ELSEVIER

Accident Analysis and Prevention 40 (2008) 260–266

ACCIDENT ANALYSIS & PREVENTION

www.elsevier.com/locate/aap

Highway accident severities and the mixed logit model: An exploratory empirical analysis

John C. Milton[a,1], Venky N. Shankar[b,2], Fred L. Mannering[c,*]

[a] Washington State Department of Transportation, Urban Corridors, PO Box 47330, Olympia, WA 98504, United States

[b] Department of Civil and Environmental Engineering, Pennsylvania State University, 226C Sackett Building, University Park, PA 16802, United States

[c] School of Civil Engineering, 550 Stadium Mall Drive, Purdue University, West Lafayette, IN 47907-2051, United States

Received 8 December 2006; received in revised form 2 May 2007; accepted 16 June 2007

Abstract

Many transportation agencies use accident frequencies, and statistical models of accidents frequencies, as a basis for prioritizing highway safety improvements. However, the use of accident severities in safety programming has been often been limited to the locational assessment of accident fatalities, with little or no emphasis being placed on the full severity distribution of accidents (property damage only, possible injury, injury)—which is needed to fully assess the benefits of competing safety-improvement projects. In this paper we demonstrate a modeling approach that can be used to better understand the injury-severity distributions of accidents on highway segments, and the effect that traffic, highway and weather characteristics have on these distributions. The approach we use allows for the possibility that estimated model parameters can vary randomly across roadway segments to account for unobserved effects potentially relating to roadway characteristics, environmental factors, and driver behavior. Using highway-injury data from Washington State, a mixed (random parameters) logit model is estimated. Estimation findings indicate that volume-related variables such as average daily traffic per lane, average daily truck traffic, truck percentage, interchanges per mile and weather effects such as snowfall are best modeled as random-parameters—while roadway characteristics such as the number of horizontal curves, number of grade breaks per mile and pavement friction are best modeled as fixed parameters. Our results show that the mixed logit model has considerable promise as a methodological tool in highway safety programming.

Keywords: Mixed logit model; Random parameters; Accident injury severity prediction

1. Introduction

In the United States, highway-safety improvement programs have primarily relied on accident-frequency reduction approaches for prioritizing safety projects. Limitations with sole reliance on accident frequency approaches (simply the number of accidents), or the use of frequency-dominated approaches (the number of accidents with consideration to resulting injury severity usually only at the fatality level) are numerous. For example, accident frequency approaches tend to favor urban and suburban locations at the expense of rural locations that may have fewer but more severe accidents. Frequency-dominated approaches (typically those that consider accident frequencies along with fatalities) often make the forecasting assumption that fatality rates are identical across low- and high-volume accident locations—a process that can introduce considerable error. And, neither frequency nor frequency-dominated approaches assess accident frequency and severity with an integrated methodology, which is an important consideration for accuracy, guidance of public opinion, and defining appropriate highway-agency policies.

In terms of methodological approaches, many states have implemented accident-tracking systems in an effort to identify roadway segments with high accident frequencies and target them for improvement. For example, in Washington State, a systematic approach has been used to rank accident locations using both negative binomial accident-frequency models (Milton and Mannering, 1998) and zero-altered accident-frequency models (Shankar et al., 1997). Use of such accident-frequency

* Corresponding author. Tel.: +1 765 496 7913; fax: +1 765 494 0395.
E-mail addresses: miltonj@wsdot.wa.gov (J.C. Milton), shankarv@engr.psu.edu (V.N. Shankar), flm@ecn.purdue.edu (F.L. Mannering).

[1] Tel.: +1 360 705 7299.

[2] Tel.: +1 814 865 9434.

0001-4575/$ – see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2007.06.006

models, estimated with historical data, has allowed the State to forecast accident frequencies over roadway segments' lifetimes—offering a significant improvement over safety-programming approaches that rely exclusively on observed accident histories. Aggregate accident data seem to support the effectiveness of this safety-programming approach with accident fatalities declining by over 50% in Washington State since its introduction.

However, using accident-injury severities as a basis for safety programming presents a different problem. One possible modeling technique is a hybrid injury/frequency approach. With this technique, separate severity-frequency models are developed for each injury-severity level. Thus, for example, a series of negative binomial accident frequency models could be developed for each severity (property damage only, injury, fatality) to predict the number of accidents of each severity level on roadway segments. Unfortunately, such an approach can introduce significant estimation errors in that it implicitly assumes that the factors generating the occurrence of an accident are independent across severity outcomes. That is, such an approach implies that the frequencies of the various severity outcomes are independent of each other, which is not the case because as the frequency of one severity type changes, the frequency of other types will also likely change. As a result of severity-modeling limitations, many states simply use the observed frequency of fatal accidents as the primary severity input into their safety programming process.

