A joint econometric analysis of seat belt use and crash-related injury severity

Naveen Eluru [1], Chandra R. Bhat [*]

*The University of Texas at Austin, Department of Civil, Architectural & Environmental Engineering,
1 University Station C1761, Austin, TX 78712-0278, United States*

Abstract

This paper formulates a comprehensive econometric structure that recognizes two important issues in crash-related injury severity analysis. First, the impact of a factor on injury severity may be moderated by various observed and unobserved variables specific to an individual or to a crash. Second, seat belt use is likely to be endogenous to injury severity. That is, it is possible that intrinsically unsafe drivers do not wear seat belts and are the ones likely to be involved in high injury severity crashes because of their unsafe driving habits.

The preceding issues are considered in the current research effort through the development of a comprehensive model of seat belt use and injury severity that takes the form of a joint correlated random coefficients binary-ordered response system. To our knowledge, this is the first instance of such a model formulation and application not only in the safety analysis literature, but in the econometrics literature in general. The empirical analysis is based on the 2003 General Estimates System (GES) data base. Several types of variables are considered to explain seat belt use and injury severity levels, including driver characteristics, vehicle characteristics, roadway design attributes, environmental factors, and crash characteristics. The results, in addition to confirming the effects of various explanatory variables, also highlight the importance of (a) considering the moderating effects of unobserved individual/crash-related factors on the determinants of injury severity and (b) seat belt use endogeneity. From a policy standpoint, the results suggest that seat belt non-users, when apprehended in the act, should perhaps be subjected to both a fine (to increase the chances that they wear seat belts) as well as mandatory enrollment in a defensive driving course (to attempt to change their aggressive driving behaviors).

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Seat belt use; Crash injury severity; Random coefficients; Selective recruitment; Discrete choice models with endogeneity

1. Introduction

Traffic crashes result in several fatalities everyday on U.S. roadways, and those who manage to survive crashes are faced with such potential consequences as mental trauma, pain, expensive medical costs, and increased insurance premiums (Cohen and Einav, 2003; Chang and Mannering, 1999). The society as a whole is also at a loss, both economically and emotionally, because of these incidents.

The injury severity sustained by individuals in traffic crashes is influenced by a multitude of factors, including vehicle characteristics, roadway design characteristics, driver behavior and physiological characteristics, angle of collision, driver use of alcohol or drugs, and driver use of restraint systems. It is essential to quantify the relative magnitudes of the impact of these factors on accident severity, so that measures to prevent or reduce accident severity can be identified and implemented. The current study contributes toward this end by formulating, and estimating, a comprehensive model of injury severity.

The methodology in the paper recognizes two important econometric issues in safety analysis. First, the impact of a factor on injury severity may be moderated by various observed and unobserved variables specific to an individual or to a crash. For instance, the effectiveness of seat belt use in reducing injury severity may be higher for teenagers with their relatively unconventional driving styles. This is a case of age, an attribute available in crash data bases, impacting the influence of seat belt use on injury severity. In a similar vein, the physical frame or precise sitting posture of an individual may have an association with seat belt effectiveness. This is an instance where

* Corresponding author. Tel.: +1 512 471 4535; fax: +1 512 475 8744.
  E-mail addresses: naveeneluru@mail.utexas.edu (N. Eluru),
bhat@mail.utexas.edu (C.R. Bhat).
  [1] Tel.: +1 512 471 4535; fax: +1 512 475 8744.

unobserved characteristics (physical frame and sitting posture) moderate the effectiveness of seat belt use in reducing injury severity. In general, one could argue that there are several subtle, unobserved, characteristics that moderate the effect of factors influencing injury severity. Ignoring such unobserved heterogeneity can, and in general will, result in inconsistent estimates in nonlinear models (see Chamberlain, 1980; Bhat, 2001). The second issue addressed in our econometric framework is the endogeneity of seat belt use to injury severity (also referred to as selective recruitment in the safety analysis literature; see Evans, 1996; Derrig et al., 2000). That is, seat belt non-users may be intrinsically unsafe drivers. For example, Janssen (1994) provides empirical evidence that seat belt non-users drive significantly faster than seat belt users, after controlling for gender, age, annual kilometerage, and number of years of possession of driver's license. Petridou and Moustaki (2000) and Shinar (1993) also allude to this risk taking behavior of seat belt non-users. Thus, it is quite likely that seat belt non-users are the ones likely to be involved in high injury severity crashes because of their unsafe driving habits. If this sample selection is ignored (as has been done in several previous studies), the result is an artificially inflated estimate of the effectiveness of the seat belt use.

The methodology used in the paper to address the two econometric issues discussed above takes the form of a joint binary logit-ordered logit structure with random coefficients. In particular, seat belt use is modeled using a binary logit structure, while injury severity is modeled using an ordered response structure. A host of driver characteristics, vehicle characteristics, roadway design attributes, environmental characteristics, and crash characteristics, and the interactions of these characteristics, are considered in the joint model. The moderating influence of unobserved factors associated with the impact of these attributes is accommodated by imposing a random coefficients structure in the ordered logit model. The potential self selection in seat belt use based on injury severity propensity is considered by tying the binary seat belt use component and the ordered response injury severity component of the joint model through a common unobserved random term. The joint model is subsequently applied in an empirical analysis that uses data from the 2003 General Estimates System (GES), a nationally representative sample of police-reported crashes of all types in the U.S.

The rest of this paper is structured as follows. The next section provides an overview of the crash-related injury severity modeling literature, and positions the current study in the context of earlier studies. Section 3 presents the econometric framework. Section 4 discusses the data source and sample used in the empirical analysis. Section 5 presents the empirical results. Section 6 concludes the paper by summarizing the important findings and identifying policy implications.

2. Earlier studies

Crash injury severity has been extensively researched in the safety analysis literature. Section 2.1 reviews earlier injury severity studies that do not consider seat belt use as an endogenous variable, while Section 2.2 discusses earlier research studies that either model, or extensively discuss the need to consider, seat belt use as an endogenous variable in analyzing injury severity. Section 2.3 provides a summary and positions the current research.

2.1. Seat belt use not considered endogenous to the modeling framework

A number of studies have examined crash-related injury severity, while considering seat belt use and several other attributes as exogenous variables. Most of these injury severity studies undertake the analysis at the level of individual accidents, rather than using an aggregate-level dependent variable such as the number of annual accidents in a county or state (but see Lourens et al., 1999; Doherty et al., 1998; Derrig et al., 2000, for examples of aggregate-level studies). The reason for using a disaggregate-level analysis (i.e., an analysis at the level of individual accidents) is that it better captures the fundamental relationship between accident severity and its determinants, rather than capturing spurious correlations from ignoring the heterogeneity of accidents in an aggregate-level analysis (see Kassoff and Deutschman, 1969 for an extensive discussion). Within the group of disaggregate-level injury severity studies, the early research efforts (those before 2000) applied frameworks such as log-linear analysis (Golob et al., 1986; Kim et al., 1994; Abdel-Aty et al., 1998), factorial analysis (see McLellan et al., 1996) and descriptive analysis (Evans, 1990; Evans and Frick, 1988; Cooper, 1994; Huelke and Compton, 1995). In the past several years, however, almost all injury severity studies have used a discrete variable framework because accident reports collect injury severity in discrete categories.

The discrete variable studies of crash-related injury severity have used one or more of the following five categories of variables: (1) driver attributes (including demographics and such behavioral characteristics as seat belt use and drug/alcohol use), (2) characteristics of the vehicle(s) involved in the crash (vehicle weight and type of vehicle(s)), (3) roadway design attributes (number of lanes, grade, alignment, presence of shoulders, lane widths and speed limits), (4) environmental factors (weather, lighting conditions, time of day, etc.), and (5) crash characteristics (manner of collision, role of vehicle in crash, whether there was a rollover of one or more vehicles, whether driver was ejected, etc.). A review of the earlier discrete choice studies of injury severity, and the categories of variables considered in each study, is presented in Table 1.

