obstacles that obscure the view for drivers and posted train speed at crossing.

Table 3
LSOLII model estimates of vehicle driver injury severity.

Variables considered	Segment 1 (high risk)		Segment 2 (low risk)	
	Estimate	*t*-Stats	Estimate	*t*-Stats
Highway railway crossing segmentation component				
Constant	–	–	0.9398	4.828
Crossing characteristics				
Total no. of trains through the crossing	–	–	0.0044	1.556
Roadway classification (base is rural and urban interstate)				
Rural local highway	–	–	0.2954	2.559
Rural minor collector	–	–	0.3413	2.008
Urban minor arterial	–	–	0.2839	1.851
Urban collector	–	–	0.4945	2.928
Urban local highway	–	–	0.4017	3.008
Pavement markings				
Stop sign	–	–	0.6036	2.285
Obstacles to road drivers near the crossing				
Permanent structure	–	–	−0.6882	−2.205
Posted train speed for the crossing				
Maximum	–	–	−0.0050	−1.790
Minimum	–	–	0.0093	3.967
Crossing safety equipment				
Type of warning device present (base is other)				
Cantilever flashing light signals	–	–	−0.1898	−1.561
Stop sign	–	–	0.1959	1.860
Crossbucks	–	–	−0.3291	−3.211
Gates	–	–	1.3012	7.818
Injury severity component				
Threshold parameters				
Threshold 1	2.4090	8.194	3.0104	0.173
Threshold 2	7.5172	10.990	4.7865	0.193
Driver demographics				
Male	–	–	−0.2165	−3.663
Age	0.0330	6.719	0.0121	7.758
Occupancy of the roadway vehicle involved in the crash	–	–	0.3610	11.481
Vehicle characteristics				
Vehicle type				
Van	−0.3121	−1.897	–	–
Environmental factors				
Time period of the day (remainder of the day is base)				
12 AM–6 AM	1.3132	4.826	−0.3753	−3.741
7 PM–12 AM	0.8694	4.055	−0.2048	−2.455
Temperature (≤32F is base)				
33F–60F	–	–	−0.2008	−2.145
>60F	–	–	−0.1038	−1.158
Weather conditions (clear weather is base)				
Rain			−0.1507	−1.248
Snow	−1.9862	−3.287	−0.2891	−1.290
Crash characteristics				
Role of vehicle in the crash (struck by the vehicle is base)				
Struck by the train	0.2243	1.169	0.3985	5.428
Motorist action at the event of a crash (base is other action)				
Drove around or through the gate	1.0973	1.987	0.4128	5.316
Motorist stopped on the crossing	−1.2178	−6.376	−1.6868	−13.348
Motorist did not stop	−0.8009	−2.207	−0.1829	−1.337
Estimated train speed	0.1301	10.016	0.0401	20.408
Log-likelihood at *constants*	−12896.8			
Log-likelihood at *convergence*	−11268.3			

An increase in the total number of trains passing through the highway-railway crossing increases the likelihood of assigning the driver to the "low risk" segment. When railway traffic at a crossing is high, roadway drivers are more alert to the possibility of encountering a train and are more likely to be attentive. Roadway type classification effects indicate that railway crossings with low class roadways (including rural local highway, rural minor collector, urban minor arterial, urban collector, and urban local highway) increase the likelihood of assigning the driver to the "low risk" segment. On these roads, the operating speeds and posted speed limits are expected to be lower than the reference category thus allowing drivers with more time to react in the event of a collision with the train. Thus, roadway facilities that are not highways increase the likelihood of being assigned to the "low risk" segment.

The presence of pavement markings for a stop sign increases the chance that the driver is assigned to the "low risk" segment. The result is expected because the presence of stop sign markings alerts the drivers to the approaching crossing thus ensuring that they reach the crossing at a lower speed compared to situation when the markings are missing. The presence of a permanent structure (that obscures the view for vehicle drivers) closer to the highway-railway crossing increases the likelihood that the driver is assigned to the "high risk" segment. The obstruction reduces visibility thus reducing reaction time for drivers increasing the likelihood of a potential high risk collision event.

