Accident Analysis and Prevention 196 (2024) 107446

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Efficient and robust estimation of single-vehicle crash severity: A mixed logit model with heterogeneity in means and variances

Zhenning Li [a,*,1], Chengyue Wang [b,1], Haicheng Liao [b], Guofa Li [c], Chengzhong Xu [b]

[a] State Key Laboratory of Internet of Things for Smart City and Departments of Civil and Environmental Engineering and Computer and Information Science, University of Macau, Macao Special Administrative Region of China

[b] State Key Laboratory of Internet of Things for Smart City and Department of Computer and Information Science, University of Macau, Macao Special Administrative Region of China

[c] College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing, China

ARTICLE INFO

Keywords:
Mixed logit
Single vehicle crash
Heterogeneity in means and variances
Halton sequence

ABSTRACT

This study delves into the factors that contribute to the severity of single-vehicle crashes, focusing on enhancing both computational speed and model robustness. Utilizing a mixed logit model with heterogeneity in means and variances, we offer a comprehensive understanding of the complexities surrounding crash severity. The analysis is grounded in a dataset of 39,788 crash records from the UK's STATS19 database, which includes variables such as road type, speed limits, and lighting conditions. A comparative evaluation of estimation methods, including pseudo-random, Halton, and scrambled and randomized Halton sequences, demonstrates the superior performance of the latter. Specifically, our estimation approach excels in goodness-of-fit, as measured by ρ^2, and in minimizing the Akaike Information Criterion (AIC), all while optimizing computational resources like run time and memory usage. This strategic efficiency enables more thorough and credible analyses, rendering our model a robust tool for understanding crash severity. Policymakers and researchers will find this study valuable for crafting data-driven interventions aimed at reducing road crash severity.

1. Introduction

Single-vehicle crashes pose a significant public health concern and are associated with a high rate of severe or even fatal injuries compared to other types of vehicle crashes (Yu et al., 2021a; Se et al., 2021b). While various countermeasures have been implemented to reduce the incidence and severity of these crashes, the challenge remains complicated and multifaceted (Fisa et al., 2022). This underscores the pressing need for advanced and effective methodologies to rigorously examine the factors influencing the severity of single-vehicle crashes.

Utilizing police-reported crash records has been a cornerstone in studying the causes and influencing factors of driver injury severity in single-vehicle crashes (Ding et al., 2023; Alogaili and Mannering, 2022). Discrete choice models, such as logit and probit models, are frequently employed to analyze these data, given their suitability for the discrete nature of crash outcome data (Behnood and Mannering, 2017; Li et al., 2019a). Through the dedicated efforts of numerous researchers, various factors have been identified as having a significant impact on driver injury severity in police-reported single-vehicle crashes, ranging from vehicle types and road conditions to driver behavior and characteristics (Fountas et al., 2020; Yu et al., 2021b; Song et al., 2020). For instance, in a study investigating single-vehicle crashes in Thailand from 2011 to 2017, both young and older drivers were found to contribute to increased injury severity. This analysis utilized uncorrelated and correlated mixed logit models with heterogeneity in means and variances to assess the impact of driver age on crash outcomes (Se et al., 2021b). Additionally, another study employing a zero-inflated hierarchical ordered probit approach with correlated disturbances revealed that older vehicles and those with engine capacities of 1800cc or greater tend to exacerbate injury severity. This study focused on the joint effects of weather and lighting conditions on police-reported crash severities (Fountas et al., 2020). Moreover, factors such as alcohol consumption, curved roadways, and wet or waterlogged surfaces were found to consistently influence injury severity in crashes. This conclusion was drawn from an analysis using a random thresholds random parameters hierarchical ordered probit (HOPIT) approach, applied to police-reported crash data collected from 2014 to 2017 in the State of North Carolina (Yu et al., 2021b).

While police-reported data make significant contributions to traffic safety research, they have limitations in scope and detail. These

* Corresponding author.
E-mail addresses: zhenningli@um.edu.mo (Z. Li), czxu@um.edu.mo (C. Xu).

[1] Equally contributed.

https://doi.org/10.1016/j.aap.2023.107446
Received 27 September 2023; Received in revised form 16 November 2023; Accepted 23 December 2023
Available online 28 December 2023
0001-4575/

datasets often lack essential information, such as specific behavioral attributes and physiological characteristics of drivers, crucial for fully understanding crash outcomes (Mannering et al., 2016). Key factors like momentary driver distractions, fatigue levels, and cognitive impairments, which can greatly influence crash dynamics, are usually not included in these reports. Additionally, physiological aspects such as reflex response times, visual acuity, and overall health conditions, crucial in determining a driver's reaction during crashes, are often not recorded. This omission of behavioral and physiological data can lead to significant biases in statistical models, risking skewed or incomplete findings. This gap in data not only affects the accuracy of crash severity assessments but also hinders the development of effective safety interventions (Islam et al., 2020).

To address these challenges, contemporary research has turned to sophisticated modeling techniques that incorporate both fixed and random effects, capturing unobserved heterogeneity among observations. Notable methods include finite mixture random parameters (Li et al., 2018), random thresholds random parameters hierarchical ordered probit (Yu et al., 2021b), and mixed logit models with heterogeneity in means and variances (Ahmed et al., 2023). For instance, Li et al. (2018) utilized a finite mixture random parameters model to unravel factors affecting driver injury severity in low visibility-related crashes. Hamed and AlShaer (2023) employed a mixed logit model with heterogeneity in means and variances, offering a comprehensive analysis of driver crash involvement over time, which revealed distinct crash risk patterns across genders and generations, as well as variations in driver behavior post-crashes. Further, Alnawmasi and Mannering (2022) applied a mixed logit model with heterogeneity in means and variances to explore the effects of increased speed limits on single-vehicle highway crashes, using data spanning periods before and after speed limit changes. Their findings, consistent with other studies (Ahmed et al., 2023), underscored the model's effectiveness.

Building on the previous discussion, the implementation of advanced statistical models such as the mixed logit with heterogeneity in means and variances is particularly pertinent in the context of single-vehicle crash analysis (Al-Bdairi et al., 2020; Se et al., 2021b). These models offer a sophisticated framework for understanding the multifaceted nature of crash dynamics, especially in cases where conventional datasets like police-reported data fall short (Li et al., 2021). The mixed logit model's strength lies in its ability to incorporate a diverse range of variables, from observable factors like vehicle type and road conditions to more subtle, often unrecorded factors like driver behavioral tendencies and physiological states. For instance, by integrating variables related to driver attention and response times, the model can provide insights into how these unobserved factors might influence crash severity. This is particularly crucial in single-vehicle crashes, where driver behavior plays a more pronounced role (Zamani et al., 2021). Given the consistently high performance of the mixed logit model with heterogeneity in means and variances, our research adopts this framework. This decision allows for a detailed examination of factors influencing single-vehicle crash severity, utilizing the model's established analytical strength and depth (Huo et al., 2020).

While the mixed logit model with heterogeneity in means and variances holds promise for providing a nuanced understanding of the factors influencing single-vehicle crash severity, it is not without challenges. The first hurdle is computational: the model demands significant resources, especially when handling large datasets (Hou et al., 2022). This often necessitates specialized software or a high-performance computing setup for efficient analysis. Another challenge lies in the risk of introducing bias or inaccuracies. For example, if the model's assumptions about the distribution of unobserved factors are incorrect, the results could be skewed. Additionally, if important variables affecting crash severity are left out, the model's insights may fall short of capturing the full picture (Alogaili and Mannering, 2020).

While the mixed logit model with heterogeneity in means and variances offers a comprehensive framework for analyzing single-vehicle crash severity, it poses several challenges that need careful consideration. Firstly, there is a significant computational demand. Analyzing large datasets with this model requires considerable computational power and resources, often necessitating the use of advanced software and high-performance computing environments (Hou et al., 2022). This can pose a barrier, especially for research teams with limited access to such resources. Another challenge is the potential for bias or inaccuracies in the model's outcomes. If the assumptions regarding the distribution of unobserved heterogeneity are not aligned with the actual underlying patterns, the model could produce misleading results. The accuracy of the mixed logit model heavily relies on these assumptions, making it crucial to critically evaluate and justify the chosen distributions (Alogaili and Mannering, 2020). Moreover, the model's effectiveness is contingent on the inclusion of all relevant variables that impact crash severity. Omitting crucial variables, either due to data limitations or oversight, can lead to incomplete analyses. This is particularly pertinent in traffic safety research, where factors such as environmental conditions, vehicle characteristics, and driver behavior play critical roles (Hou et al., 2022).

To navigate the intricacies of single-vehicle crash severity analysis, our study adopts an advanced computational approach, utilizing a mixed logit model with heterogeneity in means and variances. This sophisticated model is capable of integrating a wide array of influential factors, encompassing both observed and typically unobserved variables. Such comprehensive coverage is crucial for accurately identifying and understanding the diverse elements that contribute to crash severity. Incorporating unobserved heterogeneity, the model accounts for individual-specific effects and random variations that standard logistic regression models might overlook. This allows for a more nuanced analysis, acknowledging that factors like driver behavior, vehicle condition, and environmental influences can vary significantly across crashes and individuals. By doing so, the model provides a richer, more detailed picture of the underlying causes of crash severity. To enhance computational efficiency and accuracy, our estimation procedures incorporate scrambled and randomized Halton draws. These techniques are designed to optimize the efficiency of numerical integration in the mixed logit model. Scrambled and randomized Halton sequence, as a low-discrepancy quasi-random number sequence, offers better coverage of the multidimensional space compared to simple random sampling. This is particularly beneficial in models with a high number of random parameters, as it ensures a more thorough exploration of the parameter space with fewer iterations, thereby reducing computational time without compromising the robustness of the model's estimations. The ultimate objective of employing these advanced modeling and computational techniques is to derive meaningful insights that can inform policy and practice. By accurately identifying the key factors influencing single-vehicle crash severity and understanding their interplay, the study aims to provide valuable evidence-based recommendations. These insights are intended to guide policymakers, road safety planners, and other stakeholders in developing more effective and targeted safety interventions. The goal is to translate complex analytical findings into practical strategies that can enhance road safety, reduce the frequency and severity of crashes, and ultimately save lives.

The remainder of this paper is organized to facilitate understanding and engagement. Section 2 offers an in-depth look into the data sources and variables that serve as the backbone of this study. Section 3 details the modeling approaches employed, describing the techniques and methodologies used to delve into the intricacies of single-vehicle crash severity. Section 4 discusses the computational strategies for efficient and robust parameter estimation. Section 5 unveils the experimental findings, interpreting the results and their broader implications. Finally, Section 6 concludes the paper by summarizing the key takeaways and suggesting avenues for future research and intervention strategies.

2. Data

This study utilizes data from the STATS19 database, an official repository for road traffic casualty information in Great Britain. Access and processing were facilitated through the R package `stats19`. Managed by the Department for Transport, STATS19 provides a rich set of variables capturing the nuances of road traffic crashes that result in personal injury. Its structured format is one of its key strengths, streamlining the data collection process and making the database well-suited for statistical analysis. This structured approach enables the identification of long-term patterns and trends in road safety. Moreover, the data in STATS19 is police-reported, ensuring a high level of accuracy and reliability.

