Transportation Research Part B 49 (2013) 39–54

Contents lists available at SciVerse ScienceDirect

Transportation Research Part B

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

The heterogeneous effects of guardian supervision on adolescent driver-injury severities: A finite-mixture random-parameters approach

Yingge Xiong[a,1], Fred L. Mannering[b,*]

[a] Purdue University, School of Civil Engineering, 550 Stadium Mall Drive, West Lafayette, IN 47907-2051, United States
[b] Purdue University, School of Civil Engineering and the Center for Road Safety, 550 Stadium Mall Drive, West Lafayette, IN 47907-2051, United States

ARTICLE INFO

Article history:
Received 5 January 2013
Received in revised form 5 January 2013
Accepted 5 January 2013

Keywords:
Adolescent crashes
Graduated driver licensing
Heterogeneity model
Gaussian mixture
Bayesian inference
Permutation sampler

ABSTRACT

One of the key aspects of graduated driver licensing programs is the new-driver experience gained in the presence of a guardian (a person providing mandatory supervision from the passenger seat). However, the effect that this guardian-supervising practice has on adolescent drivers' crash-injury severity (should a crash occur) is not well understood. This paper seeks to provide insights into the injury-prevention effectiveness of guardian supervision by developing an appropriate econometric structure to account for the complex interactions that are likely to occur in the study of the heterogeneous effects of guardian supervision on crash-injury severities. As opposed to conventional heterogeneity models with standard distributional assumptions, this paper deals with the heterogeneous effects by accounting for the possible multivariate characteristics of parameter distributions in addition to allowing for multimodality, skewness and kurtosis. A Markov Chain Monte Carlo (MCMC) algorithm is developed for estimation and the permutation sampler proposed by Frühwirth-Schnatter (2001) is extended for model identification. The econometric analysis shows the presence of two distinct driving environments (defined by roadway geometric and traffic conditions). Model estimation results show that, in both of these driving environments, the presence of guardian supervision reduces the crash-injury severity, but in interestingly different ways. Based on the findings of this research, a case could easily be made for extending the time-requirement for guardian supervision in current graduated driver license programs.

1. Introduction

Motor vehicle crashes have long been recognized as a leading cause of death for adolescents in the United States and throughout the motorized world. For example, from 2004 to 2008 a report conducted for the Centers for Disease Control and Prevention, based on the Fatality Analysis Report System data, determined that a total of 9644 passenger-vehicle drivers aged 16 or 17 years were involved in fatal crashes during 2004–2008 (Shults, 2010). To better understand contributing factors of crashes involving adolescent drivers, numerous studies have been carried out in the context of psychology (Halpern-Felsher et al., 2001; Ericsson, 2005), health science (Williams et al., 1995; Millstein and Halpern-Felsher, 2002), medical science (Allen and Brown, 2008) and general transportation (Doherty et al., 1998; Bianchi and Summala, 2004). Factors including driver inexperience (Insurance Institute for Highway Safety, 2009), risk taking (Beullens and Van den Bulck, 2008),

* Corresponding author. Tel.: +1 765 496 7913; fax: +1 765 494 0395.
E-mail addresses: xiong0@purdue.edu (Y. Xiong), flm@purdue.edu (F.L. Mannering).
[1] Tel.: +1 765 430 7898; fax: +1 765 494 0395.

0191-2615/$ - see front matter
http://dx.doi.org/10.1016/j.trb.2013.01.002

distracted driving (National Highway Traffic Safety Administration, 2008a), nighttime driving (Centers for Disease Control and Prevention, 2009), drunk driving (Shope et al., 1996; Dols et al., 2010), cell-phone use while driving (Hafetz et al., 2010), safety-belt use (National Highway Traffic Safety Administration, 2008b), presence of young passengers (Preusser et al., 1998; Chen et al., 2000; Simons-Morton et al., 2005) and drivers involved non-driving delinquent behavior (Buckley et al., 2012) have all been identified as elements contributing to improper teenage-driving habits and severe crash outcomes.

To improve adolescent-driving safety, graduated driver licensing (GDL) programs have gained widespread acceptance in the US and elsewhere. In addition to placing restrictions on risky driving behaviors (underage drinking, talking or texting on a cellular phone while driving), such programs typically include requirements for the presence of an experienced licensed driver (for some period of time) to help safeguard adolescent drivers from crashes and resulting injuries. The idea is that the presence of such guardian supervision can help reinforce proper driving behavior and enhance adolescent driver skills both technically and psychologically by delivering useful information relating to the detection of hazards, estimating the crash risks, and providing real-time guidance to avoid identified crash threats.

With regard to baseline driving tendencies of adolescent drivers, research indicates that teenagers tend to have initial driving habits that closely mirror those of their parents due to genetic and environmental similarities, as well as childhood parent to offspring learning (Ferguson et al., 2001; Bianchi and Summala, 2004; Wilson et al., 2006). Parents further enforce these tendencies in their guardian role. Despite the variance that one may expect among the population of parents in terms of their driving habits and the amount of information they convey in a guardian role, there is a general consensus that guardian supervision (often parental) is a beneficial component of graduated driver licensing programs (Williams et al., 2002; Simons-Morton, 2007). However, there is at least some evidence to the contrary. For example, the American Automobile Association (2012) reports that recent naturalistic driving studies have shown that some adolescents can become nervous and hypersensitive to guardian feedback, particularly when they first begin driving. Despite the general consensus that guardian supervision is beneficial in improving adolescent driving safety, strong empirical evidence to support this consensus is missing in the extant literature. There clearly exists a need to better understand the effect that guardian supervision has on adolescent driver behavior and performance. The intent of this paper is to carefully investigate the effects of guardian supervision on the crash-injury outcomes of adolescent drivers.

2. Elements of guardian supervision

Intuitively, the occurrence of a vehicle crash can be decomposed into three stages: (1) driver assessment of stimuli from the driving environment and the possible initiation of action (such as braking and steering to avoid potentially hazardous situations); (2) the vehicle response to driver actions which will be influenced by roadway and pavement characteristics, car attributes, etc.; (3) vehicle crash and resulting damage (assuming avoidance maneuvers fail to completely avoid the crash). The eventual injury outcome of the crash is obviously strongly influenced by stages 1 and 2 above, because proper driver actions can greatly reduce the impact forces even if such actions do not allow the driver to avoid a crash. In this three-step decomposition, the guardian's primary influence will be in stage 1.

Adolescent driver assessment of stimuli from the driving environment and possible actions are likely to be determined by a number of factors including driver attributes, driving-environment conditions and guardian attributes. Driver attributes such as age, gender, education, occupation, and health conditions are all factors that are correlated with driving skills, risk assessment, and habits. Driving-environment variables, including traffic conditions, weather conditions, and roadway geometrics, are the primary factors motivating driver actions and stimulating guardian input. Driving-environment variables play both a direct (through driver response to stimuli) and an indirect role in influencing adolescent driving behavior, with the indirect role coming from guardian input. There is also the very real possibility that discrete driving environments may exist where the combination of roadway geometric and traffic conditions on specific roadway segments create distinct decision-making environments (for example, those roadway segments where geometric/traffic conditions are not considered unusually complex and others where they may be considered complex). For guardian attributes, in addition to their direct input on adolescent drivers from guardian responses to driving-environment variables, the mere presence of a guardian has an indirect influence on adolescent driving behavior through factors such as confidence, psychological comfort, and anxiety levels. Thus the net effect of the driving environment and the guardian is quite complex, and this effect is further complicated by the fact that adolescent-driver interaction may vary considerably across driving environments and the various driver/guardian personalities.

The complexity of the interactive effects in this decision-making environment can lead to incorrect inferences if inappropriate econometric methods are used. In this paper we seek to study the heterogeneous effects of guardian supervision on crash-injury severity by considering a sample of adolescent-driver crashes (crashes with drivers aged 16–19). To capture the potentially complex adolescent/guardian interactions, a finite-mixture random-parameters framework is developed to account for the possibility of discrete partitioning among roadway segments (discrete distinct driving environments among groups of roadway segments) as well as the interactive effects of guardian supervision and environmental factors on adolescent driver behavior. This framework allows for driving environment-level (component-level) as well as within driving environment-level (within component-level/roadway-segment-level) unobserved heterogeneity. As will be discussed below, Bayesian inference is used to estimate the proposed model and, to identify the roadway-segment categories, a modification of Frühwirth-Schnatter (2001)'s random-permutation sampler is used to implement conditional posterior draws.

3. Methodological background

Random parameters and finite mixture (latent class) methods have been two dominant approaches for characterizing unobserved heterogeneity. Random parameters models can accommodate individual unobserved heterogeneity by allowing parameters to differ across observations – implying varying influences of observed explanatory variables on response outcomes (Hensher and Greene, 2003; Wong et al., 2008). In contrast, finite mixture models adopt a semiparametric way to assume the sampled observations arise from a distinct subgroups of data with homogeneous features. To this end, the finite mixture approach replaces the continuous distribution assumed by random parameters model with a discrete distribution that is represented by a finite and specified number of mass points. This potentially frees the analyst from possibly unwarranted distributional assumptions on parameters (Depaire et al., 2008; Olaru et al., 2011; Eluru et al., 2012).

However, in many instances neither the finite mass points approximation of the finite mixture model nor the specified continuous distributions of the random parameters model can adequately account for the unobserved heterogeneity. For example, the finite mixture model does not have an obvious way of handling potentially substantial within-group variation due to the restrictive homogeneity assumption on characteristics of the within-group population. And, the continuous distributions assumed in the random parameters approach may not be able to adequately characterize the presence of group-specific features within the population. A comparison of mixed logit (random parameters) and latent class logit (finite mixture) model has been conducted by Shen (2009) and Greene and Hensher (2003), but statistically conclusive evidence as to which approach is superior is elusive and likely to vary from one database to the next.

There have been research efforts in statistics (Verbeke and Lesaffre, 1996), econometrics (Lenk and DeSarbo, 2000) and marketing research (Allenby et al., 1998) that undertake an alternate formulation that incorporates a random parameters structure with a finite mixture model. The hybrid modeling idea considers the possibility of individual-layer random parameters sampled from an assumed continuous distribution within each of the groups within a finite mixture framework. Hence it cannot only account for group-specific heterogeneity in the population but also allow for the individual level heterogeneity within each group. Verbeke and Lesaffre (1996) introduced a linear heterogeneity model that takes into account the presence of subgroups and provides estimates for the within-group parameters. Using Bayesian methods, Allenby et al. (1998) and Lenk and DeSarbo (2000) developed a finite mixture with random effects approach for multinomial logit and generalized linear models.

There have been an abundance of other applications of this general approach as well. For example, Yau et al. (2003) used a two-component mixture regression model with random effects to consider heterogeneity and dependency among observations simultaneously via an Expectation–Maximization algorithm. Bujosa et al. (2010) applied this model in recreational demand to identify different behavior groups as a function of socioeconomic characteristics and to account for taste diversity among individuals in the same group. In other work, Park et al. (2010) used a univariate Gaussian mixture distribution to accommodate the multimodality, skewness and excess kurtosis in traffic speed data. Finally, Greene and Hensher (2013) provide some empirical evidence showing that a latent class mixed multinomial logit model can outperform standard mixed logit and fixed latent class models. However, the true multivariate characteristics of random parameter distributions have rarely been addressed in the literature.

In terms of estimation of random parameters and finite mixture models, simulation-based approaches such as the Method of Simulated Moments (McFadden, 1989; Pakes and Pollard, 1989) and Maximum Simulated Likelihood estimation, facilitated by the Geweke–Hajivassiliou-Keane simulator (Börsch-Supan and Hajivassiliou, 1993) and Halton draws (Bhat, 2001, 2003), have been widely used. Motivated by the large computational requirements needed to ensure asymptotic Maximum Simulated Likelihood estimator properties, Bhat (2011) recently proposed a Maximum Approximate Composite Marginal Likelihood, which is a simulation-free estimation method, and successfully applied it in estimating mixed multinomial probit models with normal (Bhat and Sidharthan, 2011) and skewed-normal (Bhat and Sidharthan, 2012) assumptions for random parameters. Because the Maximum Approximate Composite Marginal Likelihood is an analytic approximation method, score vectors and the Hessian matrix of the heterogeneity model can be easily determined and parameter values and large-sample standard errors can be accurately and quickly estimated – resulting in considerable computational gains relative to standard simulation methods.

