A New Piecewise Multinomial Logit Model of Crash Severity with Accommodation of an Unknown Inflection Point

Di Yang[1], Zifeng Zhao[2], Kun Xie[3], Hong Yang[4], Tianyu Shen[5], Mansoureh Jeihani[6],

[1]Department of Transportation & Urban Infrastructure Studies, Morgan State University, 1700 E Cold Spring Ln, Baltimore, MD, USA, di.yang@morgan.edu
[2]Department of Information Technology, Analytics, and Operations, and Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Holy Cross Dr, Notre Dame, IN, USA, zzhao2@nd.edu
[3]Department of Civil & Environmental Engineering, Old Dominion University, 5115 Hampton Blvd, Norfolk, VA, USA, kxie@odu.edu
[4]Department of Electrical & Computer Engineering, Old Dominion University, 5115 Hampton Blvd, Norfolk, VA, USA, hyang@odu.edu
[5]Department of Transportation & Urban Infrastructure Studies, Morgan State University, 1700 E Cold Spring Ln, Baltimore, Baltimore, MD, USA, tishe7@morgan.edu
[6]Department of Transportation & Urban Infrastructure Studies, Morgan State University, 1700 E Cold Spring Ln, Baltimore, Baltimore, MD, USA, mansoureh.jeihani@morgan.edu

Extended Abstract

Multinomial logit regression (MNL)-based models have been applied extensively to model the relationship between crash severity and its contributing factors (*1*). Despite the methodological advancements around MNL, traditional MNL assumes a linear relationship between crash contributing factors and utility with a constant sensitivity among contribuing factors. However, employing a constant sensitivity scheme among contributing factors may not reflect the complex relationship between crash severity and contributing factors. To the best of our knowledge, there is still a lack of models that provide modeling flexibility to accommodate asymmetric sensitivities of crash contributing factors. Thus, this study proposes a new piecewise MNL model that can accommodate the asymmetric sensitivites by introducing the concept of unknown inflection points.

The crash data comprises traffic crashes that occurred in Maryland in 2019. The raw crash data was initially preprocessed to eliminate entries with invalid data entries. Subsequently, crash contributing factors were reclassified to generate binary variables known to affect crash severity levels. AADT (Average Annual Daily Traffic) values for each crash record were obtained by mapping each crash to the nearest road segment. The final dataset comprises 24,910 crashes and includes the following crash contributing factors: daylight presence, lights on in darkness, rain, snow, fog, presence of horizontal curves, presence of vertical curves, presence of road defects, presence of road divisions, airbag deployment, driver impairment, driver's gender, seat belt usage, and AADT.

In the general case of a multinomial logit model of crash injury severity outcomes (*2*), the probability of crash i of severity category k is represented by severity probability function, T_{ki}, as shown below: $T_{ki} = \alpha_k + \sum_{j=1}^{J} \beta_{kj} x_{ji} + \varepsilon_{ki}$, where, α_k is a constant parameter for crash severity category k; β_{kj} represents the jth variable for crash severity category k; $k = 1,\ldots,K$ ($K = 3$ in this paper) representing all the three severity levels: property damage only (PDO), injury, and fatal; x_{ji} represents the jth variable for crash i; ε_{ki} is a random error term following the Type 1 generalized extreme value (i.e., Gumbel) distribution; $i = 1,\ldots,n$ where n is the total number of crash events.

Without loss of generality, let's assume that the last independent variable x_J will have the unknown inflection points. Assuming that there exists an *unknown* inflection point $\tau_J > 0$ such that

$$T_{ki} = \alpha_k + \sum_{j=1}^{J-1} \beta_{kj} x_{ji} + \beta_{kJ1} x_{Ji} + (\beta_{kJ2} - \beta_{kJ1})(x_{Ji} - \tau_J)_+ + \varepsilon_{ki}$$
$$= \begin{cases} T_{ki1} := \alpha_k + \sum_{j=1}^{J-1} \beta_{kj} x_{ji} + \beta_{kJ1} x_{jI} + \varepsilon_{ki}, & x_{Ji} \le \tau_J \\ T_{ki2} := \alpha_k + \sum_{j=1}^{J-1} \beta_{kj} x_{ji} + \beta_{kJ2} x_{Ji} - (\beta_{kJ1} - \beta_{kJ2}) \tau_J + \varepsilon_{ki}, & x_{Ji} \ge \tau_J \end{cases} \quad (1)$$

where, $(x_{Ji} - \tau_J)_+ = \max(x_{Ji} - \tau_J, 0)$ and $\beta_{kJ1} \ne \beta_{kJ2}$ are two distinct regression coefficients for x_J. The flexibility of the proposed model, named MNL with unknown inflection points (MNL-UIP) in (1) stems from the unknown inflection point τ_J, which allows for changes in the effect of x_J at an unrestricted location. By design, T_{ki} is a continuous function, however, it is non-differentiable at τ_J due to the slope change (i.e., both the left and right derivatives exist but not the same). Thus, in this study, we adopt a profile likelihood strategy to estimate regression parameters.

To validate the proposed MNL-UIP model, we compare MNL-UIP with the traditional MNL model using the Akaike Information Criterion (AIC) (*3*). The AIC values of MNL and MNL-UIP are 30,820 and 30,800, respectively. Thus, MNL-UIP model is the better model since its AIC value is lower than that of MNL. The estimated coefficients for log(AADT) and the inflection point are shown below. All estimated regression coefficients are statistically significant according to 0.05 significance level. It can be seen that compared to PDO as the baseline, for each crash severity level, the effects of log(AADT) are different between log(AADT) < 9.941 and log(AADT) > 9.941. Furthermore, for $\hat{\beta}_{kJ1}$, the signs even flipped from injury to fatal.

Crash Severity Levels (with PDO as the baseline)	$\hat{\beta}_{kJ1}$	$\hat{\beta}_{kJ2}$	τ_J
Injury	0.016	-0.172	9.941
Fatal	-0.011	-0.097	

Findings from this study suggest that the proposed MNL-UIP model can capture more variations in the effects of crash contributing factors compared with the traditional MNL, while at the same time maintains model interpretability and tractability, which is often required in crash severity modeling. Our results also indicate that ignoring varying sensitivities among crash-contributing factors could result in biased estimations of regression coefficients and impede our understanding of the relationship between crash severity and these contributing factors. This study adds to the crash severity literature a new methodological approach to capture asymmetric sensitivities among crash contributing factors, which has the potential to enhance our understanding of the relationship between contributing factors and crash severity levels and facilitate the development of new safety-related countermeasures.

Keywords: Multinomial logic regression, inflection points, piecewise linear, crash severity.

References

[1] Savolainen, P., F. Mannering, D. Lord, and M. Quddus. The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. *Accident Analysis and Prevention,* Vol. 43, No. 5, 2011, pp. 1666-1676.
[2] Ye, F., and D. Lord. Comparing three commonly used crash severity models on sample size requirements: Multinomial logit, ordered probit and mixed logit models. *Analytic methods in accident research,* Vol. 1, 2014, pp. 72-85.
[3] Akaike, H. A new look at the statistical model identification. *IEEE transactions on automatic control,* Vol. 19, No. 6, 1974, pp. 716-723.