Three important observations may be made from Table 1. First, except for the studies by Chang and Mannering (1999) and Ulfarsson and Mannering (2004), none of the earlier studies has comprehensively considered all the five categories of variables. Second, the two most prevalent structures used to examine injury severity are logistic regression models and ordered-response models. The logistic regression models are binary logit models that focus on whether or not there is a severe injury associated with a crash (severe injury is defined either as a fatality or some other severe characterization of injury). The ordered-response models consider the entire range of injury severity levels and, therefore, capture and provide more injury

Table 1
Summary of existing discrete choice studies of crash injury severity

Paper	Research methodology	Accident characteristics considered in the empirical framework				
		Driver attributes	Vehicular characteristics	Roadway design attributes	Environmental factors	Crash characteristics
Shibata and Fukuda (1993)	Logistic regression	Yes	–	–	–	Yes
Farmer et al. (1996)	Logistic regression	Yes	Yes	–	–	Yes
Khattak et al. (1998)	Ordered and binary probit models	–	–	–	Yes	–
Renski et al. (1999)	Ordered probit model	–	–	Yes	–	–
O'Donnell and Connor (1996)	Ordered logit and probit models	Yes	Yes	–	–	Yes
Chang and Mannering (1999)	Nested logit model	Yes	Yes	Yes	Yes	Yes
Krull et al. (2000)	Logistic regression	Yes	Yes	Yes	–	Yes
Al-Ghamdi (2002)	Logistic regression	–	Yes	Yes	Yes	Yes
Kockelman and Kweon (2001)	Ordered probit model	Yes	Yes	Yes	–	Yes
Bedard et al. (2002)	Multivariate logistic regression	Yes	Yes	–	–	Yes
Dissanayake and Lu (2002)	Logistic regression	Yes	–	Yes	Yes	–
Ulfarsson and Mannering (2004)	Multinomial logit	Yes	Yes	Yes	Yes	Yes
Kweon and Kockelman (2002)	Ordered probit and poison models	Yes	Yes	–	–	–
Khattak et al. (2002)[a]	Ordered probit model	–	Yes	Yes	Yes	Yes
Srinivasan (2002)	Random thresholds ordered logit model	Yes	Yes	–	Yes	Yes
Toy and Hammitt (2003)	Logistic regression	Yes	Yes	–	–	Yes
Khattak and Rocha (2003)[b]	Ordered logit model	–	–	–	–	Yes
Abdel-Aty and Abdelwahab (2004)	Nested logit model	Yes	Yes	–	Yes	Yes
Wang and Kockelman (2005)	Heteroscedastic ordered logit model	Yes	Yes	Yes	Yes	–

[a] The analysis is restricted to driver aged 65 and above.
[b] The analysis is confined to sports utility vehicles.

severity information (relative to the logistic regression models). The ordered-response models used in the past for injury severity analysis take the form of either an ordered-response logit or an ordered-response probit structure. Both these ordinal model forms are essentially equivalent, and differ only in whether a logistic or a normal distribution is used for the stochastic component in the latent propensity that is assumed to underlie the observed injury severity.[2] Third, none of the existing studies allow randomness in the effects of injury severity determinants due to the moderating influence of unobserved factors. Srinivasan (2002) allows randomness due to unobserved factors in the threshold bounds that relate the underlying latent injury severity propensity to the observed injury severity categories, but does not address the randomness in the effects of injury severity determinants. Of course, none of the studies in Table 1 also consider seat belt as being endogenous in their modeling frameworks.

---

[2] While the ordered-response models have been used only within the past 7–8 years in the safety analysis literature, they have a long history of use in other transportation contexts; see Kitamura and Bunch (1990), Bhat (1991), and Bhat and Koppelman (1993). The reader will also note that the ordered-response model is perhaps more suited than the multinomial logit model for injury severity because of the correlation between adjacent injury severity levels. However, a limitation of the ordered-response structure is that it imposes a certain kind of monotonic effect of exogenous variables on injury severity levels (see Bhat and Pulugurta, 1998 for a detailed exposition of the relationship between ordered and unordered response models). Ideally, one would consider an ordered generalized extreme value model for injury severity that combines the flexibility offered by the unordered-response structure with the proximate covariance characteristic due to the ordinality in the injury severity levels. The authors are currently undertaking a research study to compare such an OGEV structure with an ordered-response structure.

2.2. Seat belt endogenous to the modeling framework

A number of earlier studies have alluded to the "selective recruitment" of seat belt non-users in crashes involving severe injuries. One of the early studies that discusses the selective recruitment (or sample selection) issue conceptually is Evans (1985). However, the first empirical validation of the sample selection hypothesis appears to have been undertaken by Evans (1996), who used a probability sample of police-reported crashes in the U.S. between 1982 and 1991 from the National Accident Sampling System (NASS) to examine the relationship between crash severity and seat belt use. Evans measured crash severity in terms of the change in velocity due to the crash, which itself was inferred using structural equations based on the level of vehicle deformation in the crash. Evans' results indicated an over-representation of unbelted drivers in high crash severity accidents. To the extent that crash severity level is correlated with injury severity level, Evans' results provide evidence that unbelted drivers are intrinsically more likely to be involved in high injury severity crashes. Evans concludes that seat belt effectiveness is overestimated by a large amount if the sample selection is not accounted for.

Another study that indirectly provides support for the sample selection hypothesis is Dee (1997), who examined why seat belt laws that increased seat belt usage sharply in the late 1980s and early 1990s had a relatively small impact on crash-related fatalities. One of the hypotheses he considered to explain this apparent paradox was that of sample selection. That is, unsafe drivers are more likely than the general population to continue not to wear seat belts even after passage of seat belt laws. If such unsafe drivers are also more likely to be involved in severe crashes, the net result would only be a small impact on

crash-related fatalities. To test the hypothesis, Dee used the Center for Disease Control and Prevention's (CDC) annual Behavioral Risk Factor Surveillance System (BRFSS) telephone surveys collected between 1985 and 1993. Dee compared the reported seat belt usage of crash-prone individuals and the general population after the passage of seat belt laws. His analysis provides evidence that crash-prone individuals are more likely not to wear seat belts than the general population after the enactment of seat belt laws, a finding consistent with the sample selection hypothesis.

Cohen and Einav (2003) examined the impact of seat belt usage on crash-related vehicle occupant fatalities using data from the fatality analysis reporting system (FARS) collected between 1983 and 1997. The FARS data on traffic fatalities were aggregated to obtain the total number of annual fatalities by U.S. state. The authors then used a log-linear regression model to relate the logarithm of the number of occupant fatalities per vehicle mile of travel in each state to (1) the seat belt usage rate in the state, (2) a set of demographic, traffic density, crime and fuel tax rate control variables in the state, (3) fixed state effects to control for the potential endogeneity of usage rate (for example, states with high crash-related fatalities may institute enforcement strategies that influence usage rates) and (4) fixed year effects. In addition, to address endogeneity of seat belt usage rates, the authors instrumented the usage rate through variables related to mandatory seat belt laws. The overall finding from this aggregate level analysis is that ignoring seat belt usage rate endogeneity leads to a substantial bias in the effect of seat belt usage rate on the logarithm of per-capita vehicle occupant fatalities.

It is interesting that the three sample selection studies discussed above have been based on a simple univariate descriptive analysis (Evans, 1996), or a simple examination of seat belt usage between pre-defined accident prone groups and the general population (Dee, 1997), or an aggregate level analysis that can mask heterogeneity in crash outcomes and characteristics (Cohen and Einav, 2003).