In our post-processing analysis, we meticulously examined a subset of 39,788 records documenting single-vehicle crashes that occurred from January 1, 2017, to December 31, 2019. The injury severity in these incidents was categorized as fatal (2.7%), serious (26.4%), and slight (70.9%). The relatively small proportion of fatal outcomes can be attributed to several factors. Firstly, advancements in vehicle safety technology and improved road safety measures have contributed to reducing the likelihood of crashes resulting in fatalities. Additionally, emergency response and medical intervention have become more efficient, often preventing serious injuries from escalating to fatal outcomes. The dataset encompasses a wide array of explanatory variables spanning multiple categories—crucial for a holistic analysis of crash circumstances. These categories include crash specifics, vehicle features, and driver demographics. Key variables such as the time and location of the crash, road type and speed limit, lighting and weather conditions, road surface quality, vehicle type, as well as the driver's gender and age, are considered. Each of these variables offers valuable insights into the crash context and can significantly influence the severity of the outcome. A detailed summary of these variables, complete with descriptive statistics, is presented in Table 1.

3. Mixed logit model with heterogeneity in means and variances

Our modeling begins with the classical multinomial logit approach. Suppose that the driver n suffered a i-th ($i = 1, \ldots, I$) level of injury severity during the crash. Let S_{in} denote the corresponding injury-severity function, which is given as

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

where $\boldsymbol{X}_{in} \in \mathbb{R}^p$ is the vector of explanatory variables that affect the injury severity of the driver n, $\boldsymbol{\beta}_i$ is the vector of the corresponding estimable parameters for injury severity i, and ε_{in} is the error term. A common way in the field of traffic safety analysis is to assume that the error term is distributed in generalized extreme values, leading to the widely used multinomial logit model (McFadden, 1981), that is,

$$P_n(i) = \frac{\exp(\boldsymbol{\beta}_i \boldsymbol{X}_{in})}{\Sigma_{i=1}^{I} \exp(\boldsymbol{\beta}_i \boldsymbol{X}_{in})} \tag{2}$$

where $P_n(i)$ is the probability that the driver n may suffer i-th level of injury severity. Given this setting, the parameters $\boldsymbol{\beta}_i$ are allowed to vary between different injury severities but are fixed between observations. In order to account for unobserved heterogeneity, the above model can be relaxed by assuming the parameters $\boldsymbol{\beta}_i$ to vary across observations with specified distributions, for example, in the most widely used mixed logit model, $\boldsymbol{\beta}_{in}$ is assumed to follow a multivariate normal distribution $N(\boldsymbol{\beta}_{in}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (Li et al., 2019a). A further refinement to the model is to allow the mean and variance of $\boldsymbol{\beta}_{in}$ to vary across the observations instead of only having a predefined distributional form. Then, this more general setting of the random parameters $\boldsymbol{\beta}_{in}$ can be given as

$$\boldsymbol{\beta}_{in} = \boldsymbol{\beta}_i + \boldsymbol{\delta}_i^T \boldsymbol{z}_{in} + \boldsymbol{\sigma}_i \exp(\boldsymbol{\omega}_i^T \boldsymbol{w}_{in}) \boldsymbol{v}_{in} \tag{3}$$

where $\boldsymbol{\beta}_n$ is the vector of the constant term which is fixed across all drivers who got involved in the i-th level of injury severity, $\boldsymbol{z}_{in}$ is

Table 1
Summary of variables and descriptive statistics.

Variable	Count	Mean	Standard deviation
Severity			
Slight[a]	28 221	0.709	0.454
Serious	10 504	0.264	0.441
Fatal	1063	0.027	0.161
Year			
2017[a]	14 560	0.366	0.482
2018	12 972	0.326	0.469
2019	12 256	0.308	0.462
Week			
Monday	5289	0.133	0.340
Tuesday	5200	0.131	0.337
Wednesday[a]	5207	0.131	0.337
Thursday	5381	0.135	0.342
Friday	5815	0.146	0.353
Saturday	6278	0.158	0.365
Sunday	6618	0.166	0.372
Road type			
Dual Carriageway	6814	0.171	0.377
One Way Street	570	0.014	0.119
Roundabout	1987	0.050	0.218
Single Carriageway[a]	29 726	0.747	0.435
Slip Road	691	0.017	0.131
Speed limit			
≤30 mph[a]	16 341	0.411	0.492
40–50 mph	6211	0.156	0.363
≥60 mph	17 236	0.433	0.496
Lighting condition			
Darkness with light	7521	0.189	0.392
Darkness without light	8189	0.206	0.404
Dawn	402	0.010	0.100
Daylight[a]	23 676	0.595	0.491
Weather condition			
Fine with high winds	505	0.013	0.112
Fine with no high winds[a]	29 990	0.754	0.431
Fog or mist	405	0.010	0.100
Snow or rain	7655	0.192	0.394
Other	1233	0.031	0.173
Road surface			
Dry[a]	23 154	0.582	0.493
Ice or snow	2310	0.058	0.234
Wet	14 324	0.360	0.480
Area			
Rural[a]	24 794	0.623	0.485
Urban	14 994	0.377	0.485
Vehicle type			
Bus	165	0.004	0.064
Car[a]	26 641	0.670	0.470
Truck	657	0.017	0.127
Motorcycle	12 325	0.310	0.462
Maneuver			
Lane Changing	463	0.012	0.107
Going Straight[a]	34 851	0.876	0.330
Overtaking	842	0.021	0.144
Slowing	1425	0.036	0.186
Turning Left	1095	0.028	0.164
Turning Right	1021	0.026	0.158
U Turn	91	0.002	0.048
Gender			
Female	11 171	0.281	0.449
Male[a]	28 617	0.719	0.449

(continued on next page)

Table 1 (*continued*).

Age			
≤25	6048	0.152	0.359
26–45[a]	15 776	0.397	0.489
46–65	13 922	0.350	0.477
>65	4042	0.102	0.302

Note:
[a] The base variables in the model.

the vector of observed variables that measure the heterogeneity in the means, $\boldsymbol{\delta}_i^T$ is the corresponding vector of estimable parameters, $\boldsymbol{w}_{in}$ is the vector of observed variables that measure the heterogeneity in the standard deviation $\boldsymbol{\sigma}_i$ with corresponding parameter vector $\boldsymbol{\omega}_i^T$, and $\boldsymbol{v}_{in}$ are vectors of randomly distributed unobserved terms with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{I}$. With the above settings, the means and standard deviations of the random parameters $\boldsymbol{\beta}_{in}$ are $\boldsymbol{\beta}_n + \boldsymbol{\delta}_i^T \boldsymbol{z}_{in}$ and $\boldsymbol{\sigma}_i \exp(\boldsymbol{\omega}_i^T \boldsymbol{w}_{in})$, respectively. Assume $P_n(i|\boldsymbol{v}_{in})$ is the probability that the driver n suffered i-th level of injury severity conditioned on $\boldsymbol{\omega}_i^T$, then the corresponding unconditional probability, $P_n(i)$, is

$$P_n(i) = \int_{\boldsymbol{v}_{in}} P_n(i|\boldsymbol{v}_{in}) \cdot f(\boldsymbol{v}_{in}) \mathrm{d}\boldsymbol{v}_{in} = \int_{\boldsymbol{v}_{in}} \frac{\exp(\boldsymbol{\beta}_{in}\boldsymbol{X}_{in})}{\sum_{i=1}^{I} \exp(\boldsymbol{\beta}_{in}\boldsymbol{X}_{in})} \cdot f(\boldsymbol{v}_{in}) \mathrm{d}\boldsymbol{v}_{in} \quad (4)$$

where $f(\boldsymbol{v}_{in})$ is the density function of $\boldsymbol{v}_{in}$. Eqs. (3) and (4) denote the (uncorrelated) mixed logit model with heterogeneity in means and variances, which can be regarded as a generalization of the multinomial logit model. There are several possible degradations of the model, including (1) fixed parameters multinomial logit model (when $\boldsymbol{\delta}_i^T = \mathbf{0}$, $\boldsymbol{\sigma}_i = \mathbf{0}$, and $\boldsymbol{\omega}_i^T = \mathbf{0}$), (2) mixed logit model (without mean–variance heterogeneity) (when $\boldsymbol{\delta}_i^T = \mathbf{0}$, $\boldsymbol{\sigma}_i \neq \mathbf{0}$, and $\boldsymbol{\omega}_i^T = \mathbf{0}$), and (3) mixed logit model with heterogeneity in means (when $\boldsymbol{\delta}_i^T \neq \mathbf{0}$, $\boldsymbol{\sigma}_i \neq \mathbf{0}$, and $\boldsymbol{\omega}_i^T = \mathbf{0}$).

The methodology adopted in this study reflects a meticulous approach to model building, emphasizing the principle of parsimony to balance the simplicity and explanatory power of the statistical model. Our model selection was informed by a stringent threshold of 95% confidence for statistical significance, a standard that is widely accepted in statistical analyses. This threshold was deliberately chosen to focus our attention on variables with a meaningful impact on crash severity. We incorporated robust statistical testing methods, such as the analysis of p-values and confidence intervals, to rigorously evaluate each variable's contribution to the model. This testing played a pivotal role in identifying the most influential factors, ensuring that our model retained only those elements that significantly affect crash severity.

4. Efficient and robust parameter estimation

4.1. Simulated maximum likelihood for model estimation

The classical framework for model estimation is illustrated by multinational logit models. The log-likelihood, $\mathcal{L}$, is given by:

$$\mathcal{L} = \Sigma_{n=1}^{N} \Sigma_{i=1}^{I} y_{in} \ln P_n(i) \quad (5)$$

where $y_{in} = 1$ if the driver n got involved in an i-th level of injury and 0 otherwise. Let $\boldsymbol{\Theta} = \{\boldsymbol{\beta}_i, \boldsymbol{\delta}_i^T, \boldsymbol{\sigma}_i, \boldsymbol{\omega}_i^T\}$ denote the set of all parameters to be estimated, and they can be estimated by maximizing Eq. (5) with respect to the parameters by finding $\hat{\theta} = \arg\max_{\theta \in \boldsymbol{\Theta}} \mathcal{L}(\theta; \boldsymbol{y})$, where θ represents any element in $\boldsymbol{\Theta}$, and $\boldsymbol{y}$ denote the observed data. As the article of McFadden (1974) demonstrated, the log-likelihood function is globally concave for linear parametric utility models with fixed parameters. It means that there is always a global solution that the optimization algorithm is expected to arrive at when minimizing the negative log-likelihood of a preference space model with fixed parameters. Therefore, $P_n(i)$ in Eq. (2) can be used directly to estimate parameters in the fixed parameters multinomial logit model.

However, the mixed logit models have a non-convex log-likelihood function and therefore there is no guarantee that a global solution will be reached. In the literature, many applications estimate the model using the simulated maximum likelihood (SML) method due to its ease of implementation and compatibility with various statistical software platforms. Moreover, it can accommodate complex models with many parameters and address situations where data may exhibit correlation or heteroscedasticity.

