# An Enhanced Aggregate Framework To Model Crash Frequency By Accommodating Zero Crashes By Crash Type

**Md Istiak Jahan**
Graduate Student
Department of Civil, Environmental & Construction Engineering
University of Central Florida
Email: mdistiakjahan@knights.ucf.edu
ORCiD number: 0000-0002-4056-7816

**Tanmoy Bhowmik***
Postdoctoral Scholar
Department of Civil, Environmental & Construction Engineering
University of Central Florida
Tel: 407-927-6574
Email: tanmoy78@knights.ucf.edu
ORCiD number: 0000-0002-0258-1692

**Naveen Eluru**
Professor
Department of Civil, Environmental & Construction Engineering
University of Central Florida
Tel: 407-823-4815
Email: naveen.eluru@ucf.edu
ORCiD number: 0000-0003-1221-4113

Submitted to: Transportation Research Record: Journal of the Transportation Research Board

* Corresponding Author

**ABSTRACT**
In recent years, joint count and fractional split model structure-based approaches have emerged as a credible alternative for multivariate crash frequency dependent variables. However, current approaches in the fractional split theme have a limitation. The fractional split component of these frameworks allocates proportion to all crash configurations. It is possible that across spatial units, several crash configurations might have a large share of zero crashes. In the traditional multivariate context, in the presence of high share of zeros, researchers employ zero-inflated or hurdle variants such as zero inflated negative binomial model. The current research effort improves the fractional split based multivariate model systems with an explicit consideration for the potential presence of zeros by crash configuration. The newly included binary component can be employed to identify safer (or riskier) zones by crash configuration. The framework also accommodates for unobserved heterogeneity across the components of the model system. The proposed model structure is estimated using zonal data from the Central Florida region for 2016. The model considered 6 crash types including rear-end, angle, sideswipe, single-vehicle, multi-vehicle (3 or more), and non-motorized crashes. The model estimation is conducted using an exhaustive set of independent variables. The model results clearly highlight the importance of accommodating zero crashes by crash type in the analysis. The model exercise is further augmented with a validation analysis.

**Keywords:** Negative binomial- multinomial fractional split model, binary logit model, zero crash region, crash type, and crash count.

## INTRODUCTION

In the United States, after a period of sustained reduction in fatalities, the number of motor vehicle crash associated fatalities have started to increase in recent years. The number of roadway crash fatalities amounted to more than 40,000 in 2021; an increase of 10.5% from the previous year (*1*). Given the significant societal, emotional, and economic impacts of roadway crashes, it is important that evidence-based solutions are applied to reduce the number of crashes and their potential consequences. A major tool employed to advance road safety is the development of data-driven econometric models to identify factors affecting crash occurrence and crash consequence. Crash frequency analysis has been employed widely for macroscopic and microscopic safety analysis. Macroscopic analysis examines factors influencing crash frequency at an aggregate spatial resolution (such as TAZ, roadway segment, and intersection). In microscopic analysis, frequency models are developed at the facility resolution (such as segments and roadways). Macroscopic analysis is useful for identifying long-term planning and safety issues influencing crash occurrence. The current paper contributes to furthering crash frequency literature by developing improved methodology for macroscopic crash frequency analysis.

Macroscopic crash frequency analysis has been widely applied in safety literature to quantify the impact of various spatially aggregated independent variables on crash occurrence (*2*–*7*). Initial approaches employed Poisson and Negative Binomial (NB) models for analyzing total crash frequency. In recent years, other advanced approaches such as generalized linear model hurdle models (Poisson and NB), geographic weighted regression models (Poisson and NB) and hierarchical Bayesian spatiotemporal random parameters approach) and machine learning (such as random forest) tools have been employed (*2*, *5*, *8*–*12*). However, as recognized in different studies modeling total crashes as a homogenous dependent variable can neglect the different impacts of independent variables on different crash configurations (by crash type and/or severity)(see (*13*, *14*) for details). Hence, the total crash dependent variable was partitioned into crash frequency by crash configuration (type or severity). The partitioning results in multiple dependent variables for each spatial unit. The suite of univariate models is not appropriate for modeling multiple dependent variables by spatial unit. Hence, researchers employed multivariate models that recognize the presence of common unobserved factors that influence spatial unit specific dependent variables. The approaches employed can be categorized into two groups: <u>first,</u> model systems employing multivariate versions of univariate models such as multivariate Poisson-lognormal model (*15*, *16*), random parameter multivariate NB model (*17*), copula-based multivariate NB model (*17*, *18*), copula-based random parameter multivariate NB model (*17*), multivariate multiple risk source regression model (*19*), and Bayesian multivariate hierarchical spatial joint model (*20*). <u>Second,</u> model systems employing variants of the joint count and fractional split framework such as negative binomial-ordered probit fractional split model (*21*–*23*), negative binomial-ordered logit fractional split model (*24*), and negative binomial-multinomial logit fractional split model (*25*–*27*).