The coefficient corresponding to the posted speed limit of the train at the highway-railway crossing indicates that as the "maximum" posted speed limit increases the likelihood of being assigned to the "high risk" segment is higher while at the same time increase in the "minimum" posted speed limit increases the likelihood of being assigned to the "low risk" segment. The former result is intuitive while the latter impact might appear counter-intuitive. A possible explanation for this is that in places where "minimum" speed limit is explicitly posted, the posted speed minimum speed limit variable acts as a proxy for crossing characteristics that promote safety. The variable influence could be explored further in future research.

The crossing safety equipment attribute – type of warning device present – has an important influence on assigning individuals to the different segments. At most crossings more than one type of warning device is present i.e. the presence of these devices is not mutually exclusive. Hence, it is necessary to view the influence of these variables as a total effect rather than just the impact of a single device. For instance, the presence of cantilever flashing signal lights and crossbucks increases the likelihood of being assigned to segment 1. This does not indicate that presence of these devices is actually harmful to drivers, but shows that their sole presence (without gates or stop sign) will result in less safe conditions. The presence of gates and stop sign on the other hand increases the safety of the crossing. This is expected because, the presence of gates is a stringent safety measure compared to the other alternatives discussed. In summary, the influence of type of warning devices is computed as the sum of coefficients of all warning devices that are present at the crossing. The computed value determines the likelihood of assigning the individual to different segments.

Overall, the "low risk" crossing segment is characterized by higher no. of trains, roadway classification of smaller roads, pavement markings for stop signs, absence of permanent structures obscuring the view, lower maximum posted train speed limits and presence of gates and stop signs.

5.4.2. *Injury severity component: segment one*

The ordered logit model corresponding to segment one (high risk segment) is described in this section. The interpretation of the coefficients follows the usual ordered response frameworks. The positive coefficients represent increased propensity to sustain severe injury while negative coefficients represent reduced propensity to sustain injury.

The only driver characteristic influencing injury severity for the "high risk" segment is age. The result indicates that the propensity to sustain a severe injury increases with age. This is expected because older individuals (compared to younger individuals) are more likely to be injured severely in the event of a crash. The examination of non-linear impact of age on injury severity did not result in significant parameter estimates. In terms of vehicle characteristics, drivers in vehicle type classified as Van are likely to sustain less severe injuries compared to drivers in other vehicle types.

The impact of environmental factors on injury severity is along expected lines. It is very interesting to note that "high risk" segment collisions occurring in the night time (7 PM–6 AM) are likely to result in severe injuries. This is potentially because vehicle drivers are less aware of the existence of a railway crossing during the night time due to lack of visibility. Further, lower traffic on the roadways at night encourages drivers to travel at higher speeds thus worsening the impact in the event of a collision.

The presence of snow at highway-railway crossings reduces the likelihood of severe injury. The result, though counter intuitive at first glance, is relatively easy to explain. The presence of snow causes the drivers to be cautious and drive slowly and the subsequent collisions occurring during snow result in less severe injuries. A similar result on the influence of snow on driver injury severity has been reported earlier in safety literature (see Eluru and Bhat, 2007).

Crash characteristics that significantly influence injury severity include: role of the vehicle and/or train, motorist action prior to collision, and estimated train speed. We find that vehicles struck by the train are more likely to involve individuals that sustain injury compared to the cases where the driver strikes the train. The result is quite intuitive because the train with its larger momentum is likely to cause more damage to a vehicle compared to the situation when a vehicle collides with the train.

Motorist action also has important implications for injury severity. The drivers that are involved in aggressive acts such as driving around or through the gate are likely to sustain severe injuries. Drivers that stopped on the crossing are likely to sustain the less severe injury. Drivers who have stopped on the crossing will have an opportunity to leave the car prior to the impact and thus reduce injury risk. Further, drivers that did not stop sustain injuries less severe than those involved in aggressive acts or drivers that participate in other motorist actions. Estimated train speed at the time of impact has a positive impact on injury severity. This is along expected lines. The faster the train is traveling the severe is the injury to the driver.

Thresholds in the ordered response model form the boundary point for the different injury severities. In our first segment, when the latent propensity of the individual is less than 2.4090 the driver sustains no injury. The driver sustains a serious injury when the propensity is between 2.4090 and 7.5172. The driver is fatally injured when the propensity value is greater than 7.5172.