A common way of SML is to use random draws to simulate $P_n(i)$ in Eq. (4) (Train, 2009). Over a series of iterations, parameters are drawn from $f(\boldsymbol{v}_{in})$ and used to compute the conditioned probability $P_n(i|\boldsymbol{v}_{in})$. The average probabilities over all the iterations, $\bar{P}_n(i|\boldsymbol{v}_{in})$, are then used in place of $P_n(i|\boldsymbol{v}_{in})$ in Eqs. (4) and (5) to compute the simulated log-likelihood. Formally, this simulation is given as

$$P_n(i) \approx E\left(\bar{P}_n(i|\boldsymbol{v}_{in})\right) \approx \frac{1}{R} \Sigma_{r=1}^{R} P_n(i|\boldsymbol{v}_{in,r}) \quad (6)$$

where R is the total number of draws of $\boldsymbol{v}_{in}$, and $\boldsymbol{v}_{in,r}$ is the r–th draw for $\boldsymbol{v}_{in}$. Interested readers are referred to the article by Train (2009) for a more detailed introduction and discussion.

The employment of the SML approach for estimating mixed logit models, despite its inherent computational complexity, is widespread. Its utility stems from the capacity to accommodate unobserved heterogeneity and to enhance the accuracy of parameter estimates. It should be duly noted, however, that simulation error is an inevitable byproduct of the simulation process, which hinges on the quantity and type of draws employed. The simulation of the log-likelihood function's value enables researchers to factor in the inherent simulation error stemming from the number and type of draws utilized. While the adoption of different sets of draws or alterations in the explanatory variable order may lead to disparate estimation results in the log-likelihood function, parameter estimates, standard errors, and their corresponding z-statistics, it is crucial to comprehend the extent of such variation and the number of draws employed in the estimation process to ensure accuracy and reliability (Train, 2009). Fig. 1 illustrates that as the number of draws increases, the magnitude of simulation errors tends to decline.

4.2. Scrambled and randomized Halton sequence

Quasi-Monte Carlo (QMC) techniques, exemplified by the Halton sequence, have pervasively emerged as a tool to reduce simulation-induced variability in statistical model outcomes. Such approaches have demonstrated superiority over pseudorandom methods, with initial investigations revealing that 100 Halton draws could yield parameter estimates with smaller bias and standard deviation than 1000 pseudorandom draws (Train, 2000; Bhat, 2001). This is attributable to the deterministic nature of QMC methods that generates a fixed sequence of points within the integration domain, in contrast to the randomness inherent in Monte Carlo methods. Consequently, QMC-based approaches can achieve higher accuracy with fewer function evaluations, leading to more precise estimates. Following Kocis and Whiten (1997), the Halton sequence in relatively prime bases $b_1, b_2, \dots, b_s$ is defined as sequence

$$x_n = \left(\boldsymbol{\Phi}_{b_1}(n), \dots, \boldsymbol{\Phi}_{b_j}(n), \dots, \boldsymbol{\Phi}_{b_s}(n)\right) \quad (7)$$

where $\boldsymbol{\Phi}_{b_j}(n)$ is the jth radical inverse function:

$$\boldsymbol{\Phi}_{b_j}(n) = \Sigma_{i=0}^{\infty} \alpha_i(j,n) b_j^{-i-1} \quad (8)$$

where $\alpha_i(j,n) \in [0, b_j)$ and is an integer obtained from digit expansion of n in base b_j:

$$n = \Sigma_{i=0}^{\infty} \alpha_i(j,n) b_j^{i} \quad (9)$$

While subsequent research has shown that the differences between the Halton sequence and pseudorandom methods may not be as significant, the Halton sequence as well as other QMC methods have remained the preferred choice for many modelers due to their consistent superior performance. However, there are limitations to the use of the Halton sequence, including the inability to evaluate error through variance

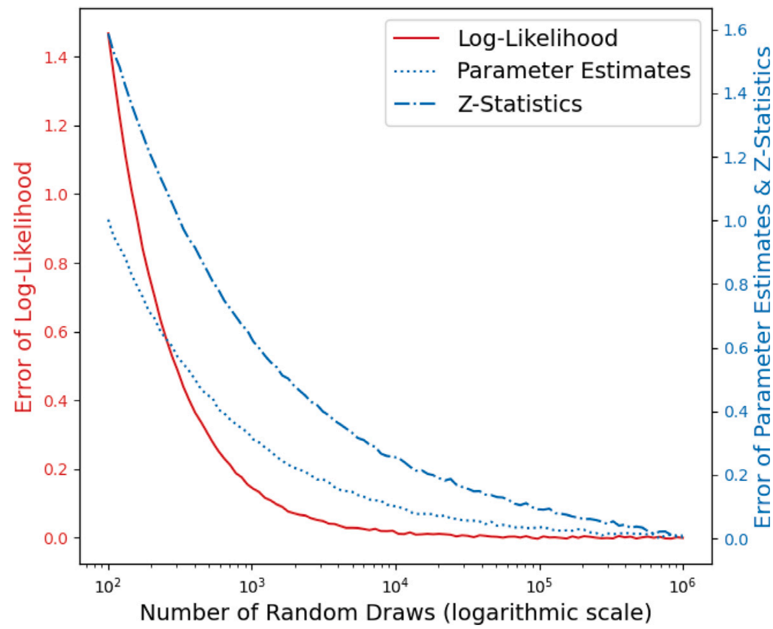

Fig. 1. Simulation error of log-likelihood function value, parameter estimates, and Z-statistics for 10^2 – 10^6 pseudo-random draws (Generated using the yogurt data set from Jain et al. (1994), which comprises 2412 choice observations from a two-year study of yogurt purchases by a panel of 100 households in Springfield, Missouri.).

analysis and the "curse of dimensionality" where the coverage of the integration domain decreases quickly as the dimensionality increases. One problem with using the Halton sequence is that, by definition, it is purely deterministic, and therefore it is not possible to evaluate the error by applying variance analysis, as in classical Monte Carlo simulations. Another problem is its poor performance in higher dimensions because the sequences generated using high prime numbers as bases tend to be highly correlated. A possible way to address this issue is to use scrambled Halton draws instead of the classical Halton sequences, which can deal with the "curse of dimensionality" issue by randomly permuting the digits in the sequence (Bhat, 2003). Formally, the scrambled technique, in other words, applying a generalized radical inverse function, can be given as

$$\Phi_{b_j}(n) = \Sigma_{i=0}^{\infty}\sigma\left(\alpha_i(j,n)\right) b_j^{-i-1} \tag{10}$$

where $\sigma(\cdot)$ is an operator of permutations on $\alpha_i(j,n)$. One of the most common choices for σ in the literature is the reverse Radix algorithm (Braaten and Weller, 1979), which we adopted in this study. This algorithm scrambles the coefficients by reversing their digits in the base representation. The algorithm consists of four steps: (1) Convert the index to the base representation for each dimension, (2) Reverse the order of the digits in each base representation, (3) Divide each reversed number by the base to the power of its length, and (4) The resulting numbers are the coordinates of the point. For instance, consider a three-dimensional Halton point set with base 2, 3, and 5 for each dimension. The first four points are: [0 0 0], [0.5 0.3333 0.2], [0.25 0.6667 0.4], [0.75 0.1111 0.6]. Applying reverse-radix scrambling to this point set (with a reverse permutation for base 5) yields: [0 0 0], [0.5 0.3333 0.8], [0.25 0.6667 0.6], [0.75 0.1111 0.4]. This technique can lower the correlation among different dimensions of the sequence and prevent some undesirable patterns.

Reverse-radix scrambling reduces correlation within the sequence's dimensions, making it better suited for high-dimensional problems, such as injury severity analysis with mixed logit models. This leads to increased accuracy in model estimation and improved convergence rates in the model fitting process. Moreover, the scrambling technique bolsters the robustness of Halton sequences, making them more resistant to artifacts or patterns potentially introduced by the original sequence's structure. This can be particularly beneficial in injury severity analysis, where small errors in the model's parameter estimation can have a significant impact on the accuracy and reliability of policy recommendations or safety interventions.

The random shift technique is another approach that can be employed to improve the performance of Halton sequences. While scrambling focuses on modifying the coefficients of the Halton sequence by permuting the digits in the base representation, random shift focuses on adding a random vector to each point in the Halton sequence to create a more diverse and uniform distribution. Random shift introduces an element of randomness into the deterministic Halton sequence, thereby enhancing its ability to explore the sample space more effectively. This is particularly beneficial in high-dimensional problems, where the uniformity of coverage can degrade rapidly due to the curse of dimensionality. By combining the low-discrepancy nature of the Halton sequence with the added randomness from the random shift, the resulting sequence can better handle higher-dimensional problems. The procedure for applying a random shift to the Halton sequence is relatively simple: (1) Generate a Halton sequence with the desired dimensionality and number of points. (2) Create a random shift vector with the same dimensionality, whose components are uniformly distributed in the interval [0, 1). (3) Add the random shift vector to each point in the original Halton sequence and take the modulus 1 of the resulting coordinates to obtain the shifted Halton sequence.

One advantage of using random shifts is that they are easy to implement and computationally efficient. Furthermore, since the random shift vector is chosen independently for each dimension, it allows for more flexible control over the distribution of points within the sample space. This is particularly important in applications where certain dimensions may require more exploration than others. When combining random shifts with scrambling techniques, the overall performance of the Halton sequence can be greatly improved (Owen, 2017). The combined effect of both methods results in a more uniformly distributed, low-discrepancy sequence that better suits higher-dimensional applications, such as numerical integration or Monte Carlo simulations. Moreover, the added randomness helps to facilitate error estimation through variance analysis. It is important to note that the choice of the random shift vector can have a significant impact on the resulting sequence. Therefore, it is crucial to use an appropriate random number generator that provides uniformly distributed random numbers. In some cases, it may be necessary to experiment with different random

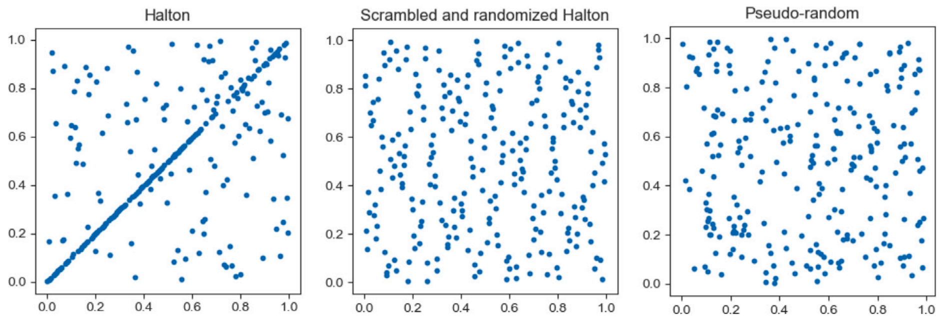

Fig. 2. First 100 points of the dimensional projection over the 49th and 50th coordinates of different sequences. (standard Halton sequence (left), scrambled and randomized Halton sequence (middle), and pseudo-random points (right)).

shift vectors to achieve the desired level of uniformity and correlation reduction.