To be sure, there have been many research efforts that have focused on the analysis and modeling of accident-injury severities—and various methodological approaches have been applied. For example, Yau (2004) and Lui et al. (1988) used a logistic regression approach. Bivariate models of injury outcomes have been applied by Saccomanno et al. (1996) and Yamamoto and Shankar (2004). One of the more popular approaches has been a discrete ordered-probability approach, such as an ordered probit model. This approach has considerable appeal because possible highway accident-injury outcomes are discrete and have an order from lower severity to higher severity (no injury, possible injury, evident injury, disabling injury, fatality). This class of models has been applied by a number of researchers with considerable success (see for example, O'Donnell and Connor, 1996; Duncan et al., 1998; Renski et al., 1999; Khattak, 2001; Kockelman and Kweon, 2002; Khattak et al., 2002; Kweon and Kockelman, 2003; Abdel-Aty, 2003). Still other studies have employed a variety of multinomial and nested logit structures to evaluate accident-injury severities (see Shankar and Mannering, 1996; Shankar et al., 1996; Chang and Mannering, 1999; Carson and Mannering, 2001; Lee and Mannering, 2002; Ulfarsson and Mannering, 2004; Khorashadi et al., 2005).[3] However, many of the previous studies on accident-injury severity have relied on detailed data from accident reports. These detailed, accident-specific data have been used to develop models with many statistically significant explanatory variables. While these models have certainly provided important new insights into the factors determining accident-injury severity, they have proved to be difficult to use in safety programming because of the large number of event-specific explanatory variables that need to be estimated to produce useable severity forecasts.

We seek to add to the growing body of literature on accident-injury severity by demonstrating another methodological approach. To do this, we explore the analysis of injury-severity proportions on various roadway segments. That is, we will assume reported accident frequencies on specific roadway segments are known[4] and we will develop a model that will estimate the proportions of various injury-severity levels. This differs from traditional accident-injury severity analyses which look at the injury-severity outcomes of specific accidents once they have occurred. While our proposed injury-proportion approach is more aggregate in that specific accident characteristics are not to be used (such as driver characteristics, vehicle characteristics, restraint usage, alcohol consumption, and so on), the approach has the advantage of allowing for a more general, non-event-specific interpretation of factors that determine injury-severity outcomes. When combined with accident frequency models, the proposed injury-proportions model will enable prediction of the severity distribution of accidents on roadway segments, which can then be used to asses the effect that competing projects will have on both the frequency and severity of accidents for a given segment. This represents an improvement over many current safety-programming approaches that focus severity-related concerns on observed fatalities.

2. Methodological approach

Because our proposed injury-severity proportions approach does not consider the details of specific accidents (with a focus on accident severities aggregated across roadway segments), additional heterogeneity across roadway segments could be introduced (in addition to the unobserved heterogeneity we expect relating to roadway, traffic and environmental factors). These accident-specific details can include factors such as driver gender, driver age, vehicle types, collision types, number of vehicles involved—all of which have been shown to have a significant effect on injury severities when analyzing individual-accident data (for example, see the recent work by Islam and Mannering, 2006 and Ulfarsson and Mannering, 2004). The reason that these factors influence the severity of accidents can

[3] Relative to ordered probability models, multinomial and nested logit structures do not account for the ordering of injury-severity data. However, multinomial and nested logit structures do offer a more flexible functional form by providing consistent parameter estimates in the presence of the possible under-reporting of accidents that do not involve injury. Also, they relax the parameter restriction imposed by ordered probability models that does not allow a variable to simultaneously increase (or decrease) both high and low injury severities. The monotonic effect of variables imposed by ordered probability models is discussed in Eluru and Bhat (2007), Bhat and Pulugurta (1998) and Washington et al. (2003).