2.3. Summary of earlier studies and the current research

The overview of the literature indicates the substantial earlier research on crash-related injury severity determinants. Increasingly, the methodology of choice for modeling injury severity is the ordered-response framework, which recognizes the ordinal nature of injury severity in police-reported accidents. However, the ordered-response models need to be enhanced to (1) comprehensively consider interactions among groups of potential determinants of injury severity, (2) allow randomness in the effects of injury severity determinants due to the moderating influence of unobserved factors, (3) recognize the potential, and very likely, endogeneity of seat-belt use in injury severity modeling, and (4) accommodate the potential randomness in the effect of seat belt use on injury severity. It is indeed surprising, in particular, that there have been very few studies to date that recognize the potential endogeneity of seat belt use. The handful of studies that do so are focused toward testing the selective recruitment hypothesis using univariate, descriptive, and aggregate analy-

ses, rather than the multivariate, methodologically rigorous, and disaggregate discrete choice framework adopted by the studies that do not consider seat belt endogeneity.

In this paper, we bring the two streams of earlier work (those that do not consider seat belt endogeneity and those that do) together by developing a comprehensive, multivariate, methodologically rigorous, and disaggregate-level model of seat belt use and injury severity that takes the form of a joint correlated random coefficients binary-ordered response system. This joint system is formulated as a mixing model that conveniently, and at once, considers all the issues of (1) systematic interaction effects among variables, (2) random unobserved effects in the influence of injury severity determinants, (3) potential endogeneity of seat belt use in modeling injury severity level, and (4) random variations in seat belt use effectiveness. To our knowledge, this is the first instance of such a model formulation and application not only in the safety analysis literature, but in the econometrics literature in general. In addition to the methodological considerations, we consider a comprehensive set of potential determinants of injury severity in the empirical analysis. The focus in the analysis is exclusively on driver injury severity (as opposed to the injury severity of other vehicle occupants).

3. Econometric framework

3.1. Model structure

Let q ($q = 1, 2, \ldots, Q$) be an index to represent drivers and let k ($k = 1, 2, 3, \ldots, K$) be an index to represent injury severity. The index k, for example, may take values of "no injury" ($k = 1$), "possible injury" ($k = 2$), "non-incapacitating injury" ($k = 3$), "incapacitating injury" ($k = 4$), and "fatal injury" ($k = 5$), as in the empirical analysis in the current paper. The equation system for the joint driver seat belt use and injury severity model is:

$$s_q^* = (\beta' + \gamma_q')x_q + \eta_q + \varepsilon_q, \quad s_q = 1 \text{ if } s_q^* > 0; s_q = 0$$

$$\text{otherwise} \quad y_q^* = (\alpha' + \delta_q')z_q \pm \eta_q + (\theta + \mu'w_q + \lambda_q)s_q + \xi_q,$$

$$y_q = k \text{ if } \psi_{k-1} < y_q^* < \psi_k \tag{1}$$

The first equation is associated with the latent propensity s_q^* of seat belt use for driver q. s_q is the actual observed seat belt use by driver q, and x_q is an ($M \times 1$)-column vector of attributes (including a constant) associated with driver q (for example, sex, age, soberness status, etc.) and driver q's trip environment (for example, roadway speed limits, time of day, etc.). β represents a corresponding ($M \times 1$)-column vector of mean effects of the elements of x_q on seat belt use propensity, while γ_q is another ($M \times 1$)-column vector with its mth element representing unobserved factors specific to driver q and her/his trip environment that moderate the influence of the corresponding mth element of the vector x_q. η_q captures common unobserved factors influencing driver q's seat belt use propensity and the driver's injury severity propensity (for instance, an intrinsically cautious and responsible driver is likely to wear seat belts and drive defensively, incurring less severe injuries in crashes). ε_q is

an idiosyncratic random error term assumed to be identically and independently standard logistic distributed across individuals q.

The second equation is associated with the latent propensity y_q^* associated with the injury severity sustained by driver q in

$$L_q(\Omega_{-\Sigma}|c_q) = G[g_q\{(\beta' + \gamma_q')x_q + \eta_q\}]$$

$$\times \{G[\psi_k - \{(\alpha' + \delta_q')z_q + (\theta + \mu w_q + \lambda_q)s_q \pm \eta_q\}] - G[\psi_{k-1} - \{(\alpha' + \delta_q')z_q + (\theta + \mu w_q + \lambda_q)s_q \pm \eta_q\}]\}^{d_{qk}}, \quad (2)$$

the accident. This latent propensity y_q^* is mapped to the actual injury severity level y_q by the ψ thresholds ($\psi_0 = -\infty$ and $\psi_k = \infty$) in the usual ordered-response fashion. z_q is an ($L \times 1$) column vector of attributes (not including a constant and not including seat belt use) that influences the propensity associated with injury severity. α is a corresponding ($L \times 1$)-column vector of mean effects, and δ_q is another ($L \times 1$)-column vector of unobserved factors moderating the influence of attributes in z_q on the injury severity propensity for driver q. θ is a scalar constant, w_q is a set of driver/crash attributes that moderate the effect of seat belt use on injury severity, and μ is a corresponding vector of coefficients. λ_q is an unobserved component influencing the impact of seat belt effectiveness for driver q, and ξ_q is an idiosyncratic random error term assumed to be identically and independently standard logistic distributed across individuals q.

The $\pm$ sign in front of η_q in the injury severity equation indicates that the correlation in unobserved factors between seat belt use and injury severity may be positive or negative. A positive sign implies that drivers who use seat belts are intrinsically more likely to incur severe injuries in crashes, while a negative sign implies that drivers who use seat belts are intrinsically less likely to incur severe injuries in accidents. Clearly, we expect, from an intuitive standpoint, that the latter case will hold. However, one can empirically test the models with both '+' and '−' signs to determine the best empirical result. Of course, if the correlation between the seat belt use and injury severity propensities is ignored, when actually present, it results in a "corrupt" estimation of the effectiveness of seat belt use in reducing injury severity. More specifically, if the unobserved correlation between seat belt use and injury severity propensities is negative, as we expect, ignoring this correlation would result in an inflated effectiveness of seat belt use in reducing injury severity.

To complete the model structure of the system in Eq. (1), we need to specify the structure for the unobserved vectors γ_q and δ_q, and the unobserved scalars λ_q and η_q. In the current paper, we assume that the γ_q and δ_q elements, and λ_q and η_q, are independent realizations from normal population distributions: $\gamma_{qm} \sim N(0, \sigma_m^2)$, $\delta_{ql} \sim N(0, \omega_l^2)$, $\lambda_q \sim N(0, \tau^2)$, and $\eta_q \sim N(0, \upsilon^2)$.

3.2. Model estimation

The parameters to be estimated in the joint model system of Eq. (1) are the β, α and μ vectors, the θ scalar, the ψ thresholds, and the following variance terms: σ_m^2, ω_l^2, τ^2, and υ^2. Let Ω represent a vector that includes all these parameters to be estimated. Also, let c_q be a vector that vertically stacks the γ_q and δ_q vectors,

and the λ_q and η_q scalars. Let Σ be another vertically stacked vector of standard errors σ_m, ω_l, τ, and υ, and let $\Omega_{-\Sigma}$ represent a vector of all parameters except the standard error terms. Finally, let $g_q = 2s_q - 1$. Then, the likelihood function, for a given value of $\Omega_{-\Sigma}$ and error vector c_q, may be written for driver q as:

where $G(.)$ is the cumulative distribution of the standard logistic distribution and d_{qk} is a dummy variable taking the value 1 if driver q sustains an injury of level k and 0 otherwise. Finally, the unconditional likelihood function can be computed for driver q as:

$$L_q(\Omega) = \int_{c_q} (L_q(\Omega_{-\Sigma})|c_q) \, \mathrm{d}F(c_q|\Sigma), \quad (3)$$

where F is the multidimensional cumulative normal distribution. The log-likelihood function is

$$L(\Omega) = \sum_q L_q(\Omega). \quad (4)$$

The likelihood function in Eq. (3) involves the evaluation of a multi-dimensional integral of size equal to the number of rows in c_q. This multi-dimensional integration cannot be accomplished using general purpose numerical methods such as quadrature, since quadrature techniques cannot evaluate the integrals with sufficient precision and speed for estimation via maximum likelihood (see Hajivassiliou and Ruud, 1994).