As an alternative to classical estimation approaches, Bayesian procedures based on Markov Chain Monte Carlo (MCMC) simulation have been recognized as an alternative way to estimate complex discrete-outcome model structures. The approach is often able to circumvent numerically cumbersome likelihood function maximizations (Huber and Train, 2001). Since the earliest use of Bayesian methods to estimate probit models (Albert and Chib, 1993; McCulloch and Rossi, 1994), Bayesian methods have seen numerous applications and extensions in the context of discrete-outcome heterogeneity model estimation. For example, Allenby and Lenk (1994), Allenby (1997) and Allenby and Rossi (1999) have applied Bayesian techniques to estimate mixed logit models with normally distributed parameters. Estimations for non-normal random parameters in mixed logit models have been developed by Huber and Train (2001) through Bayesian inference.

However, Bayesian methods based on Markov Chain Monte Carlo (MCMC) simulation can present convergence difficulties. Although a growing number of methods like reparameterization (Nandram and Chen, 1996), blocking and collapsing (Chib and Carlin, 1999), and parameter expansion (Liu and Wu, 1999) can improve the computational efficiency of Gibbs sampling and the Metropolis-Hasting algorithm, the slow mixing of samplers can still be an issue in heterogeneity models when random effects are distributed with specific mean and variance combinations (Frühwirth-Schnatter et al., 2004).

Furthermore, the identification problem can be a concern in the estimation of finite mixture models (Stephens, 2000; Frühwirth-Schnatter, 2001). A feasible and efficient algorithm is needed in many cases when dealing with more sophisticated heterogeneity models.

4. Proposed finite-mixture random-parameter binary probit model

To begin, define $\mathbf{Y} = \{y_{iq}\}$ as a vector of binary observed driver crash-injury severity outcomes (as with some previous research, such as Winston et al., 2006, consideration will be given to two crash-severity outcomes: no-injury and injury), of adolescent driver $q(q = 1, \ldots, Q_i)$ on the roadway segment $i(i = 1, \ldots, n)$. When the severity outcome is an injury, $y_{iq} = 1$, otherwise $y_{iq} = 0$. With respect to each $y_{iq}, \mathbf{X}_{iq}^{Fixed} = \left[x_{iq1}^{Fixed}\ x_{iq2}^{Fixed}\ \ldots\ x_{iqm}^{Fixed}\right]$ is correspondingly a $1 \times m$ vector of contributing factors x_{iqj}^{Fixed} $(j = 1, 2, \ldots, m)$ with fixed parameter vector $\boldsymbol{\alpha} = [\alpha_1\ \alpha_2\ \ldots\ \alpha_m]'$ across observations. The guardian indicator x_{iq1}^{Random} and some of the driving environmental factors x_{iqj}^{Random} $(j = 2, \ldots, k)$ are defined as a $1 \times k$ vector of random variables $\mathbf{X}_{iq}^{Random} = \left[x_{iq1}^{Random}\ x_{iq2}^{Random}\ \ldots\ x_{iqk}^{Random}\right]'$ with a randomly distributed parameter vector $\boldsymbol{\beta}_i = [\beta_{i1}\ \beta_{i2} \ldots \beta_{ik}]'$ which can vary across the roadway segment population.

With these definitions, the univariate representation of the dichotomous probit model is written as:

$$
\begin{aligned}
z_{iq} &= \mathbf{X}_{iq}^{Fixed}\boldsymbol{\alpha} + \mathbf{X}_{iq}^{Random}\boldsymbol{\beta}_i + \varepsilon_{iq}, \quad \varepsilon_{iq} \overset{iid}{\sim} N(0,1) \\
y_{iq} &= \begin{cases} 0, & \text{if } z_{iq} \leqslant 0 \\ 1, & \text{if } z_{iq} > 0 \end{cases}
\end{aligned} \tag{1}
$$

where z_{iq} is a unobserved latent variable determined by a regression specification. Because of the potential identification problem of parameters in the latent specification, the idiosyncratic error term ε_{tq} is designed to be identically and independently standard normal distributed (Koop et al., 2007), and strictly orthogonal to either covariates $\mathbf{X}_{iq}^{Fixed}$ or $\mathbf{X}_{iq}^{Random}$. Specifically, the random parameter vector $\boldsymbol{\beta}_i$ is given as finite mixture of G Gaussian distributions:

$$
\boldsymbol{\beta}_i \sim \sum_{g=1}^{G} \pi_g N_g(\boldsymbol{\mu}_{\beta g}, \mathbf{V}_{\beta g}), \quad \sum_{g=1}^{G} \pi_g = 1 \tag{2}
$$

with

$$
\boldsymbol{\mu}_{\beta g} = \begin{bmatrix} \mu_{\beta_1 g} \\ \mu_{\beta_2 g} \\ \cdot \\ \mu_{\beta_k g} \end{bmatrix} \quad \text{and } \mathbf{V}_{\beta g} = \begin{bmatrix} (\sigma_1^g)^2 & \sigma_{1,2}^g & \cdots & \sigma_{1,k-1}^g & \sigma_{1,k}^g \\ \sigma_{2,1}^g & (\sigma_2^g)^2 & \cdots & \sigma_{2,k-1}^g & \sigma_{2,k}^g \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{k-1,1}^g & \sigma_{k-1,2}^g & \cdots & (\sigma_{k-1}^g)^2 & \sigma_{k-1,k}^g \\ \sigma_{k,1}^g & \sigma_{k,2}^g & \cdots & \sigma_{k,k-1}^g & (\sigma_k^g)^2 \end{bmatrix}
$$

where $\pi = [\pi_1\ \pi_2\ \ldots\ \pi_g]'$ is defined as a $g \times 1$ vector of component probabilities.

This finite-mixture random-parameter formulation defines G distinct components with homogenous parameters following identical distributions within each component yet different distributions across components. Rather than simply assuming arbitrary continuous distributions (as is often done in random parameter methods) or mass point approximations (finite mixture methods) for parameters, the proposed formulation can allow for skewness, multimodality and heavy tails in outcome distributions (Geweke and Keane, 2007; Li and Tobias, 2007). This approach is able to account for component-specific unobserved heterogeneity in the population and also allow for the individual-level unobserved heterogeneity within each component (for details please see Xiong et al., 2013). Instead of imposing an independent and identically distribution assumption for each random parameter, the random parameter vector within each component follows a multivariate distribution (the multivariate normal distribution) with mean vector $\boldsymbol{\mu}_{\beta g}$ and an unrestricted variance–covariance matrix $\mathbf{V}_{\beta g}$. To this end, this matrix is of a heteroscedastic nature with the off-diagonal elements not necessarily being zero. In the case of this study, this structure is used to (1) identify a discrete partitioning of the adolescent drivers which is determined by unobserved personal attributes and driving skills; (2) capture the heterogeneous effects of environmental factors and the guardian indicator on driving behaviors of each sub-component population; and (3) analyze the effects of the guardian indicator x_{1q}^{Random} on other crash-related explanatory variables $x_{p(p\neq 1)q}^{Random}$ (the elements in both the first row and first column of the covariance matrix are of interests and expected to be non-zero).

To facilitate computational estimation, a $1 \times i$ vector of latent component indicators is defined as:

$$
S_i = [S_{1i}\ S_{2i}\ \ldots\ S_{Gi}] \tag{3}
$$

If ith roadway segment belongs to gth component, $S_{gi} = 1$ and otherwise $S_{gi} = 0$. So that the unknown distribution of heterogeneity of $\boldsymbol{\beta}_i$ can be written as:

$$\boldsymbol{\beta}_i = \begin{cases} \boldsymbol{\beta}_{(1)i} \sim N_1(\boldsymbol{\mu}_{\beta 1}, \mathbf{V}_{\beta 1}), & \text{if } S_{1i} = 1 \\ \vdots \\ \boldsymbol{\beta}_{(g)i} \sim N_g(\boldsymbol{\mu}_{\beta g}, \mathbf{V}_{\beta g}), & \text{if } S_{gi} = 1 \\ \vdots \\ \boldsymbol{\beta}_{(G)i} \sim N_G(\boldsymbol{\mu}_{\beta G}, \mathbf{V}_{\beta G}), & \text{if } S_{Gi} = 1 \end{cases} \tag{4}$$

where label indicator $g = 1, 2, \ldots, G$. With the independence assumption of the occurrence of observed adolescent accidents, the corresponding likelihood function can be written as:

$$L(\alpha,\beta) = \prod_{i=1}^{n}\prod_{q=1}^{Q_i}\left\{\begin{array}{l} \left[\Phi\left(\mathbf{X}_{iq}^{Fixed}\boldsymbol{\alpha} + \mathbf{X}_{iq}^{Random}\boldsymbol{\beta}_{(g)i}\right)^{y_{iq}}\left[1-\Phi\left(\mathbf{X}_{iq}^{Fixed}\boldsymbol{\alpha} + \mathbf{X}_{iq}^{Random}\boldsymbol{\beta}_{(g)i}\right)\right]^{1-y_{iq}}\right]^{S_{1i}} \\ \times\left[\Phi\left(\mathbf{X}_{iq}^{Fixed}\boldsymbol{\alpha} + \mathbf{X}_{iq}^{Random}\boldsymbol{\beta}_{(g)i}\right)^{y_{iq}}\left[1-\Phi\left(\mathbf{X}_{iq}^{Fixed}\boldsymbol{\alpha} + \mathbf{X}_{iq}^{Random}\boldsymbol{\beta}_{(g)i}\right)\right]^{1-y_{iq}}\right]^{S_{2i}} \\ \times\ldots \\ \times\left[\Phi(\mathbf{X}_{iq}^{Fixed}\boldsymbol{\alpha} + \mathbf{X}_{iq}^{Random}\boldsymbol{\beta}_{(g)i})^{y_{iq}}\left[1-\Phi\left(\mathbf{X}_{iq}^{Fixed}\boldsymbol{\alpha} + \mathbf{X}_{iq}^{Random}\boldsymbol{\beta}_{(g)i}\right)\right]^{1-y_{iq}}\right]^{S_{gi}} \end{array}\right\} \tag{5}$$

where $\Phi(.)$ is the cumulative standard normal distribution function.

5. Bayesian inference and MCMC simulation

Bayesian estimation treats all estimable parameters, including the latent variable, as random variables and updating prior beliefs into posterior beliefs conditional on observed data. Since latent variable z_i is unobserved, a data augmentation step (Tanner and Wong, 1987; Chib, 1992; Albert and Chib, 1993) can be used before the implementation of the MCMC simulation in order to improve computational efficiency while keeping parameters inferences unchanged so that the augmented posterior conditional can assume a tractable form for the use of Gibbs sampling (Koop et al., 2007). Given prior independence, the augmented joint posterior density function is derived as:

$$\begin{aligned} p(\alpha,\beta,z|y) = \frac{p(y,z|\alpha,\beta)p(\alpha,\beta)}{p(y)} &\propto p(y|z,\alpha,\beta)p(z|\alpha,\beta)p(\alpha)p(\beta) = p(\boldsymbol{\alpha})\prod_{g=0}^{1}p(\boldsymbol{\mu}_{\beta g})p(\mathbf{V}_{\beta g}) \times \prod_{g=0}^{1}\prod_{i:S_i=g}^{n}\prod_{q=1}^{Q_i}[\delta(z_{iq}\leqslant 0)\delta(y_{iq}=0) \\ &+ \delta(z_{iq}>0)\delta(y_{iq}=1)]\phi\left(z_{iq}, \mathbf{X}_{iq}^{Fixed}\boldsymbol{\alpha} + \mathbf{X}_{iq}^{Random}\boldsymbol{\beta}_{(g)i}, 1\right)p(\boldsymbol{\beta}_{(g)i}|\boldsymbol{\mu}_{\beta g}, \mathbf{V}_{\beta g}) \end{aligned} \tag{6}$$