[4] As mentioned earlier, methods of accident frequency prediction are well documented in the literature and easily applied to forecast the frequency of accidents on roadway segments (see for example Shankar et al., 1995, 1997; Poch and Mannering, 1996; Milton and Mannering, 1998; Carson and Mannering, 2001; Lee and Mannering, 2002).

be physical (such as the type of vehicle driven, type of object struck, etc.), physiological (with driver age and gender correlating to the ability of the body to withstand impact), and/or behavioral (with driver age and gender correlating to decisions made before and during the collision). Note that the behavioral component can relate more generally to how age and gender correlate to risk perception and risk-taking behavior. This can also include risk-compensating behavior in which drivers adjust their driving behavior in response to situations (weather conditions, traffic volumes and so on) that are perceived as comparatively dangerous or safe (see Winston et al., 2006).

In light of the above, it is important to apply a methodological approach that allows for the possibility that the influence of variables affecting accident injury-severity proportions may vary across roadway segments. This is an important consideration because, due to variations in driver behavior for example, it may be unrealistic to assume that the effects of variables such as traffic volumes, roadway geometrics, and other factors are the same across all roadway segments. Relatively recent research conducted by Train (1997), Revelt and Train (1997, 1999), Train (1999), Brownstone and Train (1999), McFadden and Train (2000), Bhat (2001), has demonstrated the effectiveness of a methodological approach (the mixed logit model) that can explicitly account for the variations (from one roadway segment to the next) of the effects that variables have on injury-severity proportions.

The application of the mixed logit model (also called the random parameters logit model) is undertaken by considering injury-severity proportions for individual roadway segments. Severity is defined as the resulting injury level of the most severely injured person in the observed accident. To develop the modeling approach, a severity function determining the proportion of injury severities (of all reported accidents per year) on a roadway segment is defined as

$$S_{\text{in}} = \boldsymbol{\beta}_i \boldsymbol{X}_{\text{in}} + \varepsilon_{\text{in}} \tag{1}$$

where S_{in} is a severity function determining the injury-severity category i proportion (property damage only, possible injury, evident injury, disabling injury and fatality) on roadway segment n; $\boldsymbol{X}_{\text{in}}$ is a vector of explanatory variables (weather, geometric, pavement, roadside and traffic variables); $\boldsymbol{\beta}_i$ is a vector of estimable parameters; and ε_{in} is error term. If ε_{in}'s are assumed to be generalized extreme value distributed, McFadden (1981) has shown that the multinomial logit model results such that

$$P_n(i) = \frac{\text{EXP}[\boldsymbol{\beta}_i \boldsymbol{X}_{\text{in}}]}{\sum_I \text{EXP}[\boldsymbol{\beta}_i \boldsymbol{X}_{\text{in}}]} \tag{2}$$

where $P_n(i)$ is the proportion of injury-severity category i (from the set of all injury-severity categories I) on roadway segment n. To generalize this to allow for parameter variations across roadway segments (variations in $\boldsymbol{\beta}$), a mixing distribution is introduced giving injury-severity proportions (see Train, 2003):

$$P_{\text{in}} = \int \frac{\text{EXP}[\boldsymbol{\beta}_i X_{\text{in}}]}{\sum_I \text{EXP}[\boldsymbol{\beta}_i X_{\text{in}}]} f(\boldsymbol{\beta}|\varphi)\, \mathrm{d}\boldsymbol{\beta} \tag{3}$$

where $f(\boldsymbol{\beta}\,|\,\boldsymbol{\varphi})$ is the density function of $\boldsymbol{\beta}$ with $\boldsymbol{\varphi}$ referring to a vector of parameters of the density function (mean and variance), and all other terms are as previously defined. Eq. (3) is the formulation for the mixed logit model. For model estimation, $\boldsymbol{\beta}$ can now account for segment-specific variations of the effect of $\boldsymbol{X}$ on injury-severity proportions, with the density function $f(\boldsymbol{\beta}\,|\,\boldsymbol{\varphi})$ used to determine $\boldsymbol{\beta}$. Mixed logit proportions are then a weighted average for different values of $\boldsymbol{\beta}$ across roadway segments where some elements of the vector $\boldsymbol{\beta}$ may be fixed and some may be randomly distributed. If the parameters are random, the mixed logit weights are determined by the density function $f(\boldsymbol{\beta}\,|\,\boldsymbol{\varphi})$. Most studies have used a continuous form of this density function in model estimation (such as a normal distribution). We will test a variety of density functions in our empirical analysis.