We apply simulation techniques to approximate the integrals in the likelihood function and maximize the logarithm of the resulting simulated likelihood function across individuals with respect to Ω. The simulation technique approximates the likelihood function in Eq. (4) by computing the $L_q(\Omega_{-\Sigma}|c_q)$ for each q at different realizations of c_q drawn from a multivariate normal distribution, and computing the individual likelihood function by averaging over the different values of the integrand across the different realizations. Notationally, if $\mathrm{SL}_q^h(\Omega)$ is the realization of the likelihood function in the hth draw ($h = 1, 2, \ldots, H$), then the individual likelihood function is approximated as:

$$\mathrm{SL}_q(\Omega) = \frac{1}{H} \sum_{h=1}^{H} \mathrm{SL}_q^h(\Omega), \quad (5)$$

where $\mathrm{SL}_q(\Omega)$ is the simulated likelihood function for the qth observation, given the parameter vector Ω. $\mathrm{SL}_q(\Omega)$ is an unbiased estimate of the actual likelihood function $L_q(\Omega)$. Its variance decreases as H increases. It also has the appealing properties of being smooth (i.e., twice differentiable) and being strictly positive for any realization of draws.

The simulated log-likelihood is constructed as:

$$\mathrm{SL}(\Omega) = \sum_q \ln[\mathrm{SL}_q(\Omega)]. \quad (6)$$

The parameter vector Ω is estimated as the value that maximizes the above simulated function. Under rather weak regularity conditions, the maximum (log) simulated likelihood

(MSL) estimator is consistent, asymptotically efficient, and asymptotically normal (see Hajivassiliou and Ruud, 1994; Lee, 1992).

In the current paper, we use a quasi-Monte Carlo (QMC) method proposed by Bhat (2001) for discrete choice models to draw realizations for c_q from its population multivariate distribution. QMC methods are similar to the familiar Monte Carlo method in that they evaluate a multidimensional integral by replacing it with an average of values of the integrand computed at discrete points (see Eq. (5)). However, rather than using pseudo-random sequences for the discrete points, the QMC approach uses "cleverly" crafted non-random and more uniformly distributed sequences (labeled as QMC sequences) within the domain of integration. The underlying idea of the QMC methods is that it is really inconsequential whether the discrete points are truly random; of primary importance is the even distribution (or maximal spread) of the points in the integration space. Within the broad framework of QMC sequences, we specifically use the Halton sequence in the current analysis.

4. Data

4.1. Data source

The data source used in this study is the 2003 General Estimates System (GES) obtained from the National Highway Traffic Safety Administration's National Center for Statistics and Analysis. The GES consists of data compiled from a sample of police-reported accidents that involve at least one motor vehicle traveling on a traffic way and resulting in property damage, injury, or death. The GES data are drawn from accidents in about 60 areas across the U.S. that reflect the geography, population, and traffic density of the U.S. (the reader is referred to ftp://ftp.nhtsa.dot.gov/GES/GES03/SAS for comprehensive details of how the accident reports are collected and compiled). The 2003 GES includes information regarding 60,000 accidents involving about 150,000 individuals and 100,000 vehicles.

A number of accident-related attributes are collected for each accident in the GES, including the characteristics of the drivers involved, vehicle characteristics, roadway design attributes, environment attributes, and crash characteristics. The injury severity of each individual involved in the accident is collected on a five point ordinal scale: (1) no injury, (2) possible injury, (3) non-incapacitating injury, (4) incapacitating injury, and (5) fatal injury.

4.2. Sample description

In the current analysis, we examine seat belt usage and injury severity of drivers of passenger vehicles. The focus on drivers is because seat belt usage data is better recorded for drivers than for non-drivers. We also confined our attention to non-commercial drivers because of potential systematic differences between commercial and non-commercial drivers (commercial drivers are professionally trained and have to follow company-related and insurance-related driving protocols). Finally, our analysis is confined to crashes (accidents involving collision

Table 2
Cross tabulation of injury severity and seat belt use

Injury severity	Seat belt		All drivers
	Not used	Used	
No injury	27.6[a]	67.3	64.6
Possible injury	8.7	12.5	12.3
Minor injury	17.5	10.0	10.5
Serious injury	39.2	9.6	11.6
Fatality	6.9	0.6	1.0

[a] The numbers in the cell represent column percentages (the sum of the figures in each column is 100%).

with a fixed object or other vehicles rather than non-collision accidents such as rolling over) and further to the vast majority of crashes in which one or two vehicles are involved.

The final data sample of non-commercial driver crashes consisted of about 50,000 records. Of these, 11,388 records were sampled so that the distribution of injury severity in this smaller sample was about the same as the weighted distribution of injury severity in the full sample of about 50,000 records (the weighted full GES dataset is intended to replicate the overall national statistics of crashes and injury severity). The seat belt use in the weighted sample is as follows: used seat belts (93.2%) and did not use seat belts (6.8%).[3] The distribution of injury severity across the observations and by seat belt use is provided in Table 2. Clearly, the table shows a negative association between seat belt use and injury severity. One of the issues to be addressed in this research is to estimate how much of the association is due to "true" seat belt use effectiveness and how much is due to "spurious" effects.

5. Empirical analysis

5.1. Variables considered

Several types of variables were considered in the empirical analysis, including driver characteristics, vehicle characteristics, roadway design attributes, environmental factors, and crash characteristics.

Driver characteristics included driver demographics (age and sex) and driver alcohol use.[4] The only *vehicle characteristics*

---

[3] The seat belt use rate of 93.2% in the GES sample is on the high side relative to national seat belt use rates, perhaps due to potential misreporting/misrecording of seat belt use. As indicated by Schiff and Cummings (2004), police officers often classify unbelted survivors as belted when they were actually not. Given that there s a much higher proportion of survivors from crashes, the Schiff and Cummins study implies that seat belt use percentage will be much higher than it should be, as is the case in the current sample. Thus, the estimated effectiveness of seat belt use in reducing injury severity should be viewed with caution in the current study. However, this issue should not detract from the analysis in the paper of how much the seat belt effectiveness may be attributed to the "true" value of restraint systems and how much may be due to the spurious effect of seat belt non-users intrinsically being more risky drivers who get themselves into more severe accidents.

[4] The GES data included information on drug use and airbag use. However, a large fraction of records had missing information on these variables, as well as their imputed counterparts. So we excluded these driver behavior variables

included in the current study are the vehicle types involved in the crash (the vehicle types include passenger cars, sports utility vehicles, pick up trucks, and minivans). Other vehicle characteristics, such as vehicle weight, vehicle speed just before impact, and seating configuration, are either not available in, or missing for a large fraction of, the GES data. The *roadway design attributes* considered in the analysis are speed limit and roadway functional class (whether the accident occurred on an interstate highway, or arterial, or other roads). Again, additional roadway design attributes, such as number of lanes, alignment of roads, and grade and shoulder widths, could not be included because of the absence of data, or the large fraction of missing data, on these variables in the GES. *Environmental factors* related to the crash that were considered included day of the week, time of day,[5] lighting conditions (dawn, daylight, dusk, dark, and dark and lit), and weather conditions (no adverse weather, rain, snow, and fog). Finally, the *crash characteristics* included whether or not the person was ejected from the vehicle, if the vehicle rolled over, whether the crash was with a stationary object or another vehicle, and the manner of collision in crashes with another vehicle (head-on, rear-end, angle, sideswipe when traveling in the same direction, and sideswipe when traveling in opposite directions), and the role of the driver's vehicle in crashes with another vehicle (i.e., whether the driver's vehicle struck the other vehicle, or the driver's vehicle was struck by the other vehicle, or both vehicles struck each other).

In addition to the five groups of variable discussed above, we also considered several interaction effects among the variables in both the seat belt use and injury severity model. The final specification was based on a systematic process of removing statistically insignificant variable and combining variables when their effects were not significantly different. The specification process was also guided by prior research and intuitiveness/parsimony considerations. We should also note here that, for the continuous variables in the data (such as age and speed limits), we tested alternative functional forms that included a linear form, a spline (or piece-wise linear) from, and dummy variables for different ranges.