where $p(\boldsymbol{\alpha})$ and $p(\boldsymbol{\beta}_i|\boldsymbol{\mu}_{\beta g}, \mathbf{V}_{\beta g})$ are priors of a fixed and random parameter of interest; $p(\boldsymbol{\mu}_{\beta g})$ and $p(\mathbf{V}_{\beta g})$ are priors of a random parameter distribution mean and distribution variance; $\delta(.)$ is an indicator function that takes unity if condition in the parenthesis is valid; $\phi(.)$ is the probability density function of a standard normal distribution. Conjugate priors are chosen to accelerate the computation of the iterative simulation via Gibbs sampling. Meanwhile, to reduce the impacts of data sample size on the posterior distribution, non-informative priors are used which are diffuse with large variance (Gelman et al., 2004). The priors in Eq. (5) can be specified as:

$$\begin{aligned} &\boldsymbol{\alpha} \sim Normal(\underline{\boldsymbol{\mu}_\alpha}, \underline{\mathbf{V}_\alpha}) \\ &\boldsymbol{\beta}_{(g)i} \sim Normal(\boldsymbol{\mu}_{\beta g}, \mathbf{V}_{\beta g}) \\ &\boldsymbol{\mu}_{\beta g} \sim Normal(\underline{\eta_g}, \underline{\mathbf{C}_g}) \\ &\mathbf{V}_{\beta g} \sim Inverse\ Wishart([\underline{\rho_g}\underline{R_g}]^{-1}, \underline{\rho_g}) \\ &\boldsymbol{\pi} \sim Dirichlet(\underline{v_1}, \ldots, \underline{v_G}) \end{aligned} \tag{7}$$

where all the priors of the fixed parameters vector and the random parameters mean vector are sampled from multivariate normal distributions while the prior of the random parameters variance–covariance matrix follows an inverse Wishart distribution. Hyper-parameters (which are denoted with underbars) in priors are set as

$$\begin{aligned} &\underline{\boldsymbol{\mu}_\alpha} = \mathbf{0}_m, \quad \underline{\boldsymbol{\eta}_g} = \mathbf{0}_k \\ &\underline{\mathbf{V}_g} = 100\mathbf{I}_m, \quad \underline{\mathbf{C}_g} = 100\mathbf{I}_k \\ &\underline{\rho_g} = k, \quad \underline{R_g} = \mathbf{L}_k, \quad \underline{v_g} = k \end{aligned} \tag{8}$$

where $\mathbf{0}_m$ and $\mathbf{0}_k$ are $m \times 1$ and $k \times 1$ vectors of zeros; $\mathbf{I}_m$ and $\mathbf{I}_k$ are $m \times m$ and $k \times k$ identity matrices; $\mathbf{L}_k$ is a $k \times k$ square matrix with unit elements on first row, first column and main diagonal. By incorporating the priors, an augmented Gibbs sampling can proceed as sequentially drawing parameters from each full conditional posterior density. To deal with poten-

Table 1
Summary of descriptive statistics.

Variable	Mean	Min	Median	Max	Standard deviation
Dependent variable					
Crash injury indicator (1 if an injury crash, 0 otherwise)	0.277	0	0	1	0.448
Explanatory variables					
Rumble strips indicator (1 if rumble strips does not exist on outside shoulder, 0 otherwise)	0.865	0	1	1	0.342
Pavement age (in years)	9.140	4	9	29	4.456
Low pavement friction number indicator (1 if current year friction number <45, 0 otherwise)	0.195	0	0	1	0.397
Angle crash indicator (1 if crash type is angle, 0 otherwise)	0.071	0	0	1	0.257
Rut depth (in inches)	0.128	0.020	0.120	0.650	0.085
Two-lane indicator (1 if crash occurred in segment with two lane per direction, 0 otherwise)	0.984	0	1	1	0.124
Guardian indicator (1 if the driver was driving with a guardian, 0 otherwise)	0.364	0	0	1	0.482
Average daily proportion of trucks	0.116	0.010	0.098	0.283	0.084
Road lighting indicator (1 if crash occurred in darkness with road lighting, 0 otherwise)	0.295	0	0	1	0.457
Narrow median indicator (1 if median with is less than 70 ft, 0 otherwise)	0.089	0	0	1	0.285
65 mi/h speed limit indicator (1 if highest speed limit of segment is 65 mi/h, 0 otherwise)	0.692	0	1	1	0.462

tial label-switching problem inherent in the mixture model, a permutation sampler is carried out for model identification (Frühwirth-Schnatter, 2001). Further details on this approach and other estimation details are provided in the Appendix.

Finally, to interpret the estimated parameters in this model, average marginal effects are calculated to quantify the impact of independent variable on the injury outcome (to determine the average effect that 1 unit change in a variable has on the injury-outcome probability, $\Pr(y = 1|\cdot)$) over the relevant posterior distribution. For a continuous variable, the marginal effect conditional on the parameters is computed as:

$$\frac{\partial \Pr\left(y=1|\mathbf{x}_{iq}^{Fixed}, \mathbf{x}_{iq}^{Random}, \boldsymbol{\alpha}, \boldsymbol{\beta}_i'\right)}{\partial x_j^{Random}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\phi\left(\mathbf{x}_{iq}^{Fixed}\boldsymbol{\alpha} + \mathbf{x}_{iq}^{Random}\boldsymbol{\beta}_i\right)\beta_{ij} \tag{9}$$

if it is a random variable $\left(x_j^{Random}\right)$ or,

$$\frac{\partial \Pr\left(y=1|\mathbf{x}_{iq}^{Fixed}, \mathbf{x}_{iq}^{Random}, \boldsymbol{\alpha}, \boldsymbol{\beta}_i'\right)}{\partial x_j^{Fixed}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\phi\left(\mathbf{x}_{iq}^{Fixed}\boldsymbol{\alpha} + \mathbf{x}_{iq}^{Random}\boldsymbol{\beta}_i\right)\alpha_j \tag{10}$$

if it is a fixed variable $\left(x_j^{Fixed}\right)$.

For indicator variables, the marginal effect conditional on the parameters is computed as:

$$\frac{\partial \Pr\left(y=1|\mathbf{x}_{iq}^{Fixed}, \mathbf{x}_{iq}^{Random}, \boldsymbol{\alpha}, \boldsymbol{\beta}_i'\right)}{\partial x_h^{Random}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\left[\Phi\left(\mathbf{x}_{iq}^{Fixed}\boldsymbol{\alpha} + \mathbf{x}_{iq}^{Random}\boldsymbol{\beta}_i\right) - \Phi\left(\mathbf{x}_{iq}^{Fixed}\boldsymbol{\alpha} + \sum_{j\neq h}^{k} x_{iqj}^{Random}\beta_{ij}\right)\right] \tag{11}$$

if it is a random variable (x_h^{Random}) or,

$$\frac{\partial \Pr\left(y=1|\mathbf{x}_{iq}^{Fixed}, \mathbf{x}_{iq}^{Random}, \boldsymbol{\alpha}, \boldsymbol{\beta}_i'\right)}{\partial x_h^{Fixed}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\left[\Phi\left(\mathbf{x}_{iq}^{Fixed}\boldsymbol{\alpha} + \mathbf{x}_{iq}^{Random}\boldsymbol{\beta}_i\right) - \Phi\left(\sum_{j\neq h}^{m} x_{iqj}^{Fixed}\alpha_j + \mathbf{x}_{iq}^{Random}\boldsymbol{\beta}_i\right)\right] \tag{12}$$

if it is a fixed variable (x_h^{Fixed}). Note that the average marginal effect of any parameter of interest is calculated by the average of conditional marginal effects from simulation draws.

6. Empirical setting

Since the guardian supervising practice in the study is of primary interest, only crashes involving an adolescent driver (when reported as driving alone or driving with a senior guardian with a valid driving license) were considered. A total of 451 police-reported adolescent (16–19 year-old drivers) driving crash records were taken from 5795 total accidents occurring on 231 segments of rural interstate highways in Indiana from January 1995 to December 1999.[2] Of these 451 accident records, 164 adolescent drivers driving were driving with a guardian and 287 of them were driving alone.[3] Regarding the injury

[2] The Indiana graduated driver license law was implemented in January 1999.

[3] A preliminary assessment of these data was performed by using a likelihood ratio test from Maximum Likelihood Estimation (MLE) results (Washington et al., 2011) with $X^2 = -2[LL(\boldsymbol{\beta}_{combined}) - LL(\boldsymbol{\beta}_{driving\ \text{alone only}}) - LL(\boldsymbol{\beta}_{driving\ \text{with guardian only}})]$. $LL(\boldsymbol{\beta}_{combined})$, $LL(\boldsymbol{\beta}_{driving\ \text{alone only}})$, $LL(\boldsymbol{\beta}_{driving\ \text{with guardian only}})$ are log-likelihood at convergence from MLE estimates of the same fixed-parameter binary probit model (without guardian supervising indicator) based on the combined dataset (451 observations), driving alone only dataset (287 observations) and driving with guardian only dataset (164 observations) separately. The result was a statistic of $X^2 = 20.14 > \chi^2_{10,.05} = 18.31$, which suggests that the parameters are not the same in guardian and non-guardian models.

Table 2
Adolescent driver injury frequency and percentage distribution by environment factors and guardian indicator.[a]

	Driving without monitoring of guardians				Driving with monitoring of guardians			
	No-injury		Injury		No-injury		Injury	
Average daily proportion of large trucks								
0–10%	103	41.5%	46	18.5%	78	31.5%	21	8.5%
>10–20%	57	46.0%	26	21.0%	31	25.0%	10	8.1%
>20%	38	35.8%	26	24.5%	31	29.2%	11	10.4%
Road lighting indicator								
Darkness with light	125	39.3%	62	19.5%	103	32.4%	28	8.8%
Daytime/darkness without light	73	54.9%	27	20.3%	25	18.8%	8	6.0%
Narrow median indicator								
Median width < 70 ft	182	44.3%	79	19.2%	118	28.7%	32	7.8%
Median width ⩾ 70 ft	16	40.0%	10	25.0%	10	25.0%	4	10.0%
65 mi/h speed limit indicator								
Highest speed limit < 65 mi/h	69	49.6%	29	20.9%	31	22.3%	10	7.2%
Highest speed limit = 65 mi/h	129	41.3%	60	19.2%	97	31.1%	26	8.3%

[a] The calculated percentages across columns within each row sum up to 100%.

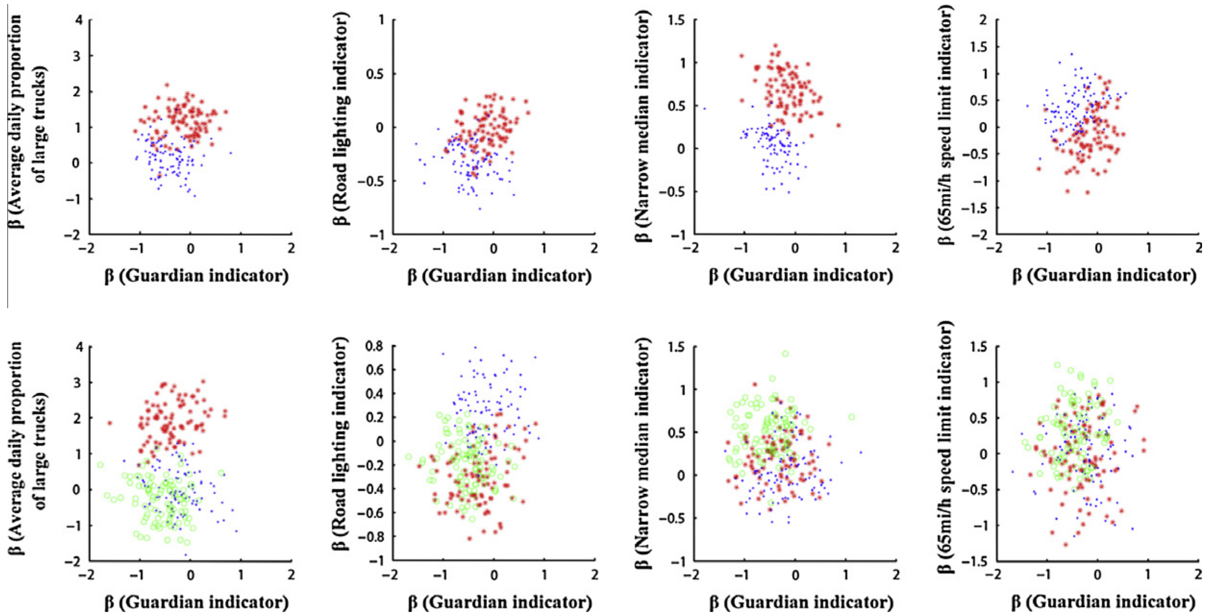

Fig. 1. Scatter plots of posterior draws for random parameters in two-state mixture (first row) and three-state mixture (second row).

severity, 324 (71.8%) of the cases are property damage only, the other 127 (28.2%) crashes involve an injury to the adolescent driver. Because no two (or more) adolescent drivers were involved in any one crash observation, each crash is assumed to occur independently. After merging the driving environment information with driver information, a wide range of potential explanatory variables were gathered including those relating to roadway characteristics, traffic conditions, driver characteristics, and crash features. Descriptive statistics and injury frequencies by key variables are presented in Tables 1 and 2 respectively.