5.4.3. *Injury severity component: segment two*

The injury severity propensity for the "low risk" segment provides variable impacts that are significantly different, in magnitude as well as sign (for a few variables), from the impacts offered by the exogenous variables in "high risk" segment. Further, we also notice that the number of variables that moderate the influence of injury severity is higher for the "low risk" segment. This again highlights the difference between the two segments. In the "high risk" segment, the injury severity is likely to be severe with very little chance to moderate injury severity through mitigating factors. On the other hand, drivers involved in crashes at the "low risk" highway-railway crossings benefit from the moderating effect of exogenous factors. So, if we can make changes to the highway-railway crossings to ensure the proportion of "high risk" segment becomes small, we can effectively reduce injury severities.

In the second segment, males are likely to sustain less severe injuries compared to females. This is expected because physiologically males are stronger. The finding in this study is similar to many findings from accident safety literature. As the individuals age increases, the likelihood of injury sustained also increases (similar to segment 1). The reader would notice that for crashes in "high risk" segment the role of gender is not important indicating that the additional physiological strength of male drivers' does not reduce injury severity.

An interesting variable that impacts injury severity, in segment 2, is the occupancy variable. The result indicates that as the occupancy of the roadway vehicle increases the likelihood of the driver sustaining a severe injury increases. It is plausible that in vehicles with multiple occupants the driver is distracted due to possible conversation and is not expecting a highway-railway crossing. Further, we have pointed out earlier that most of segment 2 highway-railway crossings are likely to be in smaller roadways where the likelihood of complacency for groups is likely to be higher (see Chang and Mannering, 1999; Paleti et al., 2010).

The influence of time period of the day has a strikingly different influence on the drivers from segment 2. The collisions occurring during the night time period are likely to result in less severe injuries in the "low risk" segment. The presence of more warning devices and reduced speeds on these facilities provide plausible explanation for this result.

The influence of environmental conditions on the crash severity for drivers from segment 2 is along expected lines. The presence of snow and rain reduces the likelihood of injury sustained for drivers (the reasoning is similar to segment 1). The injury severity of drivers is marginally influenced by temperature at the time of the crash. We find that crashes occurring at temperature greater than 32F are likely to result in less severe injuries compared to the conditions where temperature is lower than 32F.

The impact of vehicle role in the crash has a similar interpretation in segment 2. The drivers involved in crashes where they are struck by the train are likely to sustain severe injuries. The motorist action variable provides similar results as those offered by segment 1.

The impact of train speed is along expected lines with higher train speed resulting in severe injuries though the magnitude is much smaller in the second segment.

5.5. Elasticity effects

The exogenous variable coefficients presented in Table 3 do not directly provide the elasticities of the variables, that is, the magnitude of the impact of the variables on the probability of injury severity categories.

The elasticities can however be computed as the effective percentage change in aggregate shares in the entire sample due to changes to the exogenous variables. For ordinal and continuous variables the computation is straightforward. The value of the variable is increased by 1 and the resulting percentage change in probability is computed. To compute elasticity values for dummy exogenous variables, we consider two sub-samples: sub-sample with dummy variable value 0 and sub-sample with dummy variable value 1. For the first sub-sample, we change the variable value to 1 and compute the change in probability. For the second sub-sample, we change the variable value to 0 and compute the change in probability. To convert the change in the second sub-sample to the same direction as the change in the first sub-sample we reverse the signs of the value of the second sub-sample. Subsequently, the shifts from both sub-samples are added.

Table 4 provides the elasticity results by injury severity category for LSOL II model. The numbers in the table may be interpreted as the percentage change in the probability of injury category due to a change in the variable from 0 to 1. For instance, the LSOL II model in the table indicates that the probability that a male sustains a fatal injury is 10.1% lower than the probability that a female sustains a fatal injury, assuming other characteristics do not vary.

Several important observations can be made from the elasticity results presented in Table 4. *First*, crossing characteristics and crossing attributes exert significant influence on the injury severity profiles through their contributions to the segmentation component. These results clearly illustrate how the latent segmentation model assigns drivers to the two different segments with distinct injury severity profiles. *Second*, the factors that mitigate injury severity for the drivers involved in highway-railway collisions are number of trains in a day, smaller roadways, pavement markings for stop signs, minimum posted speed limit, crossing safety equipment such as stop signs and gates, male drivers, drivers vehicle type is van, temperature above 33F, presence of snow and/or rain, motorist actions including stopped on the crossing and motorist did not stop. *Third*, the factors that increase the propensity of injury severity for the drivers involved in highway-railway collisions are presence of permanent structures that obscures the view for road users, maximum posted speed limit, crashes during the time period 7 PM–6 AM, motorist being struck by train, driver involved in aggressive maneuvers, and estimated train speed. *Finally*, from the elasticity results it is evident that among dummy variables the most important determinants of injury severity are crossing safety equipment, roadway classification, pavement markings for stop signs, permanent structures obscuring the view for road users, presence of snow, and aggressive maneuvers such as "drive around or through the gate". Among the continuous variables age and estimated train speed are the important determinants.