5.2. Model specification

We estimated three different models in the research effort: (1) a simple binary choice logit for seat belt use and an independent ordered response logit for injury severity, which we will refer to as the independent binary-ordered (IBO) model, (2) a random coefficients binary choice logit for seat belt use and an independent random coefficients ordered response logit for injury severity, which we will refer to as the independent random binary-ordered model (IRBO), and (3) a random coefficients binary choice logit for seat belt use and a correlated random coefficients ordered response logit for injury severity,

Table 3
Estimates of the seat belt use component of joint model

Variables	Coefficient	t-Stats
Constant	3.351	12.77
Driver characteristics		
Male	−0.574	−6.17
Age variables (age < 25 years is base)		
25–29 years	0.327	2.25
30–64 years	0.222	2.56
65–74 years	1.226	2.92
Under the influence of alcohol	−2.255	−12.88
Vehicle characteristics (pick-up is base)		
Sedan	0.331	3.45
SUV	1.155	6.58
Minivan	0.606	3.24
Environmental factors		
9 a.m.–7 p.m.	0.601	5.96

which we will label as the correlated random binary-ordered (CRBO) model. In the context of the model formulation in Section 3.1, the IBO model imposes the assumptions that $\sigma_m^2 = 0$ for all m, $\omega_l^2 = 0$ for all l, and $\tau^2 = \upsilon^2 = 0$. The IRBO model imposes the assumption that $\upsilon^2 = 0$.

The final specifications of the random-coefficients in the seat belt use and injury severity components of the IRBO and the CRBO models were obtained after extensive testing. In the following presentation of empirical results, we will discuss only the CRBO models for the sake of presentation ease. However, we will use the IBO and IRBO models as yardsticks to evaluate the performance of the CRBO model.

5.3. Estimation results

5.3.1. Seat belt use component

Table 3 provides the results of the seat belt use component of the CRBO model (the coefficients represent the effects of the variables on the latent propensity to wear seat belts).

The specific effects of the driver characteristics indicate that men, younger individuals (age <25 years), and those driving under the influence of alcohol are less likely to use seat-belts compared to women, older individuals (age ≥25 years) and those not driving under the influence of alcohol, respectively (these results are consistent with earlier seat belt use studies; for example, see Reinfurt et al., 1996; Preusser et al., 1991).[6] The effects of the vehicle characteristics indicate that individuals driving a pick-up are the least likely ones to wear a seat belt, while sports utility vehicle (SUV) drivers are the most likely to wear seat belts. This association between vehicle type and seat belt-use is perhaps the manifestation of the link between safety-consciousness and type of vehicle owned. Finally, the time of

---

from consideration. However, data was available for almost all records for an imputed version of driver alcohol use.

[5] Time of day is represented in the following five categories: early morning (12 am–6 am), AM peak (6 am–9 am), midday (9 am–3 pm), PM peak (3 pm–7 pm), and evening (7 pm–12 pm).

---

[6] We examined differential effects of teenagers (≤19 years of age) and adults between the ages of 20 and 24 years. However, we did not find statistically different propensities to wear seat belts between these two age groups, and so combined these two age groups into a single "age <25 years" category.

day variables suggest that drivers are more likely to wear seat belts during the midday (9 a.m.–3 p.m.) and PM peak periods (3 p.m.–7 p.m.) than the early morning (12 a.m.–6 a.m.), AM peak (6 a.m.–9 a.m.) and evening (7 p.m.–12 a.m.) periods. The higher non-use of seat belts during the early morning and evening periods may be the result of fewer law-enforcement officials on the streets during these times.

5.3.2. Injury severity component

Table 4 presents the results of the injury severity component of the CRBO model (the parameters indicate the effects of variables on the latent propensity associated with injury severity). The results are discussed by variable group.

5.3.2.1. Driver characteristics.
The impact of driver characteristics show significant variations based on demographics and alcohol influence. In particular, men and young adults (<25 years of age) are less likely to sustain severe injuries relative to women and older adults, respectively, a result also observed in earlier studies of injury severity (see, for example, O'Donnell and Connor, 1996; Kim et al., 1994; Srinivasan, 2002).[7] The likelihood of being injured severely is highest for women over 74 years of age, while the likelihood of not being injured severely is highest for men younger than 25 years of age. Consistent with the findings from earlier studies and intuition, drivers under the influence of alcohol are likely to be more severely injured than those who are sober.

5.3.2.2. Vehicle characteristics.
The type of the driver's vehicle as well as the vehicle type of the other vehicle involved in dual-vehicle crashes were considered in the injury severity component of the joint model. In addition to main effects, combinations of the driver vehicle type and the other vehicle type, and interactions of vehicle type with all the four other variable groups, were considered. The final specification, however, comprised only three variables related to vehicle type (see Table 4). The results show that drivers in sedans are likely to be injured more severely in crashes compared to drivers in other vehicle types (SUVs, pick-up trucks, and minivans). This is particularly the case in the presence of snow and/or fog, and in crashes where the driver's sedan is struck by a non-sedan.

5.3.2.3. Roadway attributes.
The only roadway design attributes considered in the current analysis are speed limit and roadway functional class (and interactions of the two). However, once speed limit was controlled for, roadway functional class did not have any additional significant effects, because of the strong correlation between speed limits and roadway functional class. The results indicate that, on average, driver injury tends to be most severe for crashes on medium-to-high speed limit roads (26–64 mph). Also, driver injury from crashes tends to be,

---

[7] As for the case of seat belt use, we examined differential injury severity effects for teenagers (≤19 years of age) and adults between the ages of 20 and 24 years. However, due to the lack of statistically different injury severity propensities between the two age groups, they were combined into a single "age <25 years" category.

Table 4
Injury severity component of joint model

Variables	Coefficient	t-Stats
Driver characteristics		
Male	−0.454	−7.59
Age variables (age <25 years is base)		
25–74 years	0.207	2.81
>74 years	0.371	2.46
Male age <25 years	−0.173	−1.63
Under the influence of alcohol	0.465	2.48
Vehicle characteristics		
Sedan	0.288	4.98
Sedan × snow/fog	0.714	3.17
Sedan × struck by a non-sedan	0.131	1.62
Roadway design attributes		
Medium-to-high speed limit (26–64 mph)	0.906	9.28
High speed limit (≥64 mph)	0.358	1.11
S.D.	1.554	2.94
Environmental factors		
6 a.m.–7 p.m.	−0.308	−4.09
Lighting conditions		
Dusk	−0.237	−1.44
Dark	−0.398	−2.74
S.D.	1.500	5.53
Adverse weather conditions		
Rain	−0.144	−2.00
Snow and/or fog	−0.659	−3.80
Crash characteristics		
Driver ejected out of the vehicle	3.468	7.04
Vehicle rolled over	1.855	10.05
Crash with a stationary object (base is crash with another vehicle)		
Large object	1.509	11.34
Small object	1.201	8.76
Manner of collision in two vehicle crashes (base is rear-end collision)		
Head on	1.397	8.71
Angle	0.151	1.56
S.D.	1.066	5.41
Swipe collision when vehicles are traveling in opposite directions	−0.666	−1.96
Swipe collision when vehicles are traveling in same direction	−1.302	−9.31
Vehicle role in two vehicle crashes (base is driver strikes other vehicle)		
Driver struck by a vehicle	0.446	6.88
Driver involved in strike and struck	0.323	1.60
Seat belt	−0.752	−1.88
Standard deviation of common error component between seat belt use and injury severity propensities	0.926	3.10
Threshold parameters		
Threshold 1	2.366	3.89
Threshold 2	3.730	5.47
Threshold 3	5.218	6.99
Threshold 4	8.151	9.41
Log-likelihood at convergence		−10551.9

on average, more severe on high speed limit ($\geq$65 mph) roads relative to low speed limit ($\leq$25 mph) roads. These patterns may be a reflection of two issues. First, vehicle speeds are higher on the roads with high speed limits, because of which injury severity is higher on such roads. Second, there are several design factors associated with roads with very high speed limits ($\geq$65 mph) that can temper the seriousness of injuries from a crash. For instance, roads with high speed limits have wider lanes, more lanes, and wide shoulders that may present drivers the opportunity to take last minute evasive measures to reduce injury severity. It is also interesting, however, to note the wide variation in injury severity propensity across crashes on high speed roads (note the large standard deviation relative to the mean on the high speed limit coefficient in the table). This indicates that, while injury severity on high speed roads may be low for some crashes because of roadway characteristics, it can also be very high for some crashes because of the vehicle speed.