7. Estimation results

The estimation algorithm was coded in MATLAB/Visual C++ mixed implementation.[4] The maximum likelihood estimators of a standard probit model were used as the initial values of iterative simulation. Ten parallel chains of 20,000 iterations are adopted to simulate independently. Each of them only differs by initial parameter values overdispersed throughout parameter space (Gelman and Rubin, 1992).[5] The first 4,000 iterations are discarded as the burn-in period to diminish the effects of initial choice of parameters.

[4] Due to data parallelism inherent in hierarchical part of each iteration, conditional posterior draws of random parameters were implemented via parallel computation which could make most routine computations more efficient (for details please see Xiong et al., 2013).

[5] Parameter spaces were adopted as 90% confidence intervals of the estimated parameters of the fixed parameter model from maximum likelihood estimation.

Table 3
Estimation results of adolescent injury severity (positive parameter estimate increases the probability of an injury accident relative to a no-injury accident, negative parameter decreases it).

Variable	Fixed-parameter binary probit model			Finite-mixture random-parameter binary probit model		
	Posterior mean	Posterior SD	$P(\beta < 0\|y)$	Posterior mean	Posterior SD	$P(\beta < 0\|y)$
Fixed parameters						
Intercept	−0.642	0.510	0.89	−0.623	0.553	0.87[a]
Absence of rumble strips indicator (1 if rumble strips does not exist on outside shoulder, 0 otherwise)	0.822	0.277	0.00	0.821	0.279	0.00
Pavement age (in years)	0.042	0.016	0.01	0.035	0.018	0.03
Low pavement friction number indicator (1 if current year friction number < 45, 0 otherwise)	0.263	0.152	0.04	0.280	0.168	0.05
Rut depth (in inches)	–	–	–	−1.118	0.841	0.91
Two lane indicator (1 if crash occurred in segment with two lane per direction, 0 otherwise)	−1.005	0.586	0.95	−0.968	0.596	0.95
Angle crash indicator (1 if crash type is angle, 0 otherwise)	−0.451	0.277	0.96	−0.433	0.301	0.93
Guardian indicator (1 if the driver was driving with a guardian, 0 otherwise)	−0.314	0.138	0.99	–	–	–

Variable	Posterior mean	Posterior SD	$P(\beta < 0\|y)$	Component 1			Component 2		
				Posterior mean	Posterior SD	$P(\beta < 0\|y)$	Posterior mean	Posterior SD	$P(\beta < 0\|y)$
Random parameter distribution mean (random parameter distribution variance in the parenthesis)									
Guardian indicator (1 if adolescent is driving with a guardian, 0 otherwise, 0 otherwise)	–	–	–	−0.146	0.202	0.72[a]	−0.450	0.248	0.96
				(0.120)	(0.028)	(0)	(0.130)	(0.031)	(0)
Average daily proportion of large trucks	–	–	–	1.089	0.993	0.14	0.142	0.835	0.48[a]
				(0.221)	(0.051)	(0)	(0.239)	(0.058)	(0)
Road lighting indicator (1 if crash occurred in darkness with road lighting, 0 otherwise)	–	–	–	−0.046	0.174	0.60[a]	−0.313	0.207	0.94
				(0.033)	(0.008)	(0)	(0.035)	(0.008)	(0)
Narrow median indicator (1 if median with is less than 70 ft, 0 otherwise)	–	–	–	0.657	0.240	0	0.033	0.514	0.49[a]
				(0.050)	(0.012)	(0)	(0.054)	(0.013)	(0)
65 mi/h speed limit indicator (1 if highest speed limit of segment is 65 mi/h, 0 otherwise)	–	–	–	−0.134	0.180	0.77[a]	0.336	0.216	0.07
				(0.153)	(0.036)	(0)	(0.165)	(0.039)	(0)
Probability of a roadway segment being in component 1	–	–	–	0.518	0.094	0	–	–	–
Number of observations	451								
Maximum potential scale reduction factor	1.001			1.038					
Posterior average log-likelihood	−254.45			−42.45					
Effective number of parameters p_D	6.96			162.67					
Average discrepancy $\overline{D}$	515.86			247.57					
Deviance Information Criterion (*DIC*)	522.82			410.24					

[a] The probability of negative posterior draws ranges between 0.10 and 0.90. $P(\beta < 0|y) \geqslant 0.9$ shows at least 90% of the posterior draws of parameter β after convergence has a statistically significant negative effects while $P(\beta < 0|y) \leqslant 0.1$ denotes that at least 90% of the posteriors are ositively significant.

Scatter plots were created to provide some initial information on the number of mixture states to consider. From the scatter plots of the simulated posterior of random parameters (see Fig. 1), two "components" of roadway segments (individual crashes are assigned to the specific roadway segments on which they occurred, and these segments have identical roadway geometric and traffic conditions) are easily identified while a potential third component overlaps the other two clusters. Fig. 1 clearly suggests that a two-state mixture structure appears quite reasonable for the given data. Due to the increasing number of unknown parameters associated with a growing number of states (and the simulation problems this entails), only a two-state mixture model is considered in this paper.

The proposed finite-mixture random-parameter model is estimated along with a fixed-parameter binary probit model by using permutation sampling based on the same data set. The posterior means as well as posterior standard deviations from the simulation are reported in Table 3. In this table, the probability of negative posterior simulations associated with each parameter are provided as evidence of statistically significance at 90% level (which is considered herein to be statistically

Table 4
Average marginal effects for injury probability (marginal effects standard deviations are in the parenthesis).

Variables	Average marginal effects $P(y=1\|.)$	
Absence of rumble strips indicator (1 if rumble strips does not exist on outside shoulder, 0 otherwise)	0.201 (0.068)	
Pavement age (in years)	0.010 (0.004)	
Low pavement friction number indicator (1 if current year friction number < 45, 0 otherwise)	0.083 (0.028)	
Angle crash indicator (1 if crash type is angle, 0 otherwise)	−0.120 (0.040)	
Rut depth (in inches)	−0.258 (0.098)	
Two lane indicator (1 if crash occurred in segment with two lane per direction, 0 otherwise)	−0.330 (0.059)	
	Component 1	Component 2
Guardian indicator (1 if adolescent is driving with a guardian, 0 otherwise)	−0.027 (0.109)	−0.043 (0.158)
Average daily proportion of large trucks	0.320 (0.195)	0.054 (0.313)
Road lighting indicator (1 if crash occurred in darkness with road lighting, 0 otherwise)	−0.009 (0.028)	−0.025 (0.059)
Narrow median indicator (1 if median with is less than 70 ft, 0 otherwise)	0.019 (0.068)	0.001 (0.025)
65 mi/h speed limit indicator (1 if highest speed limit of segment is 65 mi/h, 0 otherwise)	−0.014 (0.098)	0.075 (0.149)

significant). Moreover, the diagonal elements of distribution variance–covariance matrix were found to be all positively significant, providing additional statistical evidence for the modeling approach.[6]

A potential scale reduction factor (Gelman and Rubin, 1992) is used to assess the between and the within multiple chains convergence.[7] Gelman et al. (2004) show that potential scale reduction factor values less than 1.1 indicate that the sequences have stabilized and convergence is likely to have been achieved. Table 3 shows the potential scale reduction factor values for our estimations are both less than 1.1.

The results provided in Table 3 clearly show that the finite-mixture random-parameter model outperforms the simple fixed-parameter binary probit as indicated by factors such as the posterior average log-likelihood, Deviance Information Criterion (DIC), and average discrepancy $\overline{D}$.[8] Also, please note that the guardian indicator variable produces a significant negative parameter in both the simple fixed and finite-mixture random-parameter models, indicating that the presence of a guardian reduces the probability of an injury accident. However, the findings of the finite-mixture random-parameter model show that the relationship between guardian presence and injury likelihood is far more complex than that suggested by the simple fixed parameter approach (more on this below).

Turning first to fixed-parameter estimation results of the finite-mixture random-parameters model, Tables 3 and 4 show that model estimates indicate that the absence of rumble strips indicator, pavement age, low friction number indicator, angle crash indicator, rut depth and two-lane indicator all generate significant fixed parameters. As expected, when rumble strips do not exist on an outside shoulder the likelihood of an injury crash increases. Here, the lack of feedback on eminent road departure does not provide drivers with information that can help reduce the probability that the crash will result in an injury. Increasing pavement age also was found to increase the likelihood of an injury crash. In this case, potential surface irregularities associated with older pavements may cause sporadic vehicle response to driver inputs thus increasing the likelihood of an injury crash. Similarly, a low pavement friction (friction number less than 45[9]) was found to increase the likelihood of that the crash will result in an injury.

In contrast, increasing rut depth of the pavement was found to decrease the likelihood of injury (it is notable that this variable was statistically insignificant in the simple fixed-parameter model). This finding is consistent with that of other studies that have found that pavement ruts tend to decrease crash-injury severity, likely because vehicles tend to “track” on the ruts thus making them less likely to veer off of the road and strike stationary objects or other vehicles (Anastasopoulos and Mannering, 2011).

Freeways with just two lanes in each direction showed a lower likelihood for injury (all roads are interstates, so two-lanes in each direction is the minimum number of lanes). This may reflect that roadway segments with three or more lanes may have more complex and less-predictable vehicle movements that may be difficult for adolescent drivers to comprehend and process.

Finally, during the process of a collision, crashes defined as angle-collisions are found to reduce the probability of injury relative to non-angle crashes such as head-on, rear end, and so on, which is consistent with findings in previous studies that show that in most cases angle-collisions are the least dangerous (Paleti et al., 2010).

---

[6] The statistical significance of some of the off-diagonal elements of the variance–covariance matrix also indicates the significance of the random parameters distribution covariance (interaction) among contributing factors.

[7] Suppose m chains of length n are used in simulation and ψ_{ij} is the ith simulation from chain j. Given $B = \frac{n}{m-1}\sum_{j=1}^{m}(\overline{\psi}_{.j} - \overline{\psi}_{..})^2$ and $W = \frac{1}{m}\frac{1}{n-1}\sum_{j=1}^{m}\sum_{i=1}^{n}(\psi_{ij} - \overline{\psi}_{.j})^2$, the potential scale reduction factor is defined as $\widehat{R} = \sqrt{(n-1)n^{-1} + n^{-1}B/W}$.

[8] To compare the performance of the simple fixed parameter model and the finite-mixture random parameter model, the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) is used as an estimate of how much “out-of-sample” predictive capability the model has. It can be denoted as a sum of the effective number of parameters p_D and model fitting measure $\overline{D}$. The proposed model has a smaller average discrepancy $\overline{D}$ and smaller DIC relative to the fixed model, which suggests better goodness-of-fit given the data set and better out-of-sample predictive power.

[9] Friction numbers are computed based on the coefficient of road adhesion and are scaled from 0 to 100 to assess the relative friction on different roadway surfaces (Mannering and Washburn, 2013).