Maximum likelihood estimation of mixed logit models is computationally cumbersome because of the required numerical integration of the logit formula over the distribution of the random, unobserved parameters. As a result, simulation-based maximum likelihood methods are typically employed using Halton draws, which have been shown to provide a more efficient distribution of draws for numerical integration than purely random draws (see Bhat, 2003; Train, 1999). Details of the evolution of simulation-based maximum likelihood methods for estimating mixed logit models are provided in numerous references including McFadden and Ruud (1994), Geweke, Keene and Runkle (1994), Boersch-Supan and Hajivassiliou (1993), Stern (1997) and Brownstone and Train (1999).

As a final point, note that in traditional multinomial logit models the error terms ε_{in} (unobserved effects) are assumed to be extreme-value independent and identically distributed. However, in functions that determine the injury proportions on individual roadway segments, it is important to accommodate the possibility of shared unobservables among injury outcomes. Traditional multinomial logit models assume that the alternate severity outcomes are independent and, if they are not, a model specification error will result. Some past research has shown that lower severity accidents, such as property damage and possible injury, may share unobserved effects (resulting in error term correlation). Previously this problem has been resolved in the accident-severity literature by using nested logit formulations (Shankar et al., 1996; Lee and Mannering, 2002; Savolainen and Mannering, 2007). The mixed logit averts this error term problem by allowing for a more general error-correlation structure, while obviating the need for making a priori assumptions about the structure of shared observables (such as nested structures).

3. Empirical setting

To demonstrate the applicability of the mixed logit approach, data are used from Washington State's multilane divided highways. These highways, which are part of the National Highway System, are considered critical routes because of their high economic importance—they are also known for the high speed of travel, significant traffic volumes, and congestion. To obtain roadway segments of sufficient length to allow for use in Washington State safety programming, we define segments along this multilane system by median treatments (safety barriers, cables

Table 1
Descriptive statistics of select variables

Variable	Mean	Standard deviation	Minimum	Maximum
Roadway segment length in miles	2.43	2.69	0.5	19.3
Average daily traffic	37,354	3,696	3,347	172,557
Average annual precipitation in inches	29.90	21.84	4.56	131.76
Average annual snowfall in inches	15.14	42.6	0	652
Percentage of trucks	14.16	6.68	3.20	32.00
Number of interchanges per mile	0.51	0.60	0	4
Speed limit in miles per hour	59.67	5.50	20.00	65
Friction number of pavement surface	46.82	5.63	20.00	61.5
Number of horizontal curves per mile	1.44	0.95	0	5
Number of grade breaks per mile	1.89	1.69	0	20
Average daily truck traffic	4,165	3,070	549	14,032

or landform barriers). A roadway segment's beginning point was identified where a previous run of a barrier terminated (or began) and ended where the next run of a barrier was encountered (or the current run ended). Using this approach may introduce some non-homogeneity in the geometric features for some of the roadway segments (for example shoulder widths may change over the segment length), but our modeling approach can account for this. In all, our data consist of 274 roadway segments of varying lengths with a mean segment length of roughly 2.4 miles with a standard deviation of about 2.7 miles.

Historical accident data were gathered for the 1990–1994 timeframe. For each roadway segment, accidents were sorted by year, and individual accident data reports on the roadway segments were aggregated based on the most severe person-injury in the accident. Using the total number of reported accidents per roadway segment per year, proportions of injury severities were determined.

The accident data were combined with weather data from the Western Regional Climate Center which included total precipitation (all forms) and snowfall precipitation. These weather data were observed using permanent weather stations and were assigned based on proximity of the station to a roadway segment. Data from the Washington State Department of Transportation databases were used for geometric, pavement, roadside and traffic characteristics associated with roadway segments. Geometric data included number of lanes, width of lanes, shoulder widths, median width, minimum and maximum radii of horizontal curves, central angle of horizontal curves, grade, minimum grade, maximum grade, grade differential, tangent length, number of changes in grade, number of horizontal and vertical curves per mile, presence of interchanges and presence of exit/entrance ramps. Pavement data included roadway pavement type, shoulder pavement type and friction coefficients. Roadside information included slopes, presence of vegetation, ditch information, the presence of crossovers, and information on other fixed objects (such as trees and poles). Traffic operations data included speed limit, average annual daily traffic, average daily traffic per lane, single-unit truck traffic, combination truck traffic, large truck percentages, peak-hour factors and roadway access control.