Table 4
LSOL II model elasticity effects.

Variables	No injury	Severe injury	Fatal injury
Crossing characteristics			
Total no. of trains through the crossing	0.1	−0.1	−0.1
Roadway classification			
Rural local highway	4.2	−6.7	−7.1
Rural minor collector	4.6	−7.4	−7.5
Urban minor arterial	3.9	−6.3	−6.4
Urban collector	6.5	−10.4	−10.5
Urban local highway	5.5	−8.8	−8.8
Pavement markings			
Stop sign	7.6	−12.1	−12.3
Obstacles to road drivers near the crossing			
Permanent structure	−11.6	18.5	19.5
Posted train speed for the crossing			
Maximum	−0.1	0.1	0.1
Minimum	0.1	−0.2	−0.2
Crossing safety equipment			
Type of warning device present			
Cantilever flashing light signals	−2.9	4.6	4.7
Stop sign	2.8	−4.4	−4.5
Crossbucks	−4.6	7.4	7.6
Gates	16.3	−24.9	−29.6
Driver demographics			
Male	4.4	−5.9	−10.1
Age	−0.3	0.3	1.2
Occupancy of the roadway vehicle involved in the crash	−7.6	9.9	18.7
Vehicle characteristics			
Vehicle type			
Van	1.0	0.1	−6.2
Environmental factors			
Time period of the day			
12 AM–6 AM	3.7	−13.5	14.1
7 PM–12 AM	1.5	−7.1	10.0
Temperature			
33F–60F	4.0	−5.5	−9.0
>60F	2.1	−2.8	−4.7
Weather conditions			
Rain	3.0	−4.1	−6.6
Snow	13.1	−13.4	−41.9
Crash characteristics			
Role of vehicle in the crash			
Struck by the train	−8.5	10.7	21.3
Motorist action at the event of a crash			
Drove around or through the gate	−11.6	8.6	45.7
Motorist stopped on the crossing	33.0	−44.5	−76.2
Motorist did not stop	6.3	−5.4	−22.7
Estimated train speed	−1.2	1.0	4.6

6. Conclusions

This research has attempted to examine the influence of various exogenous factors on the injury severity of motor vehicle drivers involved in highway-railway crossing collisions. Specifically, the emphasis is on understanding the effect of two sets of attributes: (a) accident attributes and (b) highway-railway crossing attributes. *Accident attributes* considered include: (1) driver demographics (including gender, age, vehicle occupancy), (2) characteristics of the vehicle involved in the collision (vehicle type), (3) environmental factors (weather, lighting conditions, time of day, etc.), and (4) crash characteristics (role of vehicle in crash, etc.). *Crossing attributes* considered include: (1) crossing characteristics (Annual traffic on the highway, railway traffic, etc.), and (2) crossing safety equipment (presence of gates, traffic signals, watchmen, etc.).

In our research effort, we propose a latent segmentation based ordered logit model to examine vehicle driver injury severity. In this approach, crossings are assigned probabilistically to segments based on a host of crossing attributes. Within each of these segments, the vehicle driver injury severity is determined based on an ordered response model that considers all accident attributes. The newly formulated model allows us to partition highway-railway crossings into segments based on their attributes and estimate the influence of accident attributes on injury severity separately within each segment. The latent segmentation model developed enables transportation safety analysts to identify the crossing attributes that contribute to or mitigate the likelihood of severe injuries for vehicle drivers.