5.3.2.4. Environmental factors. The results associated with environmental factors indicate that crashes occurring during the day (6 a.m.–7 p.m.) tend to be less severe than those occurring during other times of the day. This may be because of higher traffic volumes on the roads during the day, resulting in drivers being more alert as well as traveling at slower speeds (note that this lower injury severity crashes during the day cannot be attributed to higher seat belt use, since seat belt use is included as a variable in the analysis; see Section 5.3.2.7).

The remaining parameters characterizing the effects of the crash environment in Table 4 suggest lower injury severity levels in dusk or dark lighting conditions (relative to dawn, daylight, and dark but lit lighting conditions) and under adverse weather conditions (relative to normal weather conditions). These results, while initially counter intuitive, are presumably capturing the vehicle speed effect. That is, drivers are likely to travel slower under poor lighting and adverse weather conditions. It is also important to note the large standard deviation on the "dark" variable, which indicates a wide dispersion in injury levels sustained under dark lighting conditions.

5.3.2.5. Crash characteristics. Several crash characteristics are strong determinants of injury severity. As expected, a driver ejected from her/his vehicle or in a vehicle that rolled over is likely to sustain severe injuries. The "crash with a stationary object" group of variables and the "manner of collision in two vehicle crashes" group of variables need to be considered together. The results indicate that, on average, crashes with a stationary large object (such as a concrete traffic barrier, post, pole, culvert, ditch, trees, etc.) and head-on collisions with another vehicle are most dangerous, followed by crashes with a stationary small object (such as a fire hydrant, shrubbery, boulder, curb, guard rail, etc.) and angle collisions with another vehicle (see O'Donnell and Connor, 1996 for similar results). Rear-end collisions with another vehicle are less severe than stationary object and head-on/angle collisions with another vehicle, but more severe than swipe collisions with another vehicle. There is also a large standard deviation of the "angle" coefficient; the

mean and the standard deviation of this coefficient imply that, in a majority of cases (88%), angle crashes are less severe than head-on collisions. But, about 12% of the time, angle crashes lead to higher injury severity than head-on crashes. Finally, in the set of crash characteristics, the "vehicle role in two vehicle crashes" group of variables suggests a higher injury severity level if the driver is struck, or is struck and strikes another vehicle, relative to striking another vehicle.

5.3.2.6. Seat belt use and sample selection effect. The empirical results showed no observed or unobserved differences in seat belt effectiveness across seat belt wearers in the injury severity equation for each of the IBO, IRBO and CRBO models (i.e., the elements of μ in Eq. (1) are not significantly different from zero and the variance of λ_q is also not significantly different from zero). Table 4 shows that the remaining constant impact of seat belt-use is negative, indicating a reduced injury severity if the driver uses the seat belt. Also, the standard deviation of the common error component between the seat belt use and injury severity propensities is statistically significant. This standard deviation corresponds to the standard deviation υ of η_q in Section 3.1. The term η_q was introduced as $-\eta_q$ in the injury severity component of Eq. (1) because it provided a substantially better fit than introducing it as $+\eta_q$. This result lends very strong support for the selective recruitment (or sample selection) hypothesis: safety-conscious drivers are more likely to wear seat belts and their defensive habits also lead to less severe injuries when they are involved in crashes. The standard deviation estimate of the common error component translates to an effective correlation of -0.46 [$=-(0.93)^2/1+(0.93)^2$] between the unobserved factors impacting the seat belt use and injury severity propensities.

In contrast to the correlated random binary-ordered (CRBO) model presented here, the independent random binary-ordered (IRBO) model ignores the selective recruitment issue; that is, it ignores the unobserved factors influencing seat belt use propensity and injury severity propensity. As a result, the safety-conscious and defensive driving habits of seat-belt users gets incorrectly manifested as an inflated effectiveness of seat belt use in reducing injury severity. In fact, the coefficient estimate on the seat belt use variable in the IRBO model is -1.39 with a corresponding *t*-statistic of -13.74 (compared to coefficient of -0.75 with a corresponding *t*-statistic of -1.88 in the CRBO model). While the coefficients between the IRBO and CRBO models are not comparable, the CRBO coefficients should be generally larger in magnitude compared to the IRBO coefficient (as is the case with all other coefficients except the seat belt use coefficient).[8] Thus, there is clear, substantial, and incorrect inflation in seat belt use effectiveness when sample selection is ignored. We revisit this important point again in Section 5.3.4.

---

[8] The injury severity component of the CRBO model is normalized with respect to a smaller overall scale relative to the injury severity component of the IRBO model (due to the additional presence of the term η_q in the CRBO model). This smaller scaling should, in general, lead to larger coefficients in the CRBO model.

5.3.2.7. Threshold parameters. The threshold parameters map the injury severity latent index to the reported injury severity categories. As such, they do not have any substantive interpretation.

5.3.3. Overall likelihood-based measures of fit

The log-likelihood value at convergence of the CRBO model (with 44 parameters) is -10551.9, of the IRBO model (with 43 parameters) is -10557.6, and of the IBO model (with 40 parameters) is -10570.6. The corresponding value for the "constants only" model with only the constant in the seat belt use binary choice model and only the four thresholds in the injury severity ordered logit model is -15054.5. Likelihood ratio tests may be undertaken to compare the four models above. In particular, the test for no sample selection (CRBO versus IRBO models) yields a likelihood ratio test value of 11.4 [$=-2 \times (10557.6 - 10551.9)$], which is larger than the chi-squared table value with 1d.f. at any reasonable level of significance (of course, this is also reflected in the statistically significant *t*-statistic on the standard deviation of the common error component between the seat belt use and injury severity equations). The test for the absence of unobserved heterogeneity in the effects of exogenous variables (IRBO versus IBO) yields a likelihood ratio test value of 26, which is again larger than the critical chi-squared value with 3d.f. at even the 0.0001 level of significance.

Clearly the results indicate the importance of considering randomness in the effects of injury severity determinants due to the moderating influence of unobserved factors as well as accommodating the endogeneity of seat belt use on injury severity. Failure to accommodate these issues, as done by almost all earlier injury severity studies, will, in general, lead to poor model fits as well as biased parameter estimates.

5.3.4. Elasticity effects

The parameters on the exogenous variables in Table 4 do not directly provide the magnitude of the effects of variables on the probability of each level of injury severity. To do so, we compute the aggregate level "elasticity effects" of variables. This is achieved by first computing the probability of seat belt non-usage ($s_q = 0$) and injury severity level k ($y_q = k$) for individual q as:

$$P(s_q = 0, y_q = k) = \int_{c_q} \{G[-\{(\beta' + \gamma'_q)x_q + \eta_q\}]$$

$$\times \{G[\psi_k - \{(\alpha' + \delta'_q)z_q - \eta_q\}] - G[\psi_{k-1} - \{(\alpha' + \delta'_q)z_q - \eta_q\}]\}^{d_{qk}}\} \, \mathrm{d}F(c_q) \tag{7}$$

The corresponding probability of seat belt usage ($s_q = 0$) and injury severity level k ($y_q = k$) is computed as:

$$P(s_q = 1, y_q = k) = \int_{c_q} \{G[-\{(\beta' + \gamma'_q)x_q + \eta_q\}]$$

$$\times \{G[\psi_k - \{(\alpha' + \delta'_q)z_q + (\theta + \mu'w_q + \lambda_q) - \eta_q\}] - G[\psi_{k-1} - \{(\alpha' + \delta'_q)z_q + (\theta + \mu'w_q + \lambda_q) - \eta_q\}]\}^{d_{qk}}\} \, \mathrm{d}F(c_q) \tag{8}$$

Next, the unconditional probability that individual q sustains an injury of severity level k is obtained as

$$P(y_q = k) = P(s_q = 0, y_q = k) + P(s_q = 1, y_q = k) \tag{9}$$

The expected aggregate numbers of drivers sustaining an injury of severity level k is then computed by summing the above individual-level probability across all individuals Q.