Information on a total of 22,568 individual accidents was included in this study. Due to the limited number of accidents that resulted in disabling injury and fatality, it was not statistically possible to estimate all five injury-proportion categories (it was not possible to statistically differentiate all five categories). We thus consider three categories; property damage only, possible injury and injury (with the injury category including evident injury, disabling injury and fatality). With these categories, of the 22,568 individual accidents reported over the 5-year study period, 12,612 resulted in property damage only, 4,866 in possible injury and 5,090 in injury as the most severe outcome of the accident. Table 1 provides information on the mean, standard deviation, minimum and maximum of some selected variables.

4. Model estimation results

The mixed logit specification shown in Eq. (3) was estimated with simulation-based maximum likelihood. We begin by specifying a functional form of the parameter density function (see $f(\boldsymbol{\beta}\,|\,\boldsymbol{\varphi})$ in Eq. (3)). For our analysis, normal, lognormal (which restricts the impact of the estimated parameter to be strictly positive or negative), triangular and uniform distributions were considered. Once the functional form is specified, values of $\boldsymbol{\beta}$ are drawn from $f(\boldsymbol{\beta}\,|\,\boldsymbol{\varphi})$ and logit proportions are computed. To accomplish this, random draws and Halton draws were considered. However, as mentioned earlier, Bhat (2003) and Train (2003) have shown that Halton draws are more efficient and involve far fewer draws to achieve convergence. Halton draws were particularly effective in our case because we evaluate numerous distributional options for the parameters in the mixed logit model. Results that we report are from 200 Halton draws.

Table 2 shows the results of the mixed logit estimation. Due to missing data, the original 1370 observations (274 roadway segments multiplied by 5 years of data on each segment) were reduced to 1280 observations (each of which represent roadway segments' accident-injury severity proportions in a given year). All estimated parameters included in the model are statistically significant and the signs are plausible. Parameters found random were those that produced statistically significant standard errors for their assumed distribution. If their estimated standard errors were not statistically different from 0, the parameters were fixed to be constant across the roadway-segment population. The parameters found to be random were: the constant term,

Table 2
Mixed logit estimation results for annual accident-severity proportions on roadway segments

Variable	Parameter estimate	Standard error	*t*-Statistic
Property damage only			
Constant (standard error of parameter distribution)	−0.355 (1.776)	0.182 (0.694)	−1.95 (2.56)
Average daily traffic per lane in thousands (standard error of parameter distribution)	0.0403 (0.515)	0.0190 (0.122)	2.12 (4.23)
Average annual snowfall in inches (standard error of parameter distribution)	0.0974 (0.335)	0.0418 (0.173)	2.33 (1.93)
Possible injury			
Pavement friction (scaled 0–100), fixed parameter	−0.0124	0.00293	−4.21
Percentage of trucks (standard error of parameter distribution)	−0.129 (0.1143)	0.0309 (0.0298)	−4.18 (3.84)
Injury			
Average daily truck traffic in thousands (standard error of parameter distribution)	−0.302 (0.433)	0.0716 (0.111)	−4.22 (3.90)
Number of horizontal curves per mile, fixed parameter	−0.267	0.0547	−4.89
Number of grade breaks per mile, fixed parameter	−0.0712	0.0284	−2.51
Number of interchanges per mile (standard error of parameter distribution)	−0.601 (1.441)	0.190 (0.450)	−3.17 (3.20)
Number of observations		1,280	
Restricted log-likelihood (constant only)		−24,849.51	
Log-likelihood at convergence		−21,980.66	

the average daily traffic per lane, and average annual snowfall for property damage only; the percentage of trucks for possible injury, and the average daily truck traffic and the number of interchanges per mile for injury accidents. For all of the random parameters, the normal distribution was found to provide the best statistical fit.

Looking at the specific results in Table 2, the constant for the property-damage only proportion is normally distributed with mean −0.355 and standard deviation 1.776. Given these estimates, the constant term is less than 0 on 57.5% of the segments and greater than 0 on 42.5% of the segments. This variability is likely capturing the unobserved heterogeneity in the roadway segments that could include factors such as visual noise and other physical and environmental factors that we do not measure in our dataset.

The average daily traffic (ADT) per lane (in thousands of vehicles), which is defined for the property damage only function, results in a parameter that is normally distributed with a mean 0.0403 and standard deviation 0.515. Again, both the mean and standard deviation are statistically significant indicating that the parameter effect varies over the sample of roadway segments. With the estimated parameters, 46.9% of the distribution is less than 0 and 53.1% is greater than 0. This implies that in slightly less than half of the roadway segments result in a decrease in property-damage-only accidents (implying an increase in more severe injury outcomes) and slightly more than half result in an increase in property damage only.[5] This result is likely picking up a complex interaction among traffic volume, driver behavior and accident-injury severity. Because the roadway segments are scattered throughout Washington State, the finding that the effect of ADT per lane increases injury severity on some segments and decreases it on others may be capturing the response and adaptation of local drivers to various levels of traffic volume. This finding has important implications in that it suggests that the effect of traffic on injury-severity outcomes cannot be assumed to be uniform across geographic locations.