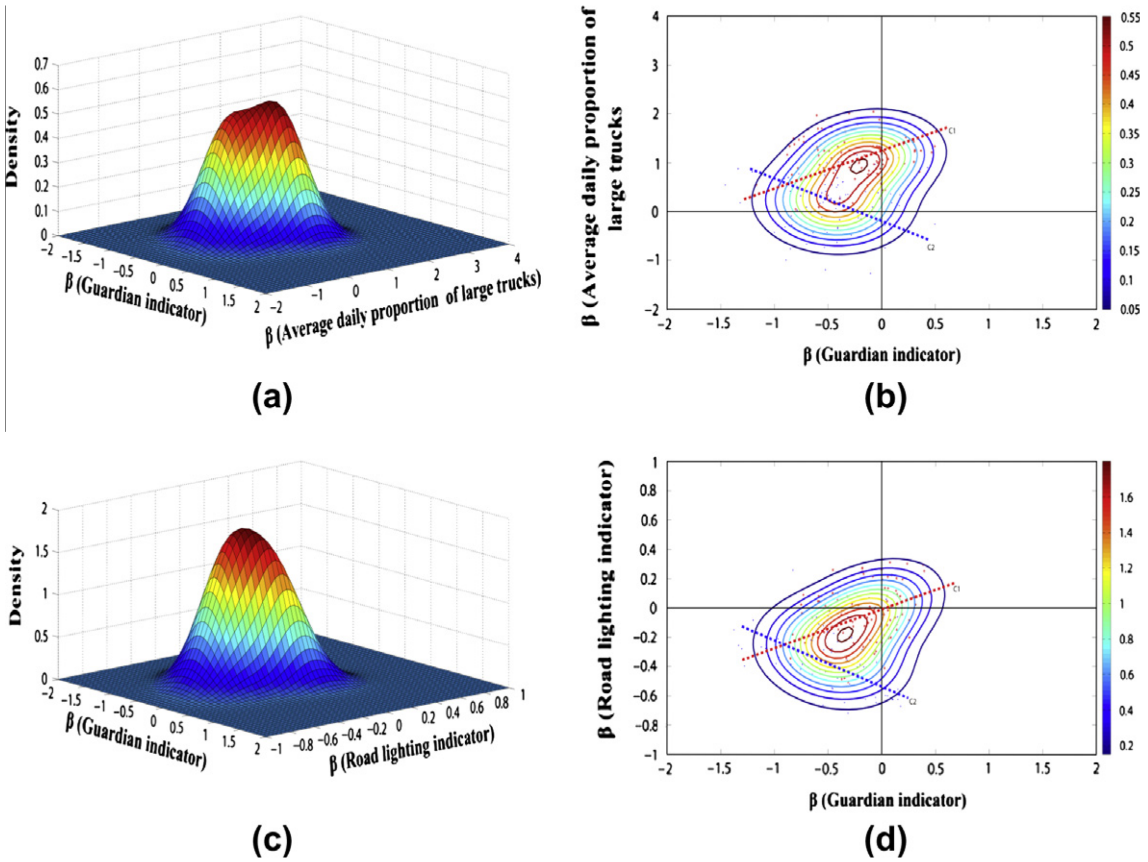

Fig. 2. Marginal posterior distribution and contour line plots of random parameter estimates (first row: guardian indicator vs. average daily proportion of large trucks; second row: guardian indicator vs. road lighting indicator) (C1 and C2 provide directions of bivariate distributions of components 1 and 2 respectively).

With regard to the effects of guardian supervision in the finite-mixture random-parameters model, the statistical significance of the probability-of-component parameter (see Table 3) shows the presence of two distinct roadway-segment components. This finding suggests that a distinct discrete partitioning of the roadway segment population exists and implies that adolescents driving in different types of roadway environments have different driving behaviors in the presence of a guardian. Component 1 has the slightly larger segment population (the mean probability of the segment being in component 1 is 0.518 as shown in Table 3). Adolescents driving under these conditions seem to be less receptive to guardian input, with the marginal effects showing that the presence of a guardian reduces the probability of an injury crash by an average of 0.027 for component 1 roadway segments as opposed to a probability reduction of 0.043 for component 2 roadway segments. This may be because the stimuli provided in the component 1 driving environment is within familiar ranges and need for, and attention given to, guardian inputs may be less. In contrast, component 2 is likely to define combination of roadway-segment factors that make guardian supervision more likely to be heeded.

Turning to other component-specific results, the effect of an increasing percent of large trucks[10] in the traffic stream is expected to increase the likelihood of an injury crash because large trucks can disrupt the traffic stream with their additional inertia and longer braking distance presenting adolescent drivers more a complex decision-making environment (this expectation is supported by past research such as Chang and Mannering, 1999). The estimation results confirm this with the probability of an injury crash increasing by 0.00332 for every 1% increase in the percent of large trucks in the traffic stream for component 1 roadway segments and 0.00054 for component 2 roadways (as shown in Table 4).

Looking more closely at the effect of large trucks across components, in the case of component 1 roadway segments, the computed positive covariance between the guardian supervising indicator and average daily proportion of large trucks indicates that the guardian supervision can help mitigate the impacts of higher truck percentages on crash-injury probabilities. To see this, consider Fig. 2. In this figure, the marginal posterior distribution of the guardian indicator and average daily proportion of large trucks is provided as a two-component mixture of bivariate Gaussian distributions (in Fig. 2a). Even though

[10] Large trucks are defined as single unit or combination tucks exceeding 10,000 lb (44.5 kN).

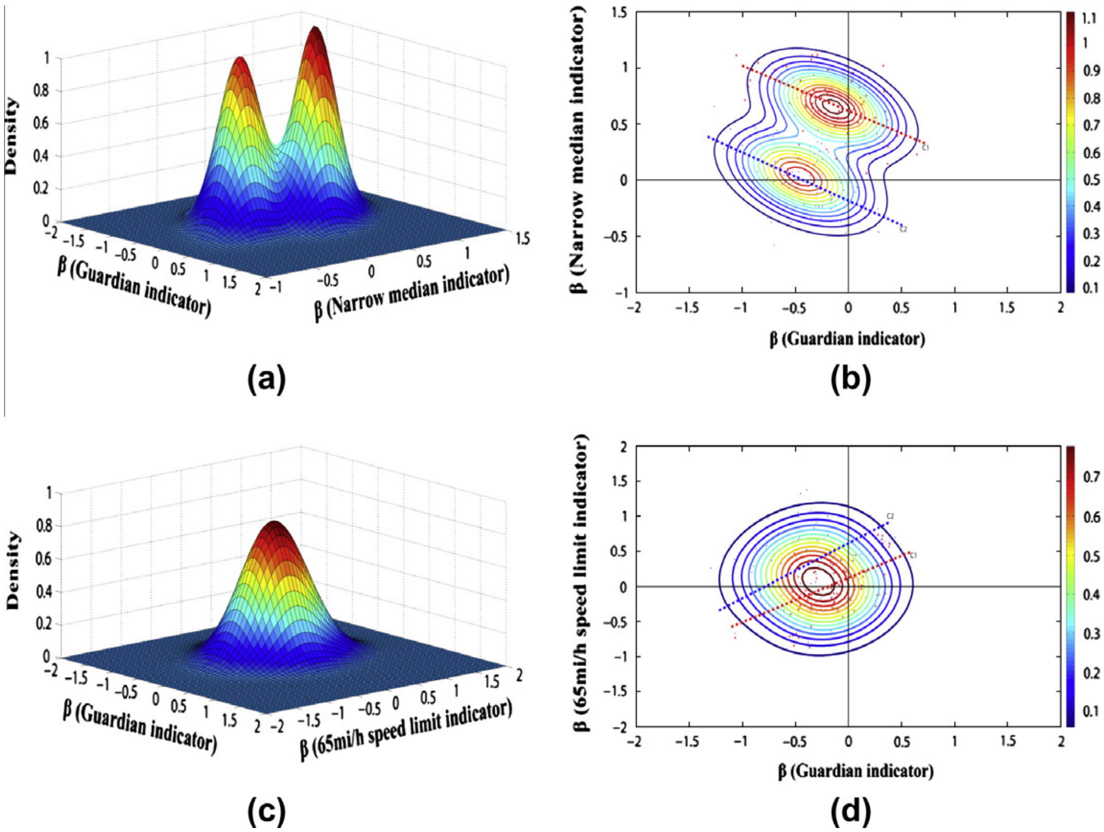

Fig. 3. Marginal posterior distribution and contour line plots of random parameter estimates (first row: guardian indicator vs. narrow median indicator; second row: guardian indicator vs. 65 mi/h speed limit indicator) (C1 and C2 provide directions of bivariate distributions of components 1 and 2 respectively).

the two distributions are mixed too closely to have distinct peaks identified, the corresponding contour line plot (Fig. 2b) demonstrates skewness and bimodality of the marginal posterior distribution. Actually, the distribution can be decomposed to one bivariate Gaussian distribution (component 1) with positive covariance (0.022) and another (component 2) with negative covariance (−0.012). Looking at Fig. 2b, for the roadway segments in component 1, the less (negative) effects the guardian supervising indicator has on injury probability, the more (positive) effect average daily proportion of trucks has – while for those roadway segments in component 2, the (positive) effects of a large truck percentage could drop along with the decline of guardian supervising (negative) effects. Furthermore, the majority of the population (about 59.8%) is distributed in the upper left-hand quadrant and has negative signs of the guardian-supervising indicator and positive values of average daily proportion of large trucks, which indicates that adolescent drivers on a substantial amount of roadway segments can benefit from supervising practice but still have increased injury probabilities resulting from higher large truck percentages.

Road lighting at night was found to decrease the probability of injury among adolescent drivers for both roadway segment components.[11] There was a positive covariance (0.023) for component 1 indicating that the presence of guardians could further decease the probability of injury on lighted roads at night (guardians are presumably able to help detect potentially hazardous situations such as animals crossing the roadway and so on). In contrast, component 2 has a negative covariance (−0.012) suggesting that in some cases guardian presence might lead to an auditory or visual distraction which could lead to a higher likelihood of injury. These findings are visually confirmed by the two-component mixture of bivariate Gaussian distributions (Fig. 2c) and its contour line plot (Fig. 2d). Fig. 2c and d show a larger proportion (63.6%) of random parameters in the lower-left quadrant reveals that guardian supervising practice as well as road lighting can keep the majority of teen drivers alert and help decrease the probability being injured in a crash.

[11] Adolescents driving at night is typically recognized as a high-risk condition, and night-time driving is often restricted by graduated driver license laws. For instance, graduated driver licenses issued in Indiana since June 2010 restricts nighttime driving of teens from 10 pm to 5 am in first 180 days of their probationary license (Nagle, 2011). But road lighting at night can mitigate the adverse effects of darkness as was found in our case.

Estimation results show that narrow medians on both components of roadway segments increase the probability of an injury crash (marginal effects show the influence of narrow medians is greater in component 1 segments than component 2 segments). Interestingly, the interaction with guardian supervision is negative (the covariance between guardian supervision and the median indicator is −0.034 and −0.036 for components 1 and 2 respectively). This is characterized by two component (bimodal) bivariate distributions with similar skewness and kurtosis in Fig. 3a and b. However, the fact that most (60.3%) of the roadway segment parameter population is in the upper-right quadrant of (3b) shows that teen drivers on most of segments still benefit from guardian supervision in the presence of narrow medians.

Finally, the effect of the 65 mi/h speed limit[12] on average results in a reduction in injury probabilities in component 1 roadway segments (−0.014 as shown in Table 4) and an increase in injury probabilities in component 2 roadway segments (0.075 as shown in Table 4). In component 1 segments, adolescents seem more cautious, possibly by adjusting their speeds on 65 mi/h segments differently than they do on roadway segments with 60 mi/h and 55 mi/h as their highest speed limits. In contrast, on component 2 segments the adjustment of speed between 65 mi/h and non-65 mi/h segments results in higher injury probabilities on the higher speed-limit segments. There is a positive covariance between guardian presence and the 65 mi/h speed limit indicator (0.022 for component 1 segments and 0.036 for component 2 segments) suggesting that guardian presence can be important in reducing injury probabilities in this case. The three dimensional plot in Fig. 3c shows skewness but little bimodality and kurtosis of the bivariate normal distribution. As indicated in Fig. 3d, guardian supervision decreases the injury probability of most crashes (see the left two quadrants in Fig. 3d) but the net effect of the 65 mi/h indicator is pretty much equally distributed between positive and negative effects (the areas of the upper vs. lower two quadrants in Fig. 3d).