With the preliminaries above, one can compute the aggregate-level "elasticity" of any dummy exogenous variable (all exogenous variables in the model are dummy variables) by changing the value of the variable to one for the subsample of observations for which the variable takes a value of zero and to zero for the subsample of observations for which the variable takes a value of one. We then sum the shifts in expected aggregate shares in the two subsamples after reversing the sign of the shifts in the second subsample, and compute an effective percentage change in expected aggregate shares in the entire sample due to change in the dummy variable from 0 to 1.

The elasticity effects are presented in Table 5 by variable category and for each of the IBO, IRBO, and CRBO models (note that the expressions in Eqs. (7) and (8) simplify in the case of the IBO and IRBO models). For ease in presentation, we provide the elasticities only for the fatal injury category. The table also presents only the effects of the non-interaction variables from Table 4 because the effect of an interaction variable is accommodated by increasing the variable whenever a component variable is increased. The numbers in the table may be interpreted as the percentage change in the probability of a fatal injury due to a change in the variable from 0 to 1. For instance, the CRBO model in the table indicates that the probability of a man being fatally injured in a crash is about 40% less than the probability of a woman being fatally injured, other characteristics being equal.

Several important observations may be made from Table 5. *First*, the major factors that are likely to lead to a fatal injury in a crash are driver ejection from vehicle, vehicle rollover, and crash into a stationary object or a head-on collision with another vehicle. On the other hand, seat belt use and a swipe collision with another vehicle traveling in the same direction are the two most important factors associated with survival in a crash. *Second*, ignoring the moderating effect of unobserved variables on the impact of factors on injury severity can lead to severely biased elasticity effects. For instance, ignoring unobserved heterogeneity leads to an overestimation of the impact of crashes on high speed limit roads by 200% (see the difference between the IBO model and the IRBO/CRBO models). Similar substantial inaccurate projection of dark lighting and angle collision are observed. Third, the elasticity effects of many variables are quite different among the CRBO model (that considers seat belt endogeneity and the other two models (the IBO and IRBO models).

Table 5
Elasticity effects for the fatal injury category

Variables	IBO	IRBO	CRBO
Driver characteristics			
Male	−36.55	−36.23	−39.33
Age variables			
25–74 years	13.13	14.03	15.05
>74 years	26.57	27.39	31.17
Under the influence of alcohol	18.55	19.67	38.86
Vehicle characteristics			
Sedan	19.83	20.47	23.27
Non-sedan (other vehicle type)	0.95	1.11	1.36
Roadway design attributes			
Medium-to-high speed limit (26–64 mph)	49.52	49.29	53.73
High speed limit (≥64 mph)	74.50	21.93	29.78
Environmental factors			
6 a.m.–7 p.m.	−19.75	−19.60	−23.40
Lighting conditions			
Dusk	−16.15	−14.94	−16.22
Dark	−1.43	−25.98	−27.76
Adverse weather conditions			
Rain	−8.76	−9.11	−10.29
Snow and/or fog	−11.38	−12.17	−13.38
Crash attributes			
Driver ejected out of the vehicle	808.90	871.37	1054.25
Vehicle rolled over	186.00	189.02	226.74
Crash with a stationary object (base is crash with another vehicle)			
Large object	120.29	119.62	147.09
Small object	107.71	101.98	127.19
Manner of collision in two vehicle crashes (base is rear-end collision)			
Head on	133.97	132.49	167.92
Angle	29.63	6.31	11.41
Swipe collision when vehicles are traveling in opposite directions	−40.12	−37.55	−39.57
Swipe collision when vehicle are traveling in same direction	−62.75	−60.50	−63.95
Vehicle role in two vehicle crashes (base is driver strikes other vehicle)			
Driver struck by a vehicle	32.05	33.13	37.20
Driver involved in strike and struck	15.00	25.94	27.28
Seat belt use	−129.45	−132.06	−64.50

For example, the likelihood of being in a fatal injury if under the influence of alcohol is underestimated in the IBO and IRBO models by 50%. Similarly, the positive effects of ejection from the vehicle, vehicle rollover, and stationary object/head-on collision with another vehicle on fatal injury are underestimated in the IBO and IRBO models by 20, 13, and 25%, respectively. *Fourth*, the elasticity effect of seat belt use from the CRBO model is about half that of the estimated effects from the IBO and IRBO models. This is, of course, because the IBO and IRBO model do not consider the endogenous nature of seat belt use. In fact, the seat belt use elasticities from the different models suggest that seat belt usage and the safety-conscious driving attitudes of those who wear seat belts are about equally important in reducing the likelihood of a fatal injury. This result is important from a policy standpoint and suggests that seat belt non-users, when apprehended in the act, should perhaps be subjected to both a

fine (to increase the chances that they wear seat belts) as well as mandatory enrollment in a defensive driving course (to attempt to change their aggressive driving behaviors). Thus, the results in our research provide support for changing the current "Click it or Ticket" campaign in several states in the US to the "Click it or Defensive Driving and Ticket" campaign.

Overall, the results indicate clear biases in the effects of variables on injury severity level when unobserved factors moderating the impact of variables is ignored and/or seat belt endogeneity is not considered.

6. Conclusions

This paper formulates a comprehensive econometric structure that recognizes two important issues in safety analysis. First, the impact of a factor on injury severity may be

moderated by various observed and unobserved variables specific to an individual or to a crash. Second, seat belt use is likely to be endogenous to injury severity. That is, it is possible that intrinsically unsafe drivers do not wear seat belts and are the ones likely to be involved in high injury severity crashes because of their unsafe driving habits. The structure of the model developed in the paper takes the form of a mixed joint binary logit-ordered response logit formulation that conveniently, and at once, considers all the issues of (1) systematic interaction effects among variables, (2) random unobserved effects in the influence of injury severity determinants, (3) potential endogeneity of seat belt use in modeling injury severity levels, and (4) random variations in seat belt use effectiveness. To our knowledge, this is the first instance of such a model formulation and application not only in the safety analysis literature, but in the econometrics literature in general.

The empirical analysis is based on the 2003 General Estimates System (GES) data base. The focus in the analysis is exclusively on non-commercial driver seat belt use and crash-related injury. The analysis is also confined to the vast majority of crashes in which one or two vehicles are involved. Several types of variables are considered in the empirical analysis, including driver characteristics, vehicle characteristics, roadway design attributes, environmental factors, and crash characteristics.

The empirical results indicate the important effects of all of the above types of variables on driver seat belt use and injury severity. In addition, the results reveal a substantial and significant negative error correlation between seat belt use propensity and injury severity propensity, which lends strong support for the selective recruitment (or sample selection) hypothesis. That is, safety-conscious drivers are more likely to wear seat belts, and their defensive habits also lead to less severe injuries when they are involved in crashes.

To summarize, ignoring the moderating impact of unobserved factors on the influence of injury severity determinants and/or the endogeneity of seat belt use in injury severity modeling leads to biased parameter estimates and elasticity effects. With respect to seat belt use specifically, our results suggest that seat belt usage and the safety-conscious driving attitudes of those who wear seat belts are about equally important in reducing the likelihood of a fatal injury (earlier research efforts do not disentangle these two different aspects of seat belt usage). Thus, from a policy standpoint, seat belt non-users should perhaps be subjected to both a fine (to increase the chances that they wear seat belts) as well as mandatory enrollment in a defensive driving course (to attempt to change their aggressive driving behaviors).