In conclusion, the use of random shifts, either alone or in combination with scrambling techniques, can significantly improve the performance of Halton sequences in various applications. As shown in Fig. 2, by adding an element of randomness to the deterministic Halton sequence, the resulting sequence is more capable of handling high-dimensional problems, while maintaining the beneficial low-discrepancy properties of the original sequence. The resulting hybrid sequence incorporates the favorable coverage properties of randomized quasi-Monte Carlo (RQMC) sequences and the convenience of estimating simulation errors with conventional Monte Carlo methods. This advanced generation sampling method can be readily implemented with the R package `ivmte` (Shea and Torgovitsky, 2021).

4.3. Enhancing computational efficiency in estimation

Several recently proposed techniques have been proven to have beneficial impacts on the computational efficiency of the estimation (Helveston, 2022). Inspired by the R package `logitr` (Helveston, 2022), a straightforward but effective trick is reformulating the injury severity probabilities $P_n(i|\boldsymbol{v}_{in})$ to diminish the computations required to compute the log-likelihood function. Instead of computing the probability of each possible outcome, the probability of the driver n having $i-$th level injury can be calculated as

$$P'_n(i|\boldsymbol{v}_{in}) = \frac{1}{1+\sum_{j\neq i}^{I}\exp(\boldsymbol{\beta}_{jn}\boldsymbol{X}_{jn} - \boldsymbol{\beta}_{in}\boldsymbol{X}_{in})} \tag{11}$$

Correspondingly, the unconditional probability and log-likelihood can be simplified, and Eq. (5) can be written as

$$\mathcal{L} = \Sigma_{n=1}^{N}\Sigma_{i=1}^{I}\ln P'_n(i) \tag{12}$$

Thus, this process makes the calculation of the log-likelihood more stable and efficient. Furthermore, consider that the data used to compute the log-likelihood function and its gradient do not change, except for the parameters. Then by precomputing several intermediate calculations that remain constant throughout the estimation process, such as $\boldsymbol{X}_{jn} - \boldsymbol{X}_{in}$, $\exp(\boldsymbol{X}_{jn} - \boldsymbol{X}_{in})$, and $\exp(\boldsymbol{\beta}_{jn}\boldsymbol{X}_{jn} - \boldsymbol{\beta}_{in}\boldsymbol{X}_{in})$, a large amount of memory can be reduced. Consequently, the analytic gradient can be quickly obtained with only a few additional calculations in each interaction.

In addition to the aforementioned technique, there are several other approaches that can be employed to enhance the optimization process when minimizing the negative log-likelihood function in mixed logit models. Utilizing the `nloptr` package in R, an interface to the `NLopt` library, enables researchers to leverage a wide array of optimization algorithms tailored to handle nonlinear optimization problems. The versatility of `NLopt` is particularly advantageous, as it can concurrently optimize the log-likelihood function and its gradient, substantially improving the efficiency of the optimization process (Ypma, 2014).

The `NLopt` library offers flexibility in terms of optimization strategies, accommodating both local and global optimization techniques. This broad scope allows researchers to select the most suitable method for their specific problem, thereby potentially achieving faster convergence and more accurate results. Moreover, `NLopt` supports both derivative-free optimization and optimization with gradient information. Consequently, users can employ gradient-based algorithms, which capitalize on the gradient (or the first derivative) of the log-likelihood function, to streamline the optimization process further. This can be particularly beneficial in statistical models, as the gradient provides valuable insights that can guide the optimization algorithm toward the optimal solution more efficiently. Furthermore, the `NLopt` library facilitates the integration of the objective function and its gradient within the same function, which can expedite the optimization process. Consequently, researchers can more effectively optimize the log-likelihood function in mixed logit models, enabling them to obtain more accurate estimates and explore complex preference patterns and heterogeneity in the decision-making process.

Moreover, the estimation of mixed logit models can be further improved by leveraging parallel computing capabilities available on multi-core machines. By utilizing the `parallelly` and `snow` (Simple Network of Workstations) packages in R, researchers can estimate multiple models with different starting points in a larger solution space simultaneously, without substantially increasing the overall estimation time (Bengtsson, 2022; Tierney et al., 2008). The `parallelly` package offers a comprehensive suite of tools for parallel and distributed computing in R, including functions for parallel evaluation, data manipulation, and communication between parallel processes (Bengtsson, 2022). This package is particularly useful for harnessing the full potential of multi-core processors, enabling researchers to execute multiple tasks concurrently and reduce the time required to explore a vast solution space. The `parallelly` package seamlessly integrates with other parallel computing frameworks in R, such as the `snow` package, facilitating the development of efficient and scalable parallel algorithms for mixed logit model estimation. On the other hand, the `snow` package is a widely-used parallel computing library in R that provides high-level functions for creating and managing clusters of R processes running on multiple cores or processors (Tierney et al., 2008). By distributing the estimation tasks across multiple cores or processors, the `snow` package allows researchers to expedite the estimation of mixed logit models with different starting points, significantly enhancing the exploration of the solution space and increasing the likelihood of identifying the global optimum.

To effectively employ the `parallelly` and `snow` packages for mixed logit model estimation, several factors can be taken into consideration, such as allocating optimal resources by determining the appropriate number of cores or processors for parallel estimation tasks, and selecting diverse starting points in the solution space to explore a wide range of potential solutions. Additionally, it is essential to monitor parallel estimation tasks to identify and address potential issues or bottlenecks in the estimation process and to establish a systematic approach for aggregating and analyzing the results. By focusing on these aspects, researchers can harness the capabilities of the packages to efficiently estimate mixed logit models, ultimately providing a deeper understanding of complex models and heterogeneity.

4.4. Marginal effects

Marginal effects measure the effect of a one-unit change in a variable on the outcome, with all other variables held constant. Since all variables in the dataset have been converted to the form of indicators, marginal effects can be directly calculated by comparing the probabilities when the variable is changed from 0 to 1. Therefore, the marginal effects of independent variable X_{ink} can be readily given as

$$\mathcal{M}_{X_{ink}}^{P_n(i)} = P_n(i)|_{X_{ink}=1} - P_n(i)|_{X_{ink}=0} \tag{13}$$

where X_{ink} is the kth binary indicator in each severity level i for individual driver n. The marginal effect of each parameter is calculated by averaging the marginal effects over all observations. It is worth noting that when calculating marginal effects for a random parameter, the model should be re-sampled with the estimation method in Section 4, rather than using their estimated means directly. This is because the random parameters vary across observations and values of explanatory variables. A detailed discussion can be found in the article by Hou et al. (2022).

5. Experimental results and discussions

5.1. Estimation details

The estimation results for the mixed logit model with heterogeneity in means and variances and its three model counterparts, namely the fixed parameters multinomial logit, mixed logit, and mixed logit with heterogeneity in means, have been carefully analyzed and are presented in Table 2. The sensitivity analysis was a cornerstone of our research methodology. We employed several highly regarded metrics in the field of statistical modeling, to evaluate the trade-offs between model complexity and accuracy. This analysis was particularly insightful for understanding the nuanced impacts of different variables. For instance, in the context of the "week" variable, our analysis revealed that days like "Friday", "Saturday", and "Sunday" possessed significant coefficients that meaningfully altered the AIC scores, thereby validating their inclusion. In contrast, days such as "Monday", "Tuesday", and "Thursday", despite their initial consideration, were shown to have negligible influence on the model's predictive accuracy and overall fit, as evidenced by their non-significant coefficients and minimal impact on model fit metrics. This selective inclusion and exclusion of variables, informed by the sensitivity analysis, enhanced the model's precision without sacrificing its comprehensiveness. In deciding which variables to exclude from the model, we took a deliberative approach, understanding that every exclusion has implications for the model's interpretability and overall findings. The exclusion of non-significant variables was not an arbitrary choice but a carefully considered decision grounded in statistical evidence and practical modeling considerations. By incorporating these categories into their respective base categories, we created a more streamlined model that remains robust and interpretable. This method aligns with current best practices in statistical modeling, as documented in contemporary literature, which emphasizes the importance of avoiding overfitting and maintaining model parsimony, especially in complex datasets. The integration of non-significant categories through reference categories allows us to maintain a comprehensive perspective, ensuring that the relative impacts of all variables are considered, albeit in a more consolidated format.

Finally, as shown in Table 2, the independent variables included in the final models demonstrate statistical significance at a confidence level of 95% or higher. To capture the heterogeneity in the population, the estimation used a distributional form of random parameters. In this regard, several probability distributions such as normal, uniform, triangular, and lognormal were tested to determine the best statistical fit. After rigorous testing and analysis, the normal distribution was found to be the most suitable for capturing the distributional form of the random parameters, given its ability to fit the observed data well. The estimation results, based on the normal distributional form of random parameters, reveal that the mixed logit model with heterogeneity in means and variances significantly outperforms its three model counterparts, in terms of model fit and predictive accuracy.

Specifically, the model generated several normally distributed random parameters such as "Friday", "Slip Road", "Darkness with Light", "Snow or rain", "Truck", "Turning Left", and "Female". This suggests that there exists substantial heterogeneity in the population, which must be accounted for in order to obtain reliable estimates of the model parameters. We conducted a chi-square (χ^2) test to evaluate the necessity of adopting a more complex model (Wang et al., 2022). The analysis reveals that the mixed logit model with heterogeneity in means and variances significantly outperforms the other three models. This is evidenced by a higher ρ^2 statistic and a confidence level exceeding 99%, strongly supporting the rejection of the null hypothesis. The null hypothesis posits no significant difference between the mixed logit model with heterogeneity in means and variances and the other three models. Given these findings, our subsequent discussions and analyses will primarily focus on the insights derived from the mixed logit model with heterogeneity in means and variances.

For the estimation process, we utilized 1000 scrambled and randomized Halton draws. This implementation was carried out using a specially tailored version of the `logitr`, `parallel`, and `snow` packages in R. To provide a consistent and reliable computational environment, all our simulations and tests were run on a dedicated system. This system was equipped with an Intel Core i7 processor and 32 GB of RAM, and it operated on the Ubuntu 20.04 LTS platform, known for its stability and performance.

Our performance evaluation involved a multi-faceted approach, the results of which are summarized in Table 3. In this table, we compared the computational efficiency and model fit of commonly used estimation techniques: pseudo-random, Halton, and scrambled and randomized Halton. Among these, our customized method demonstrated a clear edge. It showcased the highest value for ρ^2, while also proving to be the least resource-intensive. This was evidenced by its minimal AIC score, the fastest computational speed, and the smallest memory footprint. To further articulate the efficiency of our approach, Fig. 3 provides an illustrative comparison. This graphical representation plots the computational times required for different estimation methods across varying numbers of Halton draws, ranging from as few as 50 to as many as 1000. What becomes immediately evident is that our method consistently outpaces its competitors in terms of computational speed. This advantage becomes even more pronounced as the scale of the problem—the number of draws—increases. Remarkably, this speed advantage is maintained even when our method is compared with multi-core implementations of other algorithms, underscoring its robustness and suitability for large-scale, computational tasks that demand both speed and accuracy.