8. Conclusions and recommendations

In order to characterize the heterogeneous effects of guardian supervision, a finite-mixture random-parameter framework is developed. The modeling approach accounts for between-component variation as well as variation within each component. Also, a multivariate structure is considered for the distribution of random parameters so that the complex interactions between guardian supervision and other factors contributing to injury probabilities can be explored. This formulation accommodates the multimodality, skewness, kurtosis as well as multivariate characteristics of random parameters distribution in discrete outcome modeling as applied to crash-injury outcomes. A Bayesian estimation method based on data augmentation is used and a permutation sampler is extended to deal with the label switching of unknown components.

Using five-years of adolescent crash data from Indiana rural Interstates, guardian supervision is shown to be generally effective in reducing the probability of adolescent driver injuries in crashes. However, the effect of guardian supervision is complex and depends on many factors including the possibility of distinct driving environments. In this paper, the estimation results confirm the presence of two distinct driving environments that are defined by categories of roadway segments. These two roadway-segment components, while both confirming the general effectives of guardian presence in reducing injury likelihoods among adolescent drivers, show distinctly different injury-severity outcome processes when guardians are present. While the results of this study support graduated driver licensing programs in general, they also show the high variability in the effectiveness of guardian supervision under various circumstances. This variability is the result of a complex interaction among the roadway geometric and traffic conditions, the input provided by guardians, and adolescent driver response to the road environment, guardian presence, and guardian input.

In terms of better assessing the potential effectiveness of graduate licensing programs that involve guardian supervision, future research could study effectiveness of guardian supervision in each of the three stages; learners permit, probationary license and unrestricted license. From a methodological perspective, larger datasets would presumably allow for higher numbers of components in the mixture structure. Also, efforts dealing with alternate distributional assumptions and the possibility of developing a model structure that allows effects of variables to change over time are both fruitful areas for future methodological research.

Acknowledgement

The authors thank Dr. Jon Fricker for his helpful suggestions and financial support during this study.

Appendix A. Model estimation details

The augmented Gibbs sampling approach used in this paper sequentially draws parameters from each full conditional posterior density as follows:

(A) Sample $\boldsymbol{\alpha}$ from $\boldsymbol{\alpha}|\{\boldsymbol{\beta}_{(g)i}\}, \boldsymbol{\mu}_{\beta g}, \mathbf{V}_{\beta g}, \{S_i\}, \boldsymbol{\pi}, \{z_{iq}\}, \mathbf{Y}$

[12] Over the January 1995 to December 1999 time period, maximum rural interstate speed limits in Indiana were 65 mi/h. Other possible speed limits on rural interstates during this time period were 60 mi/h and 55 mi/h.

$$p(\boldsymbol{\alpha}|\{\boldsymbol{\beta}_{(g)i}\},\boldsymbol{\mu}_{\beta g},\mathbf{V}_{\beta g},\{S_i\},\boldsymbol{\pi},\{z_{iq}\},\mathbf{Y}) \propto \frac{1}{(2\pi)^{\frac{m}{2}}}\left|\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\mathbf{X}_{iq}^{Fixed\,\prime}\mathbf{X}_{iq}^{Fixed}+\underline{\mathbf{V}_{\alpha}}^{-1}\right|^{\frac{1}{2}}$$
$$\times \exp\left(\begin{array}{l} -\frac{1}{2}\left(\boldsymbol{\alpha}-\left(\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\mathbf{X}_{iq}^{Fixed\,\prime}\mathbf{X}_{iq}^{Fixed}+\underline{\mathbf{V}_{\alpha}}^{-1}\right)^{-1}\left(\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\mathbf{X}_{iq}^{Fixed\,\prime}\left(z_{iq}-\mathbf{X}_{iq}^{Random}\boldsymbol{\beta}_{(g)i}\right)+\underline{\mathbf{V}_{\alpha}}^{-1}\underline{\boldsymbol{\mu}_{\alpha}}\right)\right)^{\prime} \\ \times\left(\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\mathbf{X}_{iq}^{Fixed\,\prime}\mathbf{X}_{iq}^{Fixed}+\underline{\mathbf{V}_{\alpha}}^{-1}\right) \\ \times\left(\boldsymbol{\alpha}-\left(\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\mathbf{X}_{iq}^{Fixed\,\prime}\mathbf{X}_{iq}^{Fixed}+\underline{\mathbf{V}_{\alpha}}^{-1}\right)^{-1}\left(\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\mathbf{X}_{iq}^{Fixed\,\prime}\left(z_{iq}-\mathbf{X}_{iq}^{Random}\boldsymbol{\beta}_{(g)i}\right)+\underline{\mathbf{V}_{\alpha}}^{-1}\underline{\boldsymbol{\mu}_{\alpha}}\right)\right) \end{array}\right) \overset{ind}{\sim} Normal(D_{\alpha}d_{\alpha},D_{\alpha}) \tag{A1}$$

where

$$D_{\alpha}=\left(\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\mathbf{X}_{iq}^{Fixed\,\prime}\mathbf{X}_{iq}^{Fixed}+\underline{\mathbf{V}_{\alpha}}^{-1}\right)^{-1} \quad \text{and } d_{\alpha}=\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\mathbf{X}_{iq}^{Fixed\,\prime}\left(z_{iq}-\mathbf{X}_{iq}^{Random}\boldsymbol{\beta}_{(g)i}\right)+\underline{\mathbf{V}_{\alpha}}^{-1}\underline{\boldsymbol{\mu}_{\alpha}}$$

(B) Sample $\boldsymbol{\beta}_{(g)i}$ from $\boldsymbol{\beta}_{(g)i}|\boldsymbol{\alpha}$, $\boldsymbol{\mu}_{\beta g}$, $\mathbf{V}_{\beta g}$, $\{S_i\}$, $\boldsymbol{\pi}$, $\{z_{iq}\}$, $\mathbf{Y}$

$$p(\boldsymbol{\beta}_{(g)i}|\boldsymbol{\alpha},\boldsymbol{\mu}_{\beta g},\mathbf{V}_{\beta g},\{S_i\},\boldsymbol{\pi},\{z_{iq}\},\mathbf{Y}) \propto \frac{1}{(2\pi)^{\frac{k}{2}}}\left|\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\mathbf{X}_{iq}^{Random\,\prime}\mathbf{X}_{iq}^{Random}+\mathbf{V}_{\beta g}^{-1}\right|^{\frac{1}{2}}$$
$$\times \exp\left(\begin{array}{l} -\frac{1}{2}\left(\boldsymbol{\beta}_{(g)i}-\left(\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\mathbf{X}_{iq}^{Random\,\prime}\mathbf{X}_{iq}^{Random}+\mathbf{V}_{\beta g}^{-1}\right)^{-1}\left(\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\mathbf{X}_{iq}^{Random\,\prime}(z_{iq}-\mathbf{X}_{iq}^{Fixed}\boldsymbol{\alpha})+\mathbf{V}_{\beta g}^{-1}\boldsymbol{\mu}_{\beta g}\right)\right)^{\prime} \\ \times\left(\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\mathbf{X}_{iq}^{Random\,\prime}\mathbf{X}_{iq}^{Random}+\mathbf{V}_{\beta g}^{-1}\right) \\ \times\left(\boldsymbol{\beta}_{(g)i}-\left(\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\mathbf{X}_{iq}^{Random\,\prime}\mathbf{X}_{iq}^{Random}+\mathbf{V}_{\beta g}^{-1}\right)^{-1}\left(\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\mathbf{X}_{iq}^{Random\,\prime}(z_{iq}-\mathbf{X}_{iq}^{Fixed}\boldsymbol{\alpha})+\mathbf{V}_{\beta g}^{-1}\boldsymbol{\mu}_{\beta g}\right)\right) \end{array}\right)$$
$$\overset{ind}{\sim} Normal(D_{\beta_i}d_{\beta_i},D_{\beta_i}), \quad i=1,2,\ldots,n \tag{A2}$$

where

$$D_{\beta_i}=\left(\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\mathbf{X}_{iq}^{Random\,\prime}\mathbf{X}_{iq}^{Random}+\mathbf{V}_{\beta g}^{-1}\right)^{-1} \quad \text{and } d_{\beta_i}=\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\mathbf{X}_{iq}^{Random\,\prime}\left(z_{iq}-\mathbf{X}_{iq}^{Fixed}\boldsymbol{\alpha}\right)+\mathbf{V}_{\beta g}^{-1}\boldsymbol{\mu}_{\beta g}$$

(C) Sample $\boldsymbol{\mu}_{\beta g}$ from $\boldsymbol{\mu}_{\beta g}|\boldsymbol{\alpha}$, $\{\boldsymbol{\beta}_{(g)i}\}$, $\mathbf{V}_{\beta g}$, $\{S_i\}$, $\boldsymbol{\pi}$, $\{z_{iq}\}$, $\mathbf{Y}$

$$p(\boldsymbol{\mu}_{\beta g}|\boldsymbol{\alpha},\{\boldsymbol{\beta}_{(g)i}\},\mathbf{V}_{\beta g},\boldsymbol{\pi},\{z_{iq}\},\mathbf{Y}) \propto \frac{1}{(2\pi)^{\frac{m}{2}}|n\mathbf{V}_{\beta g}^{-1}+\underline{\mathbf{C}_{g}}^{-1}|^{-\frac{1}{2}}}$$
$$\exp\left(\begin{array}{l} -\frac{1}{2}\left(\boldsymbol{\mu}_{\beta g}-\left(n\mathbf{V}_{\beta g}^{-1}+\underline{\mathbf{C}_{g}}^{-1}\right)^{-1}\left(\sum_{i=1}^{n}\boldsymbol{\beta}_i\mathbf{V}_{\beta g}^{-1}+\underline{\mathbf{C}_{g}}^{-1}\underline{\boldsymbol{\eta}_{g}}\right)\right)^{\prime}\left(n\mathbf{V}_{\beta g}^{-1}+\underline{\mathbf{C}_{g}}^{-1}\right) \\ \times\left(\boldsymbol{\mu}_{\beta g}-\left(n\mathbf{V}_{\beta g}^{-1}+\underline{\mathbf{C}_{g}}^{-1}\right)^{-1}\left(\sum_{i=1}^{n}\boldsymbol{\beta}_i\mathbf{V}_{\beta g}^{-1}+\underline{\mathbf{C}_{g}}^{-1}\underline{\boldsymbol{\eta}_{g}}\right)\right) \end{array}\right)$$
$$\overset{ind}{\sim} Normal(D_{\boldsymbol{\mu}_{\beta}}d_{\boldsymbol{\mu}_{\beta}},D_{\boldsymbol{\mu}_{\beta}}) \tag{A3}$$

where

$$D_{\boldsymbol{\mu}_{\beta}}=\left(n\mathbf{V}_{\beta g}^{-1}+\underline{\mathbf{C}_{g}}^{-1}\right)^{-1} \quad \text{and } d_{\boldsymbol{\mu}_{\beta}}=\sum_{i=1}^{n}\boldsymbol{\beta}_i\mathbf{V}_{\beta g}^{-1}+\underline{\mathbf{C}_{g}}^{-1}\underline{\boldsymbol{\eta}_{g}}$$