The next finding relates to the effect of average annual snowfall which is defined for the property damage only severity level. We again find this to be a normally distributed parameter with mean 0.0974 and standard deviation 0.335, which results in 38.6% of the distribution less than 0, and 61.4% of the distribution grater than 0. Thus, for almost 39% of the roadway segments, increasing snowfall decreases the likelihood of property damage only, and for about 61% it increases the likelihood of property damage only. This again is likely picking up geographical differences in driving behavior in response to snow. It appears that for the majority of roadway segments, drivers respond to higher snowfalls by driving more cautiously—thus increasing the likelihood of property damage only (and consequently decreasing the likelihood of more severe injury outcomes). However for other roadway segments, this compensating behavior (driving slower) is not sufficient to overcome the reduced friction and other adverse factors associated with snow, and thus for 38.6% of the roadway segments, increasing snowfall decreases the likelihood of property damage only accidents (and consequently increasing the likelihood of more severe injury outcomes).

Pavement friction is measured on a scale of 0–100 using a standardized test. Higher values indicate better friction, and friction numbers above 30 are considered acceptable for roadways with design speeds greater than 40 mi/h (all segments in our sample have design speeds greater than 40 mi/h). The pavement friction number was found to be significant in the possible injury function. Its estimated parameter was fixed across the roadway segments (the standard error of the distribution of this parameter, when allowed to be random, was statistically insignificant). Here we find that increasing pavement friction results in more severe or less severe accidents (the negative sign decreases the likelihood of possible injury and simultaneously increases the probability of property damage only and

[5] Although not presented in this paper, the parameter values of individual roadway segments can also be determined (as opposed to simply looking at the overall distribution of parameters across roadway segments). Please see Train (2003) for a complete description of this computational procedure.

injury). Among other factors, this may be capturing the behavior of some drivers to overcompensate and driving too aggressively (thus negating the additional friction benefits) while other drivers do not.

The percentage of trucks was also found to be significant in the possible injury function. This normally distributed parameter had a mean of −0.129 and standard deviation 0.1143, both of which are statistically significant. This gives parameters being less than 0 for 87.1% of the roadway segments and greater than 0 for 12.9% of the segments. This implies that in a small proportion of roadway segments, the truck percentage increases the proportion of possible injury accidents, while in a majority of roadway segments, the proportion tends to decrease. Note that this variable implies that for 87.1% of roadway segments increasing truck percentages make the severity proportions more likely to be minor (property damage only) or major (injury). It seems that for the great majority of roadway segments, truck percentages tend to push severity proportions to low and high extremes. However, the net effect of truck percentages on severity proportions must be considered along with the "average-daily-truck-traffic variable" which was found to be statistically significant in the injury function. This parameter estimate was normally distributed with a mean of −0.302 and standard deviation of 0.443, which gives 75.2% of the roadway segments negative values (an increasing number of trucks decreases the likelihood of accidents resulting in injury) and 24.8% positive values (an increasing number of trucks increases the likelihood of accidents resulting in injury). The net effect of these two truck variables points to a fairly complex picture of the effect of trucks on accident-injury severities. On one hand, the presence of higher truck percentages and the sheer number of trucks may have a slowing effect on travel speeds which would tend to decrease the injury-severity of accidents. On the other hand, trucks' larger size, dimensions and mass may set up conditions that result in more severe accidents.

Turning to geometric variables, the number of horizontal curves per mile in the roadway segment was found to be a fixed parameter that significantly reduced the likelihood of injury accidents for all roadway segments (Shankar et al., 1996 also found this variable to be significant in their accident-specific severity model). State highways with median sections are divided and multi-lane sections that are built to the highest highway design standards (uniform design speeds and corresponding radii and superelevations). This finding may be capturing some risk-adjusting behavior on the part of drivers. That is, as the curve density increases, individuals may adjust by driving more slowly to provide more time to process information and to increase their ability to safely negotiate the curves. Thus, these curves do not present drivers with the need to significantly and dramtically adjust their speed. In other applications, where the highway segments may have horizontal curves with unexpectedly low design speeds, or where cross-sections are undivided with two or four-lane traveled ways, this finding may change.