In conclusion, our method excels in two pivotal aspects: it provides a superior statistical fit to the observed data and stands out for its computational efficiency. These strengths make it an exceptionally robust and reliable choice for implementing mixed logit models with heterogeneity in both means and variances using scrambled and randomized Halton estimation.

5.2. Parameter estimates

5.2.1. Weekdays

The marginal effects detailed in Table 4 provide a comprehensive analysis of how different days of the week influence crash severity, with

Table 2
Model estimates results.

Variables	Fixed parameters multinomial logit		Mixed logit		Mixed logit with heterogeneity in means		Mixed logit with heterogeneity in means and variances	
	Serious Est. (t-stat)	Fatal Est. (t-stat)	Serious Est. (t-stat)	Fatal Est. (t-stat)	Serious Est. (t-stat)	Fatal Est. (t-stat)	Serious Est. (t-stat)	Fatal Est. (t-stat)
Intercept	−2.583 (−12.45)	−3.214 (−17.06)	−2.845 (−11.38)	−3.402 (−16.92)	−2.478 (−13.67)	−3.189 (−14.27)	−2.492 (−14.03)	−3.225 (−15.61)
Friday	0.945 (4.87)	1.006 (5.34)	0.912 (4.73)	1.021 (5.24)	0.936 (4.98)	1.010 (5.15)	0.920 (4.81)	0.999 (5.22)
Std. dev of Friday	–	–	1.127 (2.08)	1.206 (2.24)	1.098 (2.05)	1.186 (2.17)	1.101 (2.02)	1.180 (2.20)
Saturday	2.128 (10.64)	2.254 (11.27)	2.110 (10.53)	2.242 (11.21)	2.121 (10.69)	2.240 (11.15)	2.098 (10.49)	2.231 (11.08)
Sunday	0.864 (4.32)	0.951 (4.75)	0.879 (4.38)	0.966 (4.83)	0.857 (4.28)	0.942 (4.71)	0.871 (4.34)	0.958 (4.79)
Dual Carriageway	−0.517 (−2.58)	−0.628 (−3.14)	−0.503 (−2.51)	−0.612 (−3.06)	−0.529 (−2.64)	−0.642 (−3.21)	−0.516 (−2.57)	−0.629 (−3.13)
One Way Street	−1.483 (−7.40)	−1.601 (−8.01)	−1.507 (−7.53)	−1.623 (−8.14)	−1.471 (−7.31)	−1.586 (−7.92)	−1.495 (−7.44)	−1.612 (−8.05)
Roundabout	0.302 (1.51)	0.356 (1.78)	0.319 (1.59)	0.376 (1.88)	0.292 (1.46)	0.345 (1.72)	0.308 (1.54)	0.362 (1.81)
Slip Road	0.377 (1.88)	0.402 (2.01)	0.385 (1.92)	0.410 (2.05)	0.370 (1.85)	0.395 (1.98)	0.378 (1.89)	0.403 (2.02)
Std. dev of Slip Road	–	–	1.772 (1.39)	1.963 (1.82)	1.803 (1.41)	2.004 (1.97)	1.812 (1.55)	2.010 (2.01)
40–50 mph	1.970 (9.85)	2.085 (10.42)	1.990 (9.95)	2.110 (10.55)	1.958 (9.79)	2.070 (10.35)	1.977 (9.88)	2.092 (10.46)
≥60 mph	1.157 (5.79)	1.222 (6.11)	1.172 (5.84)	1.236 (6.18)	1.147 (5.74)	1.210 (6.05)	1.162 (5.81)	1.228 (6.14)
Darkness with Light	0.759 (3.80)	0.820 (4.10)	0.772 (3.86)	0.831 (4.15)	0.749 (3.74)	0.809 (4.04)	0.762 (3.81)	0.822 (4.11)
Std. dev of Darkness with Light	–	–	1.453 (2.76)	1.601 (3.02)	1.502 (2.82)	1.614 (3.11)	1.506 (2.84)	1.619 (3.15)
Darkness without Light	2.420 (12.10)	2.556 (12.78)	2.439 (12.20)	2.574 (12.87)	2.403 (12.01)	2.540 (12.70)	2.423 (12.11)	2.558 (12.79)
Fine with High Winds	1.348 (6.74)	1.435 (7.18)	1.362 (6.81)	1.449 (7.25)	1.335 (6.67)	1.420 (7.10)	1.351 (6.75)	1.438 (7.19)
Snow or Rain	0.836 (4.18)	0.871 (4.36)	0.848 (4.24)	0.883 (4.42)	0.827 (4.14)	0.863 (4.32)	0.839 (4.20)	0.875 (4.38)
Std. dev of Snow or Rain	–	–	1.124 (2.07)	1.124 (2.07)	1.412 (3.14)	1.412 (3.14)	1.409 (3.11)	1.409 (3.11)
Ice or Snow	1.207 (6.03)	1.335 (6.68)	1.223 (6.12)	1.350 (6.75)	1.196 (5.98)	1.324 (6.62)	1.211 (6.05)	1.339 (6.69)
Urban	−0.641 (−3.20)	−0.586 (−2.93)	−0.655 (−3.27)	−0.600 (−2.99)	−0.636 (−3.18)	−0.581 (−2.90)	−0.627 (−3.29)	−0.574 (−3.02)
Bus	1.699 (8.50)	1.781 (8.90)	1.711 (8.56)	1.793 (8.96)	1.688 (8.44)	1.770 (8.84)	1.701 (8.51)	1.783 (8.92)
Truck	2.111 (10.56)	2.235 (11.18)	2.125 (10.62)	2.249 (11.24)	2.098 (10.49)	2.222 (11.10)	2.114 (10.57)	2.238 (11.19)
Std. dev of Truck	–	–	1.588 (3.04)	1.588 (3.04)	1.452 (2.96)	1.452 (2.96)	1.449 (2.94)	1.449 (2.94)
Motorcycle	0.868 (4.34)	0.936 (4.68)	0.882 (4.41)	0.950 (4.75)	0.859 (4.29)	0.927 (4.63)	0.875 (4.37)	0.943 (4.72)
Lane Changing	0.315 (1.57)	0.382 (1.91)	0.327 (1.63)	0.394 (1.97)	0.304 (1.52)	0.372 (1.86)	0.318 (1.58)	0.385 (1.93)
Overtaking	0.720 (3.60)	0.756 (3.78)	0.733 (3.67)	0.768 (3.84)	0.709 (3.54)	0.746 (3.73)	0.722 (3.61)	0.759 (3.79)
Turning Left	1.958 (9.78)	1.990 (9.95)	1.970 (9.85)	2.002 (10.01)	1.949 (9.74)	1.981 (9.91)	1.961 (9.81)	1.993 (9.98)
Std. dev of Turning Left	–	–	3.856 (1.93)	3.856 (1.93)	2.742 (1.46)	2.742 (1.46)	2.542 (1.53)	2.542 (1.53)
Turning Right	1.321 (6.60)	1.362 (6.81)	1.335 (6.67)	1.376 (6.88)	1.309 (6.54)	1.350 (6.75)	1.324 (6.62)	1.365 (6.83)
U Turn	2.024 (10.12)	2.094 (10.47)	2.038 (10.19)	2.108 (10.54)	2.012 (10.06)	2.082 (10.41)	2.027 (10.14)	2.097 (10.49)
Female	−0.735 (−3.68)	−0.673 (−3.37)	−0.751 (−3.76)	−0.688 (−3.44)	−0.721 (−3.61)	−0.661 (−3.30)	−0.738 (−3.69)	−0.676 (−3.39)
Std. dev of Female	–	–	1.337 (3.25)	1.460 (3.39)	1.352 (3.19)	1.432 (3.21)	1.357 (3.22)	1.434 (3.26)
≤25	1.701 (8.50)	1.728 (8.64)	1.713 (8.56)	1.740 (8.70)	1.690 (8.45)	1.717 (8.59)	1.703 (8.51)	1.730 (8.65)
>65	0.574 (2.87)	0.608 (3.04)	0.586 (2.93)	0.620 (3.10)	0.565 (2.82)	0.599 (2.99)	0.577 (2.89)	0.611 (3.06)
Heterogeneity in means of random parameters								
Friday: Motorcycle					0.986 (2.24)	0.986 (2.24)	0.992 (2.28)	0.992 (2.28)
Slip road: Lane Changing					0.589 (3.81)	0.588 (3.81)	0.594 (3.87)	0.594 (3.87)
Darkness with light: Urban					0.337 (2.02)	0.337 (2.02)	0.341 (2.12)	0.341 (2.12)
Truck: Turning Left					0.548 (3.65)	0.548 (3.65)	0.576 (3.69)	0.576 (3.69)
Female: Slowing					−0.708 (−7.15)	−0.708 (−7.15)	−0.696 (−7.22)	−0.696 (−7.22)
Heterogeneity in variances of random parameters								
Friday: Roundabout							0.255 (2.92)	0.255 (2.92)
Darkness with Light: Bus							−0.276 (1.97)	−0.276 (1.97)
Snow or rain: >65							0.724 (2.27)	0.724 (2.27)
Model statistics								
Number of observations	39 788							
Number of parameters	27		34		39		42	
Log-likelihood at zero, LL (0)	−8325.072		−9253.064		−9253.064		−9253.064	
Log-likelihood at convergence, LL(β)	−7012.221		−6853.106		−6822.027		−6810.141	
ρ^2	0.158		0.259		0.263		0.264	
Akaike Information Criterion (AIC)	14 078.442		13 774.212		13 722.054		13 704.282	
χ^2 test (VS mixed logit with heterogeneity in means and variances)	404.160 (15) [>99.99%]		85.930 (8) [>99.99%]		23.772 (3) [>99.99%]		–	

Table 3
Performance comparison across different estimation approaches with 1000 Draws (LL: Log-likelihood, Para.: Parameter, χ^2: χ^2 test vs. Scrambled and Randomized Halton Estimation).

Methods	LL	Para.	χ^2	ρ^2	AIC	Run time (s)	Memory (MB)
Pseudo-random	−6889.7	38	159.2 (4)	0.252	13 825.4	47.5	220.7
Halton	−6840.2	41	60.2 (1)	0.258	13 726.4	42.3	190.2
Scrambled and Randomized Halton	−6810.1	42	–	0.264	13 704.3	39.7	175.4

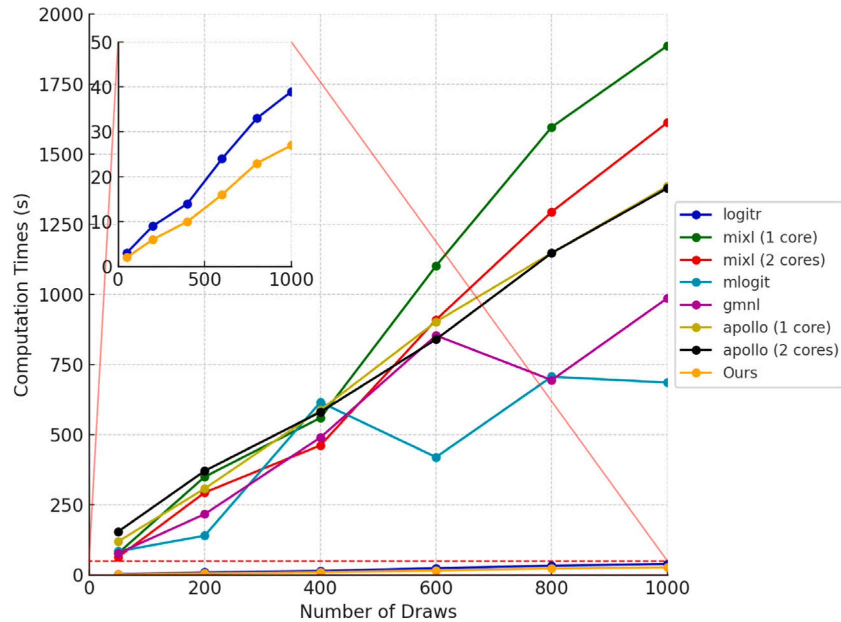

Fig. 3. Illustrative analysis of computational efficacy. This representation provides a comparative analysis of computation times across diverse estimation approaches in different R packages, emphasizing the enhanced adeptness and unparalleled computational resilience and agility of our proposed method in the realm of mixed logit models with heterogeneity in means and variances.