(D) Sample $\mathbf{V}_{\beta g}$ from $\mathbf{V}_{\beta g}|\boldsymbol{\alpha}$, $\{\boldsymbol{\beta}_{(g)i}\}$, $\boldsymbol{\mu}_{\beta g}$, $\{S_i\}$, $\boldsymbol{\pi}$, $\{z_{iq}\}$, $\mathbf{Y}$

$$p(\mathbf{V}_{\beta g}|\boldsymbol{\alpha},\{\boldsymbol{\beta}_{(g)i}\},\boldsymbol{\mu}_{\beta g},\boldsymbol{\pi},\{z_i\},\mathbf{Y}) \propto \frac{\left|\sum_{i=1}^{30}[(\boldsymbol{\beta}_{(g)i}-\boldsymbol{\mu}_{\beta g})(\boldsymbol{\beta}_{(g)i}-\boldsymbol{\mu}_{\beta g})'+\underline{\rho_g}\underline{R_g}]^{-1}\right|^{\frac{k+\underline{R_g}}{2}}}{\Gamma_k\left(\frac{k+\underline{R_g}}{2}\right)}|\mathbf{V}_{\beta g}|^{-\frac{2k+\underline{R_g}+1}{2}}\times$$

$$\exp\left(-\frac{1}{2}tr\left(\sum_{i=1}^{n}[(\boldsymbol{\beta}_{(g)i}-\boldsymbol{\mu}_{\beta g})(\boldsymbol{\beta}_{(g)i}-\boldsymbol{\mu}_{\beta g})'+\underline{\rho_g}\underline{R_g}]^{-1}\mathbf{V}_{\beta g}^{-1}\right)\right)$$

$$\overset{ind}{\sim} Inverse\ Wishart\left(\sum_{i=1}^{n}[(\boldsymbol{\beta}_{(g)i}-\boldsymbol{\mu}_{\beta g})(\boldsymbol{\beta}_{(g)i}-\boldsymbol{\mu}_{\beta g})'+\underline{\rho_g}\underline{R_g}]^{-1},k+\underline{R_g}\right) \tag{A4}$$

where $\Gamma_k(.)$ is a k-dimension multivariate Gamma distribution

(E) Sample S_i from $S_i|\boldsymbol{\alpha}$, $\{\boldsymbol{\beta}_{(g)i}\}$, $\boldsymbol{\mu}_{\beta g}$, $\mathbf{V}_{\beta g}$, $\boldsymbol{\pi}$, $\{z_{iq}\}$, $\mathbf{Y}$

$$p(S_i|\boldsymbol{\alpha},\{\boldsymbol{\beta}_{(g)i}\},\boldsymbol{\mu}_{\beta g},\mathbf{V}_{\beta g},\boldsymbol{\pi},\{z_{iq}\},\mathbf{Y}) \propto \left(\prod_{q=1}^{Q_i}\frac{\pi_1\Phi\left(\mathbf{X}_{iq}^{Fixed}\boldsymbol{\alpha}+\mathbf{X}_{iq}^{Random}\boldsymbol{\beta}_{(g)i}\right)^{y_{iq}}\left[1-\Phi(\mathbf{X}_{iq}^{Fixed}\boldsymbol{\alpha}+\mathbf{X}_{iq}^{Random}\boldsymbol{\beta}_{(g)i})\right]^{1-y_{iq}}}{\sum_{g=1}^{G}\pi_g\Phi\left(\mathbf{X}_{iq}^{Fixed}\boldsymbol{\alpha}+\mathbf{X}_{iq}^{Random}\boldsymbol{\beta}_{(g)i}\right)^{y_{iq}}\left[1-\Phi\left(\mathbf{X}_{iq}^{Fixed}\boldsymbol{\alpha}+\mathbf{X}_{iq}^{Random}\boldsymbol{\beta}_{(g)i}\right)\right]^{1-y_{iq}}}\right)^{S_{1i}}\cdots$$

$$\times\left(\prod_{q=1}^{Q_i}\frac{\pi_G\Phi\left(\mathbf{X}_{iq}^{Fixed}\boldsymbol{\alpha}+\mathbf{X}_{iq}^{Random}\boldsymbol{\beta}_{(G)i}\right)^{y_{iq}}\left[1-\Phi\left(\mathbf{X}_{iq}^{Fixed}\boldsymbol{\alpha}+\mathbf{X}_{iq}^{Random}\boldsymbol{\beta}_{(G)i}\right)\right]^{1-y_{iq}}}{\sum_{g=1}^{G}\pi_g\Phi\left(\mathbf{X}_{iq}^{Fixed}\boldsymbol{\alpha}+\mathbf{X}_{iq}^{Random}\boldsymbol{\beta}_{(g)i}\right)^{y_{iq}}\left[1-\Phi\left(\mathbf{X}_{iq}^{Fixed}\boldsymbol{\alpha}+\mathbf{X}_{iq}^{Random}\boldsymbol{\beta}_{(g)i}\right)\right]^{1-y_{iq}}}\right)^{S_{Gi}}$$

$$\overset{ind}{\sim} Multinomial\left(1,\begin{bmatrix}\frac{\pi_1\Phi\left(\mathbf{X}_i^{Fixed}\boldsymbol{\alpha}+\mathbf{X}_i^{Random}\boldsymbol{\beta}_{(g)i}\right)^{y_{iq}}\left[1-\Phi\left(\mathbf{X}_i^{Fixed}\boldsymbol{\alpha}+\mathbf{X}_i^{Random}\boldsymbol{\beta}_{(g)i}\right)\right]^{1-y_{iq}}}{\sum_{g=1}^{G}\pi_g\Phi\left(\mathbf{X}_i^{Fixed}\boldsymbol{\alpha}+\mathbf{X}_i^{Random}\boldsymbol{\beta}_{(g)i}\right)^{y_{iq}}\left[1-\Phi\left(\mathbf{X}_i^{Fixed}\boldsymbol{\alpha}+\mathbf{X}_i^{Random}\boldsymbol{\beta}_{(g)i}\right)\right]^{1-y_{iq}}}\\ \cdots\frac{\pi_G\Phi\left(\mathbf{X}_i^{Fixed}\boldsymbol{\alpha}+\mathbf{X}_i^{Random}\boldsymbol{\beta}_{(G)i}\right)^{y_{iq}}\left[1-\Phi\left(\mathbf{X}_i^{Fixed}\boldsymbol{\alpha}+\mathbf{X}_i^{Random}\boldsymbol{\beta}_{(G)i}\right)\right]^{1-y_{iq}}}{\sum_{g=1}^{G}\pi_g\Phi\left(\mathbf{X}_i^{Fixed}\boldsymbol{\alpha}+\mathbf{X}_i^{Random}\boldsymbol{\beta}_{(g)i}\right)^{y_{iq}}\left[1-\Phi\left(\mathbf{X}_i^{Fixed}\boldsymbol{\alpha}+\mathbf{X}_i^{Random}\boldsymbol{\beta}_{(g)i}\right)\right]^{1-y_{iq}}}\end{bmatrix}'\right) \tag{A5}$$

(F) Sample $\boldsymbol{\pi}$ from $\boldsymbol{\pi}|\boldsymbol{\alpha}$, $\{\boldsymbol{\beta}_{(g)i}\}$, $\boldsymbol{\mu}_{\beta g}$, $\mathbf{V}_{\beta g}$, $\{S_i\}$, $\{z_{iq}\}$, $\mathbf{Y}$

$$p(\boldsymbol{\pi}|\boldsymbol{\alpha},\{\boldsymbol{\beta}_{(g)i}\},\boldsymbol{\mu}_{\beta g},\mathbf{V}_{\beta g},\{S_i\},\{z_{iq}\},\mathbf{Y}) \propto \left[\frac{\Gamma\left(\sum_{i=1}^{n}(n_i+\underline{v_i})\right)}{\prod_{i=1}^{n}\Gamma(n_i+\underline{v_i})}\right]\prod_{i=1}^{N}\pi_i^{n_i+\underline{v_i}-1}\overset{ind}{\sim} Dirichlet(n_1+\underline{v_1},n_2+\underline{v_2},\ldots,n_G+\underline{v_G}) \tag{A6}$$

where

$$n_g \equiv \sum_{i=1}^{N}S_{gi}$$

(G) Sample z_{iq} from $z_{iq}|\boldsymbol{\alpha}$, $\{\boldsymbol{\beta}_{(g)i}\}$, $\boldsymbol{\mu}_{\beta g}$, $\mathbf{V}_{\beta g}$, $\{S_i\}$, $\boldsymbol{\pi}$, $\mathbf{Y}$

$$P(z_{iq}|\boldsymbol{\alpha},\{\boldsymbol{\beta}_{(g)i}\},\boldsymbol{\mu}_{\beta g},\mathbf{V}_{\beta g},\boldsymbol{\pi},\mathbf{Y})\overset{ind}{\sim}\begin{cases}Truncated\ Normal_{(-\infty,0]}\left(\mathbf{X}_{iq}^{Fixed}\boldsymbol{\alpha}+\mathbf{X}_{iq}^{Random}\boldsymbol{\beta}_{(g)i},1\right) & \text{if } y_{iq}=0\\ Truncated\ Normal_{(0,\infty)}\left(\mathbf{X}_{iq}^{Fixed}\boldsymbol{\alpha}+\mathbf{X}_{iq}^{Random}\boldsymbol{\beta}_{(g)i},1\right) & \text{if } y_{iq}=1\end{cases} \tag{A7}$$

Iteratively drawing from steps (A)–(G) until the iteration times satisfied the expected number – and making inferences based on the remaining draws after discarding the burn-in period.

It is worth noting that the convergence speed of the algorithm can be improved further by grouping or blocking the random parameters $\{\boldsymbol{\beta}_{(g)i}\}$ and fixed parameters $\boldsymbol{\alpha}$ together and drawing from the joint posterior, in order to accelerate mixing of chains and reduce numerical standard errors associate with the estimates. In this case, sampling from the joint posterior density $p(\boldsymbol{\alpha}, \{\boldsymbol{\beta}_{(g)i}\}|\boldsymbol{\mu}_\beta, \mathbf{V}_\beta, \{z_{iq}\}, \mathbf{Y})$ can be decomposed into drawing sequentially from the $p(\{\boldsymbol{\beta}_{(g)i}\}|\boldsymbol{\alpha}, \boldsymbol{\mu}_\beta, \mathbf{V}_\beta, \{z_{iq}\}, \mathbf{Y})$ step and from the $p(\boldsymbol{\alpha}|\boldsymbol{\mu}_\beta, \mathbf{V}_\beta, \{z_{iq}\}, \mathbf{Y})$ step. Details relating to this approach can be found in Chib and Carlin (1999), and Frühwirth-Schnatter et al. (2004).

Finally, because sampled parameters correspond to an unknown hidden state, labeling of unobserved states switches between various labeling subspaces, potentially causing the unconstrained posterior to have at most $G!$ modes – thus introducing an identification problem and slow mixing of simulations. This is an important issue in the context of Bayesian estimation via MCMC simulation and has received considerable attention to identify a unique labeling scheme from $G!$ possible modes (Celeux et al., 2000; Stephens, 2000; Frühwirth-Schnatter, 2001).

Following Frühwirth-Schnatter (2001)'s work, an indirect permutation sampler to implement a unique labeling under identifiability constraints within the course of MCMC estimation can be applied. This approach is able to explore the full unconstrained parameter space and jump between various labeling subspaces by sampling from the unconstrained posterior. The steps of the sampling scheme are the same as (A)–(G), as shown above, augmented Gibbs sampling plus an additional step (H) which is specified as:

(H) Select some permutation $\zeta(1), \ldots, \zeta(G)$ of the current labeling, and apply it by substituting component specific parameters $(\boldsymbol{\mu}_{\beta 1}, \ldots, \boldsymbol{\mu}_{\beta G})$ with $(\boldsymbol{\mu}_{\beta\zeta(1)}, \ldots, \boldsymbol{\mu}_{\beta\zeta(G)})$, $(\mathbf{V}_{\beta 1}, \ldots, \mathbf{V}_{\beta G})$ with $(\mathbf{V}_{\beta\zeta(1)}, \ldots, \mathbf{V}_{\beta\zeta(G)})$, $(S_1, \ldots, S_i)$ with $(\zeta(S_1), \ldots, \zeta(S_i))$ and $(\pi_1, \ldots, \pi_G)$ with $(\pi_{\zeta(1)}, \ldots, \pi_{\zeta(G)})$.

The selection of the permutation can be made as: (1) randomly drawing one of the $G!$ possible permutations as $\zeta(1), \ldots, \zeta(G)$ which is performed when using Bayesian selection to choose the appropriate number of states for the mixture structure and explore unconstrained and unknown parameter spaces; and (2) permutation sampling under the identifiability constraint that subspaces are unconstrained in parameter space $\Re$ in subsets $\{\Re_n\}$. Since it is fully expected that the influences of guardian supervision on teen driving is likely to vary considerably, an ordering of a component-specific distribution mean of the guardian supervising indicator is used as the constraint to confine subspace $\Re_n$ as: $\mu_{\beta_{1,1}} > \mu_{\beta_{2,1}} > \cdots > \mu_{\beta_{G,1}}$, based on the invariance of priors. Thus, the parameters can be sampled from constrained posteriors $p_{\Re_n}(\bullet \mid Y)$ to make inferences.

References

Albert, J., Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association 88 (422), 669–679.

Allen, J., Brown, B., 2008. Teens, peers, and automobiles: the perfect storm. American Journal of Preventive Medicine 35 (Suppl. 3), 289–293.