In a similar vein, the number of grade breaks per mile was also found to be a fixed parameter that had a significant and negative effect on injury-accident proportions (thus increasing the likelihood of lower-severity accidents). Grade breaks per mile is defined by the presence of a vertical curve or a vertical point of inflection for grade changes that would result in vertical curves below minimum curve lengths (see Mannering et al., 2005 for examples). As was the case with the horizontal curves variable, we speculate that this finding may again be capturing risk-adjusting behavior.

Finally, the number of interchanges per mile was found to be a normally distributed parameter with a mean of −0.601 and standard deviation 1.441, both of which are statistically significant. Given these values, 66.2% of the roadway segments have parameter values less than 0 and 33.8% greater 0. For the majority of roadway segments, interchange density reduces the likelihood of injury accidents, thus producing less severe accident types (lower speed, smaller collision angles). However, for some, perhaps those with more complex interchange geometrics and traffic patterns, interchange density increases the likelihood of injury accidents.

5. Summary and conclusions

The difficulties associated with modeling accident-injury severities have led many highway agencies to focus on safety-improvement programs that deal primarily with accident frequency. Where efforts have addressed accident-injury severity, the approach has generally been to identify locations that have an abnormally high number of fatalities. The ability to understand and address the accident injury-severity potential in a multivariate context (understanding how multiple factors affect injury-severity distributions) is a priority for many transportation agencies.

The modeling approach presented in this paper offers methodological flexibility that can be used as a basis for safety programs to move beyond simple accident-frequency and observed-fatality approaches. By using a combination of frequency models and the proposed mixed logit model to determine severity proportions, agencies can gain a much better understanding of the effect that possible safety enhancements will have on overall roadway safety.

From an econometric perspective, the mixed logit allows the flexibility to capture segment-specific heterogeneity that can arise from a number of factors relating to roadway characteristics, environmental factors, driver behavior, vehicle types, and interactions among these factors. The ability to consider geometric, pavement, traffic and weather-related factors directly as they relate to accident injury-severity proportions provides a much fuller understanding of the complex interaction of the numerous variables which determine transportation safety. The methodological approach proposed herein has potential to provide new insights into accident-severity analysis and to open the door for development and application of new data-modeling methodologies.

References

Abdel-Aty, M., 2003. Analysis of driver injury severity levels at multiple locations using ordered probit models. J. Safety Res. 34 (5), 597–603.

Bhat, C., 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. Transport. Res. Part B 17 (1), 677–693.

Bhat, C., 2003. Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. Transport. Res. Part B 37 (1), 837–855.

Bhat, C., Pulugurta, V., 1998. A comparison of two alternate behavioral choice mechanisms for household auto ownership decisions. Transport. Res. Part B 32 (1), 61–75.

Boersch-Supan, A., Hajivassiliou, V., 1993. Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. J. Econom. 58 (3), 347–368.

Brownstone, D., Train, K., 1999. Forecasting new product penetration with flexible substitution patterns. J. Econom. 89, 109–129.

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

Carson, J., Mannering, F., 2001. The effect of ice warning signs on accident frequencies and severities. Accident Anal. Prevent. 33 (1), 99–109.

Duncan, C., Khattak, A., Council, F., 1998. Applying the ordered probit model to injury severity in truck-passenger car rear-end collisions. Transport. Res. Rec. 1635, 63–71.

Eluru, N., Bhat, C., 2007. A joint econometric analysis of seat belt use and crash-related injury severity. Accident Anal. Prevent. 39, 1037–1049.

Geweke, J., Keane, M., Runkle, D., 1994. Alternative computational approaches to inference in the multinomial probit model. Rev. Econom. Stat. 76 (4), 609–632.

Islam, S., Mannering, F., 2006. Driver aging and its effect on male and female single-vehicle accident injuries: some additional evidence. J. Safety Res. 37 (3), 267–276.

Khattak, A., 2001. Injury severity in multi-vehicle rear-end crashes. Transport. Res. Rec. 1746, 59–68.

Khattak, A., Pawlovich, D., Souleyrette, R., Hallmark, S., 2002. Factors related to more severe older driver traffic crash injuries. J. Transport. Eng. 128 (3), 243–249.

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 Anal. Prevent. 37 (5), 910–921.