Acknowledgments

References

Abdel-Aty, M.A., Chen, C.L., Schott, J.R., 1998. An assessment of the effect of driver age on traffic accident involvement using log-linear models. Accid. Anal. Prev. 30 (6), 851–861.

Abdel-Aty, M.A., Abdelwahab, H., 2004. Modeling rear-end collisions including the role of driver's visibility and light truck vehicles using a nested logit structure. Accid. Anal. Prev. 36 (3), 447–456.

Al-Ghamdi, S.A., 2002. Using logistic regression to estimate the influence of accident factors on accident severity. Accid. Anal. Prev. 34 (6), 724–741.

Bedard, M., Guyatt, G.H., Stones, M.J., Hirdes, J.P., 2002. The independent contribution of driver, crash, and vehicle characteristics to driver fatalities. Accid. Anal. Prev. 34 (6), 717–727.

Bhat, C.R., 1991. Toward a model of activity program generation. PhD dissertation. Department of Civil and Environmental Engineering, Northwestern University.

Bhat, C.R., 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. Transport. Res. Part B 35, 677–693.

Bhat, C.R., Koppelman, F.S., 1993. An endogenous switching simultaneous equation system of employment, income and car ownership. Transport. Res. Part A 27, 447–459.

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

Chamberlain, G., 1980. Analysis of covariance with qualitative data. Rev. Econ. Stud. 47, 225–238.

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

Cohen, A., Einav, L., 2003. The effects of mandatory seat belt laws on driving behavior and traffic fatalities. Rev. Econ. Stat. 85 (4), 828–843.

Cooper, P.J., 1994. Estimating over involvement of seat belt nonwearers in crashes and the effect of lap/shoulder restraint use on different crash severity consequences. Accid. Anal. Prev. 26, 263–275.

Dee, T.S., 1997. Reconsidering the effects of seat belts laws and their enforcement status. Accid. Anal. Prev. 30 (1), 1–10.

Derrig, R.A., Segui-Gomez, M., Abtahi, A., Liu, L.L., 2000. The effect of population safety belt usage rates on motor vehicle-related fatalities. Accid. Anal. Prev. 34 (1), 101–110.

Dissanayake, S., Lu, J.J., 2002. Factors influential in making an injury severity difference to older drivers involved in fixed object-passenger car crashes. Accid. Anal. Prev. 34 (5), 609–618.

Doherty, S.T., Andrey, J.C., MacGregor, C., 1998. The situational risks of young drivers: the influence of passengers, time of day and day of week on accident rates. Accid. Anal. Prev. 30 (1), 45–52.

Evans, L., 1985. Fatality risk for belted drivers versus car mass. Accid. Anal. Prev. 17 (3), 251–271.

Evans, L., 1990. The fraction of traffic fatalities attributable to alcohol. Accid. Anal. Prev. 22 (6), 587–602.

Evans, L., 1996. Safety-belt effectiveness: the influence of crash severity and selective recruitment. Accid. Anal. Prev. 28 (4), 423–433.

Evans, L., Frick, M.C., 1988. Seating position in cars and fatality risk. Am. J. Pub. Health 78 (11), 1456–1458.

Farmer, C.M., Braver, E.R., Mitter, E.R., 1996. Two-vehicle side impact crashes: the relationship of vehicle and crash characteristics to injury severity. Accid. Anal. Prev. 29 (3), 399–406.

Golob, T.F., Recker, W.W., Leonard, J.D., 1986. An analysis of the severity and incident duration of truck-involved freeway accidents. Accid. Anal. Prev. 19 (5), 375–395.

Hajivassiliou, V.A., Ruud, P.A., 1994. Classical estimation methods for LDV models using simulations. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, vol. IV. Elsevier, New York, pp. 2383–2441.

Huelke, D.F., Compton, C.P., 1995. Effects of seat belts on injury severity of front and rear seat occupants in the same frontal crash. Accid. Anal. Prev. 27 (6), 835–838.

Janssen, W., 1994. Seat-belt wearing and driving behavior: an instrumented vehicle study. Accid. Anal. Prev. 26 (2), 249–261.

Kassoff, H., Deutschman, H.D., 1969. Trip generation: a critical appraisal. Highway Res. Rec. 297, 15–30.

Kim, K., Nitz, L., Richardson, J., Li, L., 1994. Personal and behavioral predictors of automobile crash and injury severity. Accid. Anal. Prev. 27 (4), 469–481.

Kitamura, R., Bunch, D.S., 1990. Heterogeneity and state dependence in household car ownership: a panel analysis using ordered-response probit models with error components. In: Koshi, M. (Ed.), Transportation and Traffic Theory, pp. 477–496.

Khattak, A.J., Rocha, M., 2003. Are SUVs 'Supremely Unsafe Vehicles'? Analysis of rollovers and injuries with sport utility vehicles. Transport. Res. Rec. 1840, 167–177.

Khattak, A.J., Kantor, P., Council, F.M., 1998. Role of adverse weather in key crash types on limited-access roadways. Transport. Res. Rec. 1621, 10–19.

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

Krull, K.A., Khattak, A.J., Council, F.M., 2000. Injury effects of rollovers and events sequence in single-vehicle crashes. Transport. Res. Rec. 1717, 46–54.

Kockelman, K.M., Kweon, Y.J., 2001. Driver injury severity: an application of ordered probit models. Accid. Anal. Prev. 34 (3), 313–321.

Kweon, Y.J., Kockelman, K.M., 2002. Overall injury risk to different drivers: combining exposure, frequency, and severity models. Accid. Anal. Prev. 35 (4), 441–450.

Lee, L.-F., 1992. On the efficiency of methods of simulated moments and maximum simulated likelihood estimation of discrete response models. Econometr. Theor. 8, 518–552.

Lourens, P.F., Vissers, J.A.M.M., Jessurun, M., 1999. Annual mileage, driving violations and accident involvement in relation to drivers' sex, age and level of education. Accid. Anal. Prev. 31 (1), 593–597.

McLellan, B.A., Rizoli, B.S., Brenneman, D.F., Boulanger, R.B., Sharkey, W.P., Szalai, P.J., 1996. Injury pattern and severity in lateral motor vehicle collisions: a Canadian experience. J. Trauma 41 (4), 708–713.

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

Petridou, E., Moustaki, M., 2000. Human factors in the causation of road traffic crashes. Eur. J. Epidemiol. 16, 819–826.

Preusser, D.F., Williams, A.F., Lund, A.K., 1991. Characteristics of belted and unbelted drivers. Accid. Anal. Prev. 23 (6), 475–482.

Reinfurt, D., Williams, A., Wells, J., Rodgman, E., 1996. Characteristics of drivers not using seat belts in a high belt use state. J. Safety Res. 27 (4), 209–215.

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

Schiff, M.A., Cummings, P., 2004. Comparison of reporting of seat belt use by police and crash investigators: variation in agreement by injury severity. Accid. Anal. Prev. 36 (6), 961–965.

Shibata, A., Fukuda, K., 1993. Risk factors of fatality in motor vehicle traffic accidents. Accid. Anal. Prev. 26 (3), 391–397.

Shinar, D., 1993. Demographic and socio-economic correlates of safety belt use. Accid. Anal. Prev. 35 (6), 745–755.

Srinivasan, K.K., 2002. Injury severity analysis with variable and correlated thresholds: ordered mixed logit formulation. Transport. Res. Rec. 1784, 132–142.

Toy, E.L., Hammitt, J.K., 2003. Safety impacts of SUVs, vans, and pickup trucks in two vehicle crashes. Risk Anal. 23 (4), 641–650.

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

Wang, X., Kockelman, K.M., 2005. Use of Heteroscedastic Ordered Logit Model to Study Severity of Occupant Injury: Distinguishing Effects of Vehicle Weight and Type. Transportation Research Record 1908, pp. 195–204.