Allenby, G., Lenk, P., 1994. Modeling household purchase behavior with logistic normal regression. Journal of the American Statistical Association 89 (428), 1219–1231.

Allenby, G., 1997. An introduction to hierarchical Bayesian modeling. Tutorial Notes, Advanced Research Techniques Forum, American Marketing Association.

Allenby, G., Arora, N., Ginter, J., 1998. On the heterogeneity of demand. Journal of Marketing Research 35 (3), 384–389.

Allenby, G., Rossi, P., 1999. Marketing models of consumer heterogeneity. Journal of Econometrics 89 (1–2), 57–78.

American Automobile Association, 2012. Distracted driving among newly licensed teen drivers. Foundation for Traffic Safety, Washington, DC. <http://www.aaafoundation.org/pdf/DistractedDrivingAmongNewlyLicensedTeenDrivers.pdf> (retrieved 14.09.12).

Anastasopoulos, P., Mannering, F., 2011. An empirical assessment of fixed and random parameter logit models using crash- and non-crash-specific injury data. Accident Analysis and Prevention 43 (3), 1140–1147.

Beullens, K., Van den Bulck, J., 2008. News, music videos and action movie exposure and adolescents' intentions to take risks in traffic. Accident Analysis and Prevention 40 (1), 349–356.

Bhat, C., 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. Transportation Research Part B 35 (7), 677–693.

Bhat, C., 2003. Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. Transportation Research Part B 37 (9), 837–855.

Bhat, C., 2011. The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. Transportation Research Part B 45 (7), 923–939.

Bhat, C., Sidharthan, R., 2011. A simulation evaluation of the maximum approximate composite marginal likelihood (MACML) estimator for mixed multinomial probit models. Transportation Research Part B 45 (7), 940–953.

Bhat, C., Sidharthan, R., 2012. A new approach to specify and estimate non-normally mixed multinomial probit models. Transportation Research Part B 46 (7), 817–833.

Bianchi, A., Summala, H., 2004. The "genetics" of driving behavior: parents' driving style predicts their children's driving style. Accident Analysis and Prevention 36 (4), 655–659.

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

Buckley, L., Chapman, R., Sheehan, M., 2012. Adolescent involvement in anti-social and delinquent behaviours: predicting future injury risk. Accident Analysis and Prevention 48, 518–522.

Bujosa, A., Riera, A., Hicks, R., 2010. Combining discrete and continuous representation of preference heterogeneity: a latent class approach. Environment and Resource Economics 4 (7), 477–493.

Celeux, G., Hurn, M., Robert, C., 2000. Computational and inferential difficulties with mixture posterior distributions. Journal of the American Statistician Association 95 (451), 957–970.

Centers for Disease Control and Prevention, 2009. Youth Risk Behavior Surveillance—United States, 2007. National Center for Chronic Disease Prevention and Health Promotion (producer). <www.cdc.gov/mmwr/preview/mmwrhtml/ss5704a1.htm> (retrieved 13.09.12).

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

Chen, L., Baker, S., Braver, E., Li, G., 2000. Carrying passengers as a risk factor for crashes fatal to 16- and 17-year old drivers. Journal of the American Medical Association 283 (12), 1578–1582.

Chib, S., 1992. Bayes inference in the Tobit censored regression model. Journal of Econometrics 51 (1–2), 79–99.

Chib, S., Carlin, B., 1999. On MCMC sampling in hierarchical longitudinal models. Statistics Computing 9 (1), 17–26.

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

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

Dols, S., González, F., Aleixandre, N., Vidal-Infer, A., Rodrigo, M., Valderrama-Zurián, J., 2010. Predictors of driving after alcohol and drug use among adolescents in Valencia (Spain). Accident Analysis and Prevention 42 (6), 2024–2029.

Eluru, N., Bagherib, M., Miranda-Morenoa, L., Fu, L., 2012. A latent class modeling approach for identifying vehicle driver injury severity factors at highway–railway crossings. Accident Analysis and Prevention 47, 119–127.

Ericsson, K., 2005. Recent advances in expertise research: a commentary on the contributions to the special issue. Applied Cognitive Psychology 19 (2), 223–241.

Ferguson, S., Williams, A., Chapline, J., Reinfurt, D., De Leonardis, D., 2001. Relationship of parent driving records to the driving records of their children. Accident Analysis and Prevention 33 (2), 229–234.

Frühwirth-Schnatter, S., 2001. Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. Journal of the American Statistical Association 96 (453), 194–209.

Frühwirth-Schnatter, S., Tüchler, R., Otter, T., 2004. Bayesian analysis of the heterogeneity model. Journal of Business and Economic Statistics 22 (1), 2–15.

Gelman, A., Rubin, D., 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7 (4), 457–472.

Gelman, A., Carlin, J., Stern, H., Rubin, D., 2004. Bayesian Data Analysis, second ed. Chapman and Hall, London.

Geweke, J., Keane, M., 2007. Smoothly mixing regressions. Journal of Econometrics 138 (1), 252–290.

Greene, W., Hensher, D., 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. Transportation Research Part B 37 (8), 681–698.

Greene, W., Hensher, D., 2013. Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model. Applied Economics 45 (14), 1897–1902.

Hafetz, J., Jacobsohn, L., García-España, J., Curry, A., Winston, F., 2010. Adolescent drivers' perceptions of the advantages and disadvatages of absention from in-vehicle cell phone use. Accident Analysis and Prevention 42 (6), 1570–1576.
Halpern-Felsher, B., Millstein, S., Ellen, J., Adler, N., Tschann, J., Biehl, M., 2001. The role of behavioral experience in judging risks. Health Psychology 20 (2), 120–126.
Hensher, D., Greene, W., 2003. Mixed logit models: state of practice. Transportation 30 (2), 133–176.
Huber, J., Train, K., 2001. On the similarity of classical and Bayesian estimates of individual mean partworths. Marketing Letters 12 (3), 257–267.
Insurance Institute for Highway Safety, 2009. Fatality facts: Teenagers 2008. The Institute, Arlington (VA). <www.iihs.org/research/fatality_facts_2008/teenagers.html> (retrieved 13.09.12).
Koop, G., Poirier, D., Tobias, J., 2007. Bayesian Econometric Methods. Cambridge Univ. Press, Cambridge, UK.
Lenk, P., DeSarbo, W., 2000. Bayesian inference for finite mixture of generalized linear models with random effects. Psychometrika 65 (1), 93–119.
Li, M., Tobias, J., 2007. Bayesian analysis of treatment effects in an ordered potential outcomes model. Advances in Econometrics 21, 57–91.
Liu, J., Wu, Y.-N., 1999. Parameter expansion for data augmentation. Journal of the American Statistical Association 94 (488), 1264–1274.
Mannering, F., Washburn, S., 2013. Principles of Highway Engineering and Traffic Analysis, fifth ed. John Wiley and Sons, New York, NY.
McCulloch, R., Rossi, P., 1994. An exact likelihood analysis of the multinomial probit model. Journal of Econometrics 64 (1–2), 207–240.
McFadden, D., 1989. A method of simulated moments for estimation of discrete response models without numerical integration. Econometrica 57 (5), 995–1026.
Millstein, S., Halpern-Felsher, B., 2002. Perceptions of risk and vulnerability. Journal of Adolescent Health 31 (1), 10–27.
Nagle, M., 2011. Effects of graduated driver licensing on crash outcomes in Indiana. Indiana University Public Policy Institute, Indianapolis. Document ID: 11-C01. <http://policyinstitute.iu.edu/PubsPDFs/TrafficBrief_GDL2011_Final.pdf> (retrieved 01.10.12).
Nandram, B., Chen, M.-H., 1996. Reparameterizing the generalized linear model to accelerate Gibbs sampler convergence. Journal of Statistical Computation and Simulation 54 (1–3), 129–144.
National Highway Traffic Safety Administration, 2008a. Statistics and Facts about Distracted Driving. Based on the 2008 National Occupant Protection Use Survey. <www.distraction.gov/stats-and-facts/#what> (retrieved 13.09.12).
National Highway Traffic Safety Administration, 2008b. Safety Belt Use in 2008—Demographic Results. Based on the 2008 National Occupant Protection Use Survey. <http://www-nrd.nhtsa.dot.gov/Pubs/811183.PDF> (retrieved 13.09.12).
Olaru, D., Smith, B., Taplin, J., 2011. Residential location and transit-oriented development in a new rail corridor. Transportation Research Part A 45 (3), 219–237.
Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimators. Econometrica 57 (5), 1027–1057.
Paleti, R., Eluru, N., Bhat, C., 2010. Examining the influence of aggressive driving behavior on driver injury severity in traffic crashes. Accident Analysis and Prevention 42 (10), 1839–1854.
Park, B., Zhang, Y., Lord, D., 2010. Bayesian mixture modeling approach to account for heterogeneity in speed data. Transportation Research Part B 44 (5), 662–673.
Preusser, D., Ferguson, S., Williams, A., 1998. The effect of teenage passengers on the fatal crash risk of teenage drivers. Accident Analysis and Prevention 30 (2), 217–222.
Shope, J., Waller, P., Lang, S., 1996. Alcohol-related predictors of adolescent driving: gender differences in crashes and offenses. Accident Analysis and Prevention 28 (6), 755–764.
Shen, J., 2009. Latent class model or mixed logit model? A comparison by transport mode choice data. Applied Economics 41 (22), 2915–2924.
Shults, R., 2010. Drivers Aged 16 or 17 Years Involved in Fatal Crashes – United States, 2004–2008. Prepared for Div. of Unintentional Injury Prevention, National Center for Injury Prevention and Control, CDC. <http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5941a2.htm?s_cid=mm5941a2_w> (retrieved 13.09.12).
Simons-Morton, B., 2007. Parent involvement in novice teen driving: rationale, evidence of effects, and potential for enhancing graduated driver licensing effectiveness. Journal of Safety Research 38 (2), 193–202.
Simons-Morton, B., Lerner, N., Singer, J., 2005. The observed effects of teenage passengers on the risky driving behavior of teenage drivers. Accident Analysis and Prevention 37 (6), 973–982.
Spiegelhalter, D., Best, N., Carlin, B., Van der Linde, A., 2002. Bayesian measures of model complexity and fit (with discussion). Journal of the Royal Statistical Society Series B 64 (4), 583–639.
Stephens, M., 2000. Bayesian analysis of mixture models with an unknown number of components – an alternative to reversible jump methods. The Annuals of Statistics 28 (1), 40–74.
Tanner, M., Wong, W., 1987. The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association 82 (398), 528–549.
Verbeke, G., Lesaffre, E., 1996. A linear mixed-effects model with heterogeneity in the random-effects population. Journal of the American Statistical Association 91 (433), 217–221.
Washington, S., Karlaftis, M., Mannering, F., 2011. Statistical and Econometric Methods for Transportation Data Analysis. Chapman and Hall/CRC, Boca Raton, FL.
Williams, A., Preusser, D., Ulmer, R., Weinstein, H., 1995. Characteristics of fatal crashes of 16-year-old drivers: implications for licensure policies. Journal of Public Health Policy 16 (3), 347–360.
Williams, A., Ferguson, S., Shope, J., 2002. Rationale for graduated licensing and the risks it should address. Injury Prevention 8 (Suppl. 2), ii9–ii16.
Wilson, R., Meckle, W., Wiggins, S., Cooper, P., 2006. Young driver risk in relation to parents' retrospective driving record. Journal of Safety Research 37 (4), 325–332.
Winston, C., Maheshri, V., Mannering, F., 2006. An exploration of the offset hypothesis using disaggregate data: the case of airbags and antilock brakes. Journal of Risk and Uncertainty 32 (2), 83–99.
Wong, K., Wong, S., Yang, H., Wu, J., 2008. Modeling urban taxi services with multiple user classes and vehicle modes. Transportation Research Part B 42 (10), 985–1007.
Xiong, Y., Tobias, J., Mannering, F., 2013. A Markov switching random parameter model for heterogeneous panel analysis of vehicle crash severity data. Working paper.
Yau, K., Lee, A., Ng, A., 2003. Finite mixture regression model with random effects: application to neonatal hospital length of stay. Computational Statistics and Data Analysis 41 (28), 359–366.