The Federal Railroad Administration (FRA) crossing database for the period 1997–2006 is employed. In this research effort, we considered three different model specifications including: (1) traditional ordered logit (OL) model, (2) latent segmentation based ordered logit model with two segments (LSOL II) and (3) latent segmentation based ordered logit model with three segments (LSOL III). The LSOL II model with two segments offered the best data fit. The segmentation component results highlight important findings: the "low risk" crossing segment is characterized by higher no. of trains, roadway classification of smaller roads, pavement markings for stop signs, absence of permanent structures obscuring the view, lower maximum posted train speed limits and presence of gates and stop signs. For the high risk segment, age, collisions during the time period 7 PM–6 AM, vehicles struck by train, aggressive driver maneuvers and estimated train speeds at the time of the collision contribute to increasing the likelihood of severe injury while driving a van and presence of snow reduce the injury severity. On the other hand, for the low risk segment, we find age, vehicle occupancy, struck by train and aggressive driver maneuvers and estimated train speed contribute to severe injury while male drivers, crashes during the time period 7 PM–6 AM, temperature 33F and above, and presence of snow and/or rain are likely to reduce injury severity. The comparison of results across the two segments is very interesting. The low risk segment injury severity profile has a large number of variables moderating the influence of exogenous variables whereas the high risk injury severity profile is characterized by very few mitigating factors.

The exogenous variable coefficients do not directly provide the elasticities of the variables, that is, the magnitude of the impact of the variables on the probability of injury severity categories. To understand the impact of various exogenous factors, elasticity effects for the exogenous variables from the LSOL II model are computed. From the elasticity results it is evident that among dummy variables the most important determinants of injury severity are crossing safety equipment, roadway classification, pavement markings for stop signs, permanent structures obscuring the view for road users, presence of snow, and aggressive maneuvers such as "drive around or through the gate". Among the continuous variables age and estimated train speed are the important determinants. These results clearly underscore the importance of allowing for impact of exogenous factors to be flexible across different segments in the data.

Acknowledgements

The authors acknowledge the insightful comments from two reviewers on an earlier version of the paper.

References

Ben-Akiva, M., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. The MIT Press, Cambridge.

Bhat, C., 1997. An endogenous segmentation mode choice model with an application to intercity travel. Transportation Science 31 (1), 34–48.

Chang, L.Y., Mannering, F., 1999. Analysis of injury severity and vehicle occupancy in truck- and non-truck-involved accidents. Accident Analysis and Prevention 31 (5), 579–592.

Depaire, B., Wets, G., Vanhoof, K., 2008. Traffic accident segmentation by means of latent class clustering. Accident Analysis & Prevention 40 (4), 1257–1266.

Desantis, S.M., Houseman, E.A., Coull, B.A., Stemmer-Rachamimov, A., Betensky, R.A., 2008. A penalized latent class model for ordinal data. Biostatistics 9 (2), 249.

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.

Greene, W., Harris, M., Hollingsworth, B., Maitra, P., 2008. A bivariate latent class correlated generalized ordered probit model with an application to modeling observed obesity levels. Department of Economics, Stern School of Business, New York University, Working Paper, 08-18.

Hu, S., Li, C., Lee, C.K., 2010. Investigation of key factors for accident severity at railroad grade crossings by using a logit model. Safety Science 48 (2), 186–194.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transportation Research – Part A 44 (5), 291–305.

Miranda-Moreno, L.F., Fu, L., Lord, D., Ukkusuri, S., 2009. How to incorporate accident severity and vehicle occupancy into the hotspot identification process? Transportation Research Record 2102, 53–60.

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.

Park, B.-J., Lord, D., 2009. Application of finite mixture models for vehicle crash data analysis. Accident Analysis and Prevention 41 (4), 683–691.

O'Donnell, C., Connor, D., 1996. Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice. Accident Analysis and Prevention 28 (6), 739–753.

Raub, R., 2009. Examination of highway-rail grade crossing collisions nationally from 1998 to 2007. Transportation Research Record: Journal of the Transportation Research Board 2122 (1), 63–71.

Saccomanno, F., Lai, X., 2005. A model for evaluating countermeasures at highway-railway grade crossings. Transportation Research Record: Journal of the Transportation Research Board 1918 (1), 18–25.

Saccomanno, F., Park, P., Fu, L., 2007. Estimating countermeasure effects for reducing collisions at highway-railway grade crossings. Accident Analysis and Prevention 39 (2), 406–416.

Washington, S., Oh, J., 2006. Bayesian methodology incorporating expert judgment for ranking countermeasure effectiveness under uncertainty: example applied to at grade railroad crossings in Korea. Accident Analysis and Prevention 38 (2), 234–247.

Windmeijer, F.A.G., 1995. Goodness-of-fit measures in binary choice models. Econometric Reviews 14, 101–116.