Kockelman, K., Kweon, Y.-J., 2002. Driver injury severity: an application of ordered probit models. Accident Anal. Prevent. 34 (4), 313–321.

Kweon, Y.-J., Kockelman, K., 2003. Overall injury risk to different drivers: combining exposure, frequency, and severity models. Accident Anal. Prevent. 35 (3), 414–450.

Lee, J., Mannering, F., 2002. Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accident Anal. Prevent. 34 (2), 149–161.

Lui, K., McGee, D., Rhodes, P., Pollock, D., 1988. An application of a conditional logistic regression to study the effects of safety belts, principal impact points, and car weights on drivers' fatalities. J. Safety Res. 19, 197–203.

Mannering, F., Kilareski, W., Washburn, S., 2005. Principles of Highway Engineering and Traffic Analysis, 3rd ed. John Wiley and Sons, New York, NY.

McFadden, D., 1981. Econometric models of probabilistic choice. In: Manski, D., McFadden (Eds.), A Structural Analysis of Discrete Data with Econometric Applications. The MIT Press, Cambridge, MA.

McFadden, D., Ruud, P., 1994. Estimation by simulation. Rev. Econ. Stat. 76 (4), 591–608.

McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. J. Appl. Econom. 15, 447–470.

Milton, J., Mannering, F., 1998. The relationship among highway geometrics, traffic-related elements and motor vehicle accident frequencies. Transportation 25 (4), 395–413.

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

Poch, M., Mannering, F., 1996. Negative binomial analysis of intersection-accident frequencies. J. Transport. Eng. 122 (2), 105–113.

Renski, H., Khattak, A., Council, F., 1999. Impact of speed limit increases on crash injury severity: analysis of single-vehicle crashes on North Carolina interstate highways. Transport. Res. Rec. 1665, 100–108.

Revelt, D., Train, K., 1997. Mixed logit with repeated choices: households' choice of appliance efficiency level. Rev. Econ. Stat. 80 (4), 647–657.

Revelt, D., Train, K., 1999. Customer specific taste parameters and mixed logit. Working Paper, University of California, Berkeley, Department of Economics.

Saccomanno, F., Nassar, S., Shortreed, J., 1996. Reliability of statistical road accident injury severity models. Transport. Res. Rec. 1542, 14–23.

Savolainen, P., Mannering, F., 2007. Probabilistic models of motorcyclists' injury severities in single- and multi-vehicle crashes. Accident Anal. Prevent. 39, 955–963.

Shankar, V., Mannering, F., Barfield, W., 1995. Effect of roadway geometrics and environmental factors on rural accident frequencies. Accident Anal. Prevent. 27 (3), 371–389.

Shankar, V., Mannering, F., 1996. An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. J. Safety Res. 27 (3), 183–194.

Shankar, V., Mannering, F., Barfield, W., 1996. Statistical analysis of accident severity on rural freeways. Accident Anal. Prevent. 28 (3), 391–401.

Shankar, V., Milton, J., Mannering, F., 1997. Modeling accident frequencies as zero-altered probability processes: an empirical inquiry. Accident Anal. Prevent. 29 (6), 829–837.

Stern, S., 1997. Simulation-based estimation. J. Econ. Literature 35 (4), 2006–2039.

Train, K., 1997. Mixed logit models for recreation demand. In: Kling, C., Herriges, J. (Eds.), Valuing the Environment Using Recreation Demand Models. Elgar Press, New York.

Train, K., 1999. Halton sequences for mixed logit. Working Paper, University of California Berkley, Department of Economics.

Train, K., 2003. Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge, UK.

Ulfarsson, G., Mannering, F., 2004. Differences in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger car accidents. Accident Anal. Prevent. 36 (2), 135–147.

Washington, S., Karlaftis, M., Mannering, F., 2003. Statistical and Econometric Methods for Transportation Data Analysis. Chapman and Hall/CRC, Boca Raton, FL.

Winston, C., Maheshri, V., Mannering, F., 2006. An exploration of the offset hypothesis using disaggregate data: the case of airbags and antilock brakes. J. Risk Uncertainty 32 (2), 83–99.

Yamamoto, T., Shankar, V., 2004. Bivariate ordered-response probit model of driver's and passenger's injury severities in collisions with fixed object. Accident Anal. Prevent. 36 (5), 869–876.

Yau, K., 2004. Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong. Accident Anal. Prevent. 36 (3), 333–340.