Table 4
Marginal effects of the significant variables in the model.

Variable	Slight	Serious	Fatal
Friday	0.0007	0.0007	−0.0014
Saturday	−0.0015	−0.0004	0.0019
Sunday	0.0001	−0.0012	0.0011
Dual Carriageway	0.0021	−0.0010	−0.0011
One Way Street	0.0001	0.0002	−0.0003
Roundabout	−0.0008	0.0007	0.0001
Slip road	−0.0019	0.0020	−0.0001
40–50 mph	0.0032	0.0008	−0.0040
≥60 mph	−0.0003	0.0002	0.0001
Darkness with light	0.0019	0.0020	−0.0039
Darkness without light	−0.0000	−0.0005	0.0005
Fine with high winds	0.0015	0.0005	−0.0020
Snow or rain	−0.0008	0.0004	0.0004
Ice or snow	0.0015	−0.0005	−0.0010
Urban	0.0005	−0.0015	0.0010
Bus	0.0017	0.0014	−0.0031
Truck	−0.0004	−0.0018	0.0023
Motorcycle	0.0010	−0.0035	0.0025
Lane Changing	−0.0013	0.0021	−0.0008
Overtaking	0.0031	−0.0014	−0.0017
Turning Left	0.0011	−0.0007	−0.0004
Turning Right	−0.0006	0.0017	−0.0011
U Turn	−0.0011	0.0004	0.0007
Female	0.0002	0.0015	−0.0018
≤25	0.0002	−0.0005	0.0003
≥65	0.0013	−0.0023	0.0010

a specific focus on “Friday”, “Saturday”, and “Sunday”. On Fridays, there is a noticeable trend of increased occurrences of slight and serious injuries, coupled with a reduced likelihood of fatal outcomes. This pattern may suggest that while crashes are more frequent on Fridays, they are less likely to be fatal. This could be attributed to factors such as increased commuting traffic, a rush to start weekend activities, or potentially different driving patterns at the end of the workweek, leading to a higher incidence of crashes that are severe but not fatal.

Saturdays present a distinct shift in crash severity dynamics. The data indicates a significant reduction in slight injuries but an increase in fatal outcomes. This shift could be influenced by various factors characteristic of weekend activities, such as increased recreational travel, possibly higher alcohol consumption, or variations in traffic patterns. These factors might contribute to fewer overall crashes but a greater severity in those that do occur.

Sundays show a nuanced trend with a slight increase in slight injuries, a decrease in serious injuries, and an elevation in fatal outcomes. This pattern might be the result of a complex interplay of factors, including reduced traffic congestion and residual effects of Saturday night activities. The lower traffic volumes could lead to higher speeds and potentially more severe crashes, while the aftermath of weekend social activities might influence driver behavior and alertness.

These observed patterns, particularly from Friday to Sunday, highlight the complexity of factors influencing crash severity on different days. They underscore the importance of considering day-specific variables such as traffic volume, driver behavior, and recreational activities in understanding and mitigating traffic-related risks. The data strongly supports the need for targeted traffic safety initiatives and interventions, tailored to the unique characteristics and risks associated with specific days of the week (Se et al., 2022; Li et al., 2019b).

5.2.2. Road type

Dual carriageways, distinguished by their higher speeds and divided lanes, present a complex pattern in crash outcomes. Notably, there is a marginal increase in the occurrence of slight injuries, but a noticeable decrease in serious and fatal incidents. This pattern could be linked to the characteristic high speeds of these roads. While these speeds might contribute to a lower overall frequency of crashes, they also have the potential to intensify the severity of the crashes that do occur. This observation underscores the intricate relationship between road design features and the resulting impact on crash severity (Wang et al., 2023b,a).

One-way streets, typically found in urban settings, demonstrate a slight increase in both slight and serious injuries, but a decrease in fatal injuries. This pattern suggests that the controlled environment of one-way streets might lower the risk of fatal crashes. However, the close proximity to urban obstacles could result in a higher number of slight and serious injuries, reflecting the dual aspects of urban road design in influencing crash outcomes (Zamani et al., 2021).

Regarding roundabouts, the data shows a decrease in slight injuries, an increase in serious injuries, and a small rise in fatal outcomes. The design of roundabouts, which generally reduces traffic speed, might contribute to the lower frequency of slight injuries. However, the required complex driving maneuvers might increase the risk of serious collisions. The slight uptick in fatal outcomes might be linked to high-risk behaviors like high-speed entries or non-compliance with traffic rules, highlighting the critical role of driver behavior in roundabout safety (Poudel and Singleton, 2021).

5.2.3. Speed limit

The analysis of speed zones, as reflected in Table 4, provides significant insights into their impact on crash severity, particularly in zones with speed limits of “40-50 mph” and “≥ 60 mph”. In 40–50 mph zones, there is a notable impact on the frequency and severity of crashes. The data suggests a considerable increase in the likelihood of slight and serious injuries. However, there is a marked decrease in fatal outcomes. This pattern could be attributed to the nature of these zones, which often serve as transitional areas bridging urban and rural settings. In such zones, drivers might retain a mindset of cautious driving typical of urban areas. While the speeds are sufficiently high to increase the incidence of slight and serious injuries, they might not be extreme enough to lead to a significant number of fatal crashes, especially if urban driving caution is maintained. This observation underscores the complex relationship between speed, driver behavior, and crash severity in such transitional zones (Alnawmasi and Mannering, 2022).

Conversely, in zones with speed limits of 60 mph or higher, a different trend emerges. There is a decrease in slight injuries, but both serious and fatal outcomes show an increase. These areas are often highways or open roads where vehicles operate at higher speeds. At such velocities, the energy involved in a collision escalates dramatically, increasing the likelihood of serious or fatal injuries. Furthermore, higher speeds reduce the reaction time available to drivers, amplifying the risk and severity of collisions (Li et al., 2019a).

These findings reiterate the established correlation between speed and crash severity. High-speed zones, while facilitating efficient travel, also introduce heightened risks. The data suggests a potential need to reassess speed limits in areas where the severity of crashes is disproportionately high. Implementing traffic calming measures, such as speed bumps, or enhancing law enforcement presence, could be effective strategies to mitigate these risks and enhance road safety (Samerei et al., 2021).

5.2.4. Lighting condition

The influence of lighting conditions on road safety is clearly highlighted in the results presented in Table 4. In scenarios described as “Darkness with light”, typically indicative of well-lit environments, there is a significant increase in the frequency of slight and serious injuries. Interestingly, these same conditions are associated with a notable decrease in fatal outcomes. This phenomenon can be attributed to the heightened sense of security that well-lit conditions provide to drivers. In such environments, drivers may exhibit a degree of overconfidence or minor lapses in judgment, leading to an increased occurrence of less severe crashes. However, the improved visibility in these conditions also plays a crucial role in reducing the likelihood of fatal crashes, as drivers are more capable of perceiving potential hazards, including pedestrians, in time to avoid catastrophic outcomes (Li et al., 2023).

Conversely, “Darkness without light” conditions, characterized by the absence of adequate lighting, show a different impact. There is a slight decrease in both slight and serious injuries, but an increase in fatal outcomes. The lack of proper lighting creates a challenging environment where the visibility of road features, other vehicles, and pedestrians is significantly compromised. This limitation might lead to heightened caution among drivers and pedestrians, resulting in fewer non-fatal incidents. However, the same factor drastically increases the risk of fatal crashes, as the inability to clearly perceive and react to road hazards in time can lead to more severe collisions (Azimi et al., 2020).

These findings underline the imperative need for proper street lighting as a public safety measure. Proper lighting not only illuminates obstacles but also significantly impacts driver and pedestrian behavior (Hu et al., 2020). It may instill a sense of security, but this should not translate into complacency. Municipalities and city planners should consider these findings when designing roadways, particularly in areas where the rate of severe or fatal crashes is high.

5.2.5. Weather condition

In conditions classified as “Fine with high winds”, there is an observed increase in the probability of slight and serious injuries, while fatal outcomes show a considerable decrease. High winds, although not impeding visibility, can substantially impact vehicle control. Drivers may struggle to maintain direction or be unexpectedly affected by sudden gusts, resulting in more frequent non-fatal, yet serious crashes. The lower incidence of fatal crashes in these conditions could be attributed to increased driver caution due to the winds, or perhaps an effect where the winds ‘disperse’ potentially fatal crashes into less severe categories (Alrejjal et al., 2021).

Conversely, in “Snow or rain” conditions, there is a noted decrease in slight injuries but increases in serious and fatal injuries. Adverse weather conditions like snow and rain significantly affect road conditions by reducing traction and visibility. These factors contribute to a higher risk of severe crashes. While drivers may reduce their speed in response to these conditions, thereby lowering the occurrence of slight injuries, the compromised road conditions can escalate the severity of collisions, transforming minor incidents into serious or fatal crashes (Yu et al., 2020).

The data underscores the critical need for adaptive driving behavior in response to varying weather conditions. Additionally, it highlights the importance of road design that considers local weather patterns. Implementing educational initiatives to inform the public about the risks associated with different weather conditions, coupled with the strategic placement of road signs to warn drivers of these risks, could significantly reduce the severity of crashes under various weather scenarios (Yu et al., 2019).

5.2.6. Area type

The type of area in which a crash occurs significantly influences its severity, as indicated by the data in Table 4. In urban settings, there is an observed increase in the likelihood of slight injuries, but a notable decrease in serious injuries and an alarming increase in fatal outcomes. This trend suggests that urban areas, while potentially safer for non-fatal crashes, harbor specific risks that elevate the probability of fatal incidents. Contributing factors could include pedestrian density, frequent stop-and-go traffic, and the prevalence of distractions for drivers, such as digital billboards and dense signage. Moreover, urban settings might have a higher occurrence of certain types of fatal crashes, such as those involving multiple vehicles or pedestrians, due to their densely populated nature (Yu et al., 2021a; Zou et al., 2023).

These findings point to the necessity for targeted safety measures in urban areas. The development of public awareness campaigns, stricter enforcement of traffic laws, and urban planning initiatives designed to identify and mitigate high-risk areas can significantly enhance safety in urban driving environments (Se et al., 2021a).

5.2.7. Vehicle type

The impact of vehicle type on crash severity is also evident from Table 4. Buses show an increase in slight and serious injuries but a substantial decrease in fatal outcomes. The larger size and lesser maneuverability of buses might result in more frequent non-fatal collisions. However, the lower rate of fatal outcomes could be attributed to the robust safety features of buses and the typically more cautious driving style of bus operators (Feng et al., 2016).

Trucks, in contrast, exhibit a decrease in slight and serious injuries but a significant increase in fatal outcomes. Advanced safety technologies in modern trucks might contribute to the reduction in non-fatal injuries. However, the high severity of truck-related crashes leading to fatalities could be due to factors like the vehicle's large mass and higher traveling speeds, which, when involved in crashes, tend to result in more severe outcomes (Behnood and Mannering, 2019).

These findings highlight the importance of vehicle-specific safety strategies. For buses, enhancing safety features and driver training could further reduce the rate of injuries. In the case of trucks, while safety technology has helped reduce less severe injuries, a deeper understanding and mitigation of factors leading to fatalities are imperative (Yang et al., 2021).

5.2.8. Maneuver

The severity of injuries resulting from crashes is significantly influenced by the vehicle maneuvers involved, as shown in Table 4. Lane changing tends to reduce the likelihood of slight injuries but increases the probability of serious injuries and slightly reduces fatal outcomes. The sudden nature of lane changes often leads to collisions that, while severe, are less likely to be fatal due to the typically lower speeds and shorter reaction times involved (Ali et al., 2022).

Overtaking maneuvers show an increased likelihood of slight injuries but a reduction in serious and fatal injuries. This trend might be due to the heightened alertness and faster reaction times typically associated with overtaking, allowing drivers to take evasive actions to prevent more severe collisions (Kovaceva et al., 2022).

Turning maneuvers also exhibit varying patterns. Left turns increase the likelihood of slight injuries but decrease serious and fatal injuries, while right turns show a reduction in slight injuries but an increase in serious injuries and a decrease in fatal outcomes. These variations could be attributed to differences in visibility, traffic flow, and compliance with right-of-way rules during these maneuvers, influencing the severity of crashes (Ali et al., 2022).

U-Turns present an increased risk for serious and fatal injuries while reducing the likelihood of slight injuries. The complexity of U-Turns, often requiring crossing multiple lanes of traffic, contributes to their higher risk, especially when executed without sufficient visibility or in inappropriate locations (Kovaceva et al., 2022).

The data underscores the importance of specialized road safety measures and educational programs focusing on safe practices for different types of maneuvers. Additionally, road designs could be optimized to minimize the risks associated with specific actions like turning or overtaking, enhancing overall road safety (Yang et al., 2021).

5.2.9. Gender and age

The "Female" gender category shows an increase in the likelihood of slight and serious injuries, but a decrease in fatal outcomes. This trend might be reflective of distinct driving behaviors or physical vulnerabilities between genders. It is important to note, however, that further research is needed to conclusively isolate and understand these effects (Yan et al., 2021).

In the age category of "≤ 25", there is a slight increase in slight injuries, a decrease in serious injuries, and an increase in fatal outcomes. Younger drivers, often associated with riskier driving behaviors, may also possess quicker reflexes, enabling them to perform evasive maneuvers. This could potentially reduce the occurrence of serious injuries but may lead to an increase in both slight and fatal crashes. The increase in fatal outcomes could be related to the high-risk behaviors commonly observed in this age group, such as speeding or distracted driving (Useche et al., 2019).

Conversely, the "> 65" age group exhibits an increase in slight injuries, a significant decrease in serious injuries, and an increase in fatal outcomes. This pattern suggests that older drivers might adopt more cautious driving habits, leading to fewer serious crashes. However, factors like slower reflexes and potential health complications might contribute to higher rates of slight and fatal injuries in this demographic. This emphasizes the complex interplay between age, physical abilities, and driving behaviors in determining crash severity (Amiri et al., 2020).

These findings highlight the importance of designing age- and gender-specific road safety strategies. Tailored interventions such as public awareness campaigns, driving education programs, and the implementation of driving aids can effectively address the unique risk profiles and safety needs of different demographic groups.

5.3. Heterogeneity in results

The analysis of crash severity dynamics, as per the coefficients and t-values presented in Table 2, reveals complex interactions and heterogeneities. The mixed logit models with heterogeneity in means and variances provide particularly insightful observations.

One key interaction is between "Friday" and "Motorcycle", with a coefficient of 0.992 and a notable t-value of 2.28. This interaction implies that the risk associated with motorcycle crashes varies significantly on Fridays, possibly due to the combination of weekend traffic patterns and the inherent vulnerabilities of motorcycles. This could reflect a scenario where motorcyclists' behavior or surrounding traffic conditions on Fridays differ from other weekdays, impacting crash severity.

The term "Slip road: Lane Changing" also demonstrates significant heterogeneity with a coefficient of 0.594 and a t-value of 3.87. This suggests that on slip roads, where speeds vary and drivers are frequently merging or exiting, the act of lane changing introduces substantial risk. The decision-making process in such scenarios, often under time pressure and requiring quick assessment of speeds and gaps, can lead to a higher risk of severe crashes.

Further, the interaction between "Darkness with Light" and "Urban" settings shows a coefficient of 0.341 and a t-value of 2.12. This indicates that while urban areas are typically well-lit, the effectiveness of lighting may vary, affecting crash severity. Factors like the positioning of lights, their intensity, or obstructions caused by urban infrastructure can lead to unpredictable and hazardous conditions, particularly in complex urban traffic scenarios.

In terms of heterogeneity in variances, the model reveals intriguing insights. The interaction of "Snow or Rain" with the age group "> 65" shows a coefficient of 0.724 with a t-value of 2.27. This implies that the variance in crash severity under snowy or rainy conditions increases significantly for drivers over 65. The reasons could include age-related factors such as reduced reaction times or visual acuity, which, when combined with challenging weather conditions, might amplify the risk of severe crashes.

These findings from the mixed logit models underscore the importance of considering a wide range of factors and their interactions in understanding crash severity. The model's ability to capture such detailed nuances highlights its efficacy in informing road safety strategies. It emphasizes the need for comprehensive approaches that account for the diverse and interconnected factors affecting road safety, enabling more targeted and effective interventions for reducing crash severity.

6. Conclusions

This study offers a comprehensive examination of the factors affecting the severity of road crashes, using a novel method that proves to be computationally more efficient than existing approaches. Our analysis uncovers distinct patterns in crash severity based on variables such as weekdays, road types, speed limits, lighting conditions, and weather. For example, we found that Fridays generally result in less severe crashes, whereas Saturdays are associated with higher severity. Dual carriageways, characterized by higher speed limits, were identified as a type of road where more severe crashes are likely to occur. Moreover, our findings substantiate the long-held belief that higher speed limits can exacerbate crash severity and underscore the importance of proper street lighting for public safety.

The model we employed in this study stands out for its multiple strengths, making it a robust tool for future road safety research. One of the most notable advantages is its computational efficiency, enabling quicker analysis with less demand on computational resources, thus paving the way for real-time applications. Alongside its efficiency, the model excels in predictive accuracy, reliably capturing underlying trends and patterns to inform better decision-making in road safety initiatives. It has the remarkable ability to identify complex interactions among variables, which provides a nuanced, holistic view of the factors influencing crash severity. This is particularly vital for creating comprehensive safety strategies. Adding to its adaptability, the model incorporates heterogeneity in both means and variances, allowing it to account for the diverse scenarios and complex nature of road crashes. This feature enhances its applicability and generalizability across different datasets and conditions. The model's scalability is another significant asset; it maintains consistent performance even when applied to larger and more diverse datasets, demonstrating its versatility. Lastly, its capability to identify significant random parameters offers deeper insights into the intricate dynamics of road safety that are often overlooked in traditional models. These collective strengths make our model a highly reliable and versatile choice for studies requiring both computational efficiency and in-depth analysis.

Looking ahead, future research could aim to refine the model by incorporating alternative optimization techniques, including a broader set of random parameters, and experimenting with advanced parallel and distributed computing platforms. Such innovative methods could further enrich our understanding of mixed logit models, paving the way for more robust and nuanced approaches to studying injury severity.

CRediT authorship contribution statement

Zhenning Li: Conceptualization, Methodology, Writing. **Chengyue Wang:** Experiment. **Haicheng Liao:** Experiment. **Guofa Li:** Methodology. **Chengzhong Xu:** Conceptualization, Funding, Review.

Declaration of competing interest

The authors declare that there are no conflicts of interest regarding the publication of this paper. No financial support or benefits from commercial sources was received to conduct this study.

Data availability

Data will be made available on request.

Acknowledgments

This research is supported by the Science and Technology Development Fund of Macau SAR (File no. 0021/2022/ITP, 0081/2022/A2, SKL-IoTSC(UM)-2021-2023/ORP/GA08/2022, SKL-IoTSC(UM)-2024 - 2026/ORP/GA06/2023), and University of Macau (SRG2023-00037-IOTSC). We extend our gratitude to the Institute for Advanced Studies in Humanities and Social Sciences (IAS) for their valuable assistance. For any correspondence regarding this paper, please contact Dr. Zhenning Li at zhenningli@um.edu.mo or Dr. Chengzhong Xu at czxu@um.edu.mo.

References

Ahmed, S.S., Corman, F., Anastasopoulos, P.C., 2023. Accounting for unobserved heterogeneity and spatial instability in the analysis of crash injury-severity at highway-rail grade crossings: A random parameters with heterogeneity in the means and variances approach. Anal. Methods Accid. Res. 37, 100250.

Al-Bdairi, N.S.S., Behnood, A., Hernandez, S., 2020. Temporal stability of driver injury severities in animal-vehicle collisions: A random parameters with heterogeneity in means (and variances) approach. Anal. Methods Accid. Res. 26, 100120.

Ali, Y., Haque, M.M., Zheng, Z., 2022. An extreme value theory approach to estimate crash risk during mandatory lane-changing in a connected environment. Anal. Methods Accid. Res. 33, 100193.

Alnawmasi, N., Mannering, F., 2022. The impact of higher speed limits on the frequency and severity of freeway crashes: Accounting for temporal shifts and unobserved heterogeneity. Anal. Methods Accid. Res. 34, 100205.

Alogaili, A., Mannering, F., 2020. Unobserved heterogeneity and the effects of driver nationality on crash injury severities in Saudi Arabia. Accid. Anal. Prev. 144, 105618.

Alogaili, A., Mannering, F., 2022. Differences between day and night pedestrian-injury severities: Accounting for temporal and unobserved effects in prediction. Anal. Methods Accid. Res. 33, 100201.

Alrejjal, A., Farid, A., Ksaibati, K., 2021. A correlated random parameters approach to investigate large truck rollover crashes on mountainous interstates. Accid. Anal. Prev. 159, 106233.

Amiri, A.M., Sadri, A., Nadimi, N., Shams, M., 2020. A comparison between artificial neural network and hybrid intelligent genetic algorithm in predicting the severity of fixed object crashes among elderly drivers. Accid. Anal. Prev. 138, 105468.

Azimi, G., Rahimi, A., Asgari, H., Jin, X., 2020. Severity analysis for large truck rollover crashes using a random parameter ordered logit model. Accid. Anal. Prev. 135, 105355.

Behnood, A., Mannering, F., 2017. Determinants of bicyclist injury severities in bicycle-vehicle crashes: A random parameters approach with heterogeneity in means and variances. Anal. Methods Accid. Res. 16, 35–47.

Behnood, A., Mannering, F., 2019. Time-of-day variations and temporal instability of factors affecting injury severities in large-truck crashes. Anal. Methods Accid. Res. 23, 100102.

Bengtsson, H., 2022. Enhancing the 'parallel' Package [R package parallelly version 1.32.1]. Compr. R Arch. Netw. (CRAN) URL: https://cran.r-project.org/web/packages/parallelly/index.html.

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

Bhat, C.R., 2003. Simulation estimation of mixed discrete choice models using randomized and scrambled halton sequences. Transp. Res. Part B: Methodol. 37 (9), 837–855.

Braaten, E., Weller, G., 1979. An improved low-discrepancy sequence for multidimensional quasi-Monte Carlo integration. J. Comput. Phys. 33 (2), 249–258. http://dx.doi.org/10.1016/0021-9991(79)90019-6, URL: https://www.sciencedirect.com/science/article/pii/0021999179900196.

Ding, H., Lu, Y., Sze, N., Antoniou, C., Guo, Y., 2023. A crash feature-based allocation method for boundary crash problem in spatial analysis of bicycle crashes. Anal. Methods Accid. Res. 37, 100251.

Feng, S., Li, Z., Ci, Y., Zhang, G., 2016. Risk factors affecting fatal bus accident severity: Their impact on different types of bus drivers. Accid. Anal. Prev. 86, 29–39.

Fisa, R., Musukuma, M., Sampa, M., Musonda, P., Young, T., 2022. Effects of interventions for preventing road traffic crashes: an overview of systematic reviews. BMC Public Health 22 (1), 513.

Fountas, G., Fonzone, A., Gharavi, N., Rye, T., 2020. The joint effect of weather and lighting conditions on injury severities of single-vehicle accidents. Anal. Methods Accid. Res. 27, 100124.

Hamed, M.M., AlShaer, A., 2023. Analysis of duration between crashes using a hazard-based duration approach with heterogeneity in means and variances: Some new evidence. Anal. Methods Accid. Res. 100283.

Helveston, J.P., 2022. Logitr: Fast estimation of multinomial and mixed logit models with preference space and willingness to pay space utility parameterizations. arXiv preprint arXiv:2210.10875.

Hou, Q., Huo, X., Leng, J., Mannering, F., 2022. A note on out-of-sample prediction, marginal effects computations, and temporal testing with random parameters crash-injury severity models. Anal. Methods Accid. Res. 33, 100191.
Hu, L., Wu, X., Huang, J., Peng, Y., Liu, W., 2020. Investigation of clusters and injuries in pedestrian crashes using GIS in Changsha, China. Saf. Sci. 127, 104710.
Huo, X., Leng, J., Hou, Q., Zheng, L., Zhao, L., 2020. Assessing the explanatory and predictive performance of a random parameters count model with heterogeneity in means and variances. Accid. Anal. Prev. 147, 105759.
Islam, M., Alnawmasi, N., Mannering, F., 2020. Unobserved heterogeneity and temporal instability in the analysis of work-zone crash-injury severities. Anal. Methods Accid. Res. 28, 100130.
Jain, D.C., Vilcassim, N.J., Chintagunta, P.K., 1994. A random-coefficients logit brand-choice model applied to panel data. J. Bus. Econom. Statist. 12 (3), 317–328.
Kocis, L., Whiten, W.J., 1997. Computational investigations of low-discrepancy sequences. ACM Trans. Math. Softw. 23 (2), 266–294.
Kovaceva, J., Bärgman, J., Dozza, M., 2022. On the importance of driver models for the development and assessment of active safety: a new collision warning system to make overtaking cyclists safer. Accid. Anal. Prev. 165, 106513.
Li, Z., Chen, C., Wu, Q., Zhang, G., Liu, C., Prevedouros, P.D., Ma, D.T., 2018. Exploring driver injury severity patterns and causes in low visibility related single-vehicle crashes using a finite mixture random parameters model. Anal. Methods Accid. Res. 20, 1–14.
Li, Z., Ci, Y., Chen, C., Zhang, G., Wu, Q., Qian, Z.S., Prevedouros, P.D., Ma, D.T., 2019a. Investigation of driver injury severities in rural single-vehicle crashes under rain conditions using mixed logit and latent class models. Accid. Anal. Prev. 124, 219–229.
Li, Z., Liao, H., Tang, R., Li, G., Li, Y., Xu, C., 2023. Mitigating the impact of outliers in traffic crash analysis: A robust Bayesian regression approach with application to tunnel crash data. Accid. Anal. Prev. 185, 107019.
Li, Y., Song, L., Fan, W.D., 2021. Day-of-the-week variations and temporal instability of factors influencing pedestrian injury severity in pedestrian-vehicle crashes: A random parameters logit approach with heterogeneity in means and variances. Anal. Methods Accid. Res. 29, 100152.
Li, Z., Wu, Q., Ci, Y., Chen, C., Chen, X., Zhang, G., 2019b. Using latent class analysis and mixed logit model to explore risk factors on driver injury severity in single-vehicle crashes. Accid. Anal. Prev. 129, 230–240.
Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Anal. Methods Accid. Res. 11, 1–16.
McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. Front. Econom. 105–142.
McFadden, D., 1981. Econometric models of probabilistic choice. Struct. Anal. Discret. Data Econom. Appl. 198272.
Owen, A.B., 2017. A randomized halton algorithm in R. arXiv preprint arXiv:1706.02808.
Poudel, N., Singleton, P.A., 2021. Bicycle safety at roundabouts: a systematic literature review. Transp. Rev. 41 (5), 617–642.
Samerei, S.A., Aghabayk, K., Mohammadi, A., Shiwakoti, N., 2021. Data mining approach to model bus crash severity in Australia. J. Saf. Res. 76, 73–82.
Se, C., Champahom, T., Jomnonkwao, S., Chaimuang, P., Ratanavaraha, V., 2021a. Empirical comparison of the effects of urban and rural crashes on motorcyclist injury severities: A correlated random parameters ordered probit approach with heterogeneity in means. Accid. Anal. Prev. 161, 106352.
Se, C., Champahom, T., Jomnonkwao, S., Karoonsoontawong, A., Ratanavaraha, V., 2021b. Temporal stability of factors influencing driver-injury severities in single-vehicle crashes: A correlated random parameters with heterogeneity in means and variances approach. Anal. Methods Accid. Res. 32, 100179.
Se, C., Champahom, T., Jomnonkwao, S., Kronprasert, N., Ratanavaraha, V., 2022. The impact of weekday, weekend, and holiday crashes on motorcyclist injury severities: Accounting for temporal influence with unobserved effect and insights from out-of-sample prediction. Anal. Methods Accid. Res. 36, 100240.
Shea, J., Torgovitsky, A., 2021. Ivmte: An R Package for Implementing Marginal Treatment Effect Methods. University of Chicago, Becker Friedman Institute for Economics Working Paper.
Song, L., Li, Y., Fan, W.D., Wu, P., 2020. Modeling pedestrian-injury severities in pedestrian-vehicle crashes considering spatiotemporal patterns: Insights from different hierarchical Bayesian random-effects models. Anal. Methods Accid. Res. 28, 100137.
Tierney, L., Rossini, A., Li, N., Sevcikova, H., 2008. Snow: Simple network of workstations. R package version 0.3-3, URL http://CRAN.R-project.org/package=snow.
Train, K., 2000. Halton sequences for mixed logit.
Train, K.E., 2009. Discrete Choice Methods with Simulation. Cambridge University Press.
Useche, S.A., Alonso, F., Montoro, L., Esteban, C., 2019. Explaining self-reported traffic crashes of cyclists: An empirical study based on age and road risky behaviors. Saf. Sci. 113, 105–114.
Wang, C., Easa, S.M., Song, D., Chen, F., Xiao, F., Cheng, J., 2023a. Alternative unobserved heterogeneity models to analyze injury severity of expressway crashes in different tunnel types. Tunn. Undergr. Space Technol. 140, 105245.
Wang, C., Ijaz, M., Chen, F., Easa, S.M., Zhang, Y., Cheng, J., Zahid, M., 2023b. Temporal assessment of injury severities of two types of pedestrian-vehicle crashes using unobserved-heterogeneity models. J. Transp. Saf. Secur. 1–50.
Wang, C., Ijaz, M., Chen, F., Zhang, Y., Cheng, J., Zahid, M., 2022. Evaluating gender differences in injury severities of non-helmet wearing motorcyclists: Accommodating temporal shifts and unobserved heterogeneity. Anal. Methods Accid. Res. 36, 100249.
Yan, X., He, J., Zhang, C., Liu, Z., Wang, C., Qiao, B., 2021. Temporal analysis of crash severities involving male and female drivers: A random parameters approach with heterogeneity in means and variances. Anal. Methods Accid. Res. 30, 100161.
Yang, C., Chen, M., Yuan, Q., 2021. The application of XGBoost and SHAP to examining the factors in freight truck-related crashes: An exploratory analysis. Accid. Anal. Prev. 158, 106153.
Ypma, J., 2014. Introduction to nloptr: an R interface to NLopt. R Package 2.
Yu, H., Li, Z., Zhang, G., Liu, P., 2019. A latent class approach for driver injury severity analysis in highway single vehicle crash considering unobserved heterogeneity and temporal influence. Anal. Methods Accid. Res. 24, 100110.
Yu, H., Li, Z., Zhang, G., Liu, P., Ma, T., 2021a. Fusion convolutional neural network-based interpretation of unobserved heterogeneous factors in driver injury severity outcomes in single-vehicle crashes. Anal. Methods Accid. Res. 30, 100157.
Yu, M., Ma, C., Shen, J., 2021b. Temporal stability of driver injury severity in single-vehicle roadway departure crashes: A random thresholds random parameters hierarchical ordered probit approach. Anal. Methods Accid. Res. 29, 100144.
Yu, H., Yuan, R., Li, Z., Zhang, G., Ma, D.T., 2020. Identifying heterogeneous factors for driver injury severity variations in snow-related rural single-vehicle crashes. Accid. Anal. Prev. 144, 105587.
Zamani, A., Behnood, A., Davoodi, S.R., 2021. Temporal stability of pedestrian injury severity in pedestrian-vehicle crashes: New insights from random parameter logit model with heterogeneity in means and variances. Anal. Methods Accid. Res. 32, 100184.
Zou, R., Yang, H., Yu, W., Yu, H., Chen, C., Zhang, G., Ma, D.T., 2023. Analyzing driver injury severity in two-vehicle rear-end crashes considering leading-following configurations based on passenger car and light truck involvement. Accid. Anal. Prev. 193, 107298.