The two methods offer different approaches to model multivariate crash frequency variables. The traditional multivariate approach employs a count propensity for each crash variable and the interaction among dependent variables is accommodated through interaction of common unobserved factors. On the other hand, as discussed in our earlier work (*17*, *26*), the fractional split model structure employs a total crash model and a fractional split model that determines the proportion of crashes by configuration. The approach allows for independent variables to directly interact across crash configurations. Thus, it provides a potentially distinct alternative model structure for crash frequency models by crash configuration.

### Current Study

Earlier research has established that fractional split model systems perform as well (if not better) than the traditional multivariate approaches with a parsimonious specification (see (*15*, *26*) for examples of comparison efforts). However, earlier approaches in the fractional split theme have a limitation. As it is employed right now, the fractional split component allocates proportion to all crash configurations. It is possible that across spatial units, several crash configurations might have large share of zero crashes. In the traditional multivariate context, when the presence of such zeros needs to be accommodated, researchers employ zero-inflated or hurdle variants such as zero inflated negative binomial model (*10*, *28–32*), zero inflated Poisson model (*10*, *11*, *32*), spatiotemporal random effect zero inflated negative binomial model (*33*), multivariate random parameters zero-inflated negative binomial regression model (*34*), zero inflated hierarchical ordered probit model (*35*), and hurdle negative binomial model (*36–39*).

The current research effort furthers safety methodology by improving the fractional split based multivariate model systems with an explicit consideration for the potential presence of zeros by crash configuration. The modified framework with an additional component for zero crashes results in three components: (1) total crash NB component, (2) binary logit (BL) component by crash configuration and (3) multinomial logit fractional split (MNLFS) crash proportion component. The binary component can be employed to identify safer (or riskier) zones by crash configuration. The framework also accommodates for unobserved heterogeneity across the three components of the model system. The study contributes to empirical literature on safety by estimating the proposed model system for 6 crash types including rear-end, angle, sideswipe, single-vehicle, multi-vehicle (3 or more), and non-motorized crashes. The model estimation is conducted using a host of independent variables including roadway-, traffic-, land use, built environment, and socio-demographic characteristics, which will contribute to evaluate their impact on crash frequency for different collision types. The model results clearly highlight the importance of accommodating zero crashes by crash type in the analysis. The model exercise is augmented with a validation analysis.

## METHODOLOGY

The Joint model developed in the study consist of three components: 1) a NB model employed to analyze the total number of crashes,2) a binary logit model for determining zones with zero or non-zero crashes by crash type, (3) a MNLFS model employed to study the proportion of crashes by crash types for non-zero crash frequency zones. For ease of presentation, we discuss the framework by component.

### Count Component

Let us assume $i(i = 1,2,3, \dots, N)$ as the index for TAZ, $j$ as an index representing each crash type and $K$ represents total crashes at TAZ level, therefore, $j = 1,2,3, \dots, J; j\epsilon K \; and \; K = \sum_{j=1}^{J} j$. In our study, the values of the $j$ were assigned as follows: rear-end ($j = 1$), angle ($j = 2$), sideswipe ($j = 3$), single-vehicle crash ($j = 4$), multiple-vehicle crash ($j = 5$), and non-motorized crash ($j = 6$).Using these notations, the equation for estimating total crash counts at different crash level using NB is as follows:

$$P(c_{iK}) = \frac{\Gamma\left(c_{iK} + \frac{1}{\alpha_K}\right)}{\Gamma(c_{iK} + 1)\Gamma\left(\frac{1}{\alpha_K}\right)} \left(\frac{1}{1 + \alpha_K \mu_{iK}}\right)^{\frac{1}{\alpha_K}} \left(1 - \frac{1}{1 + \alpha \mu_{iK}}\right)^{c_{iK}} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots. (1)$$

where, $c_{iK}$ be the index for total crash counts occurring over a period of time in TAZ $i$. $P(c_{iK})$ is the probability that TAZ $i$ has $c_{iK}$ number of crashes. $\Gamma(\cdot)$ is the gamma function, $\alpha_K$ is NB over dispersion parameter and $\mu_{iK}$ is the expected number of crashes occurring in TAZ $i$ over a given time period. We can express $\mu_{iK}$ as a function of explanatory variables by using a log-link function as follows:

$$\mu_{iK} = E(c_{iK}|\boldsymbol{x}_i) = exp\big((\boldsymbol{\theta} + \boldsymbol{\varrho}_i)\boldsymbol{x}_i + \phi_i + \psi_{ij}\big) \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (2)$$

where, $\boldsymbol{x}_i$ is a vector of explanatory variables (including the constant) associated with TAZ $i$. $\boldsymbol{\theta}$ is a vector of coefficients to be estimated. $\boldsymbol{\varrho}_i$ is a vector of unobserved factors on crash count propensity for TAZ $i$ and its associated zonal characteristics assumed to be a realization from standard normal distribution: $\boldsymbol{\varrho}_i \sim N(0, \boldsymbol{\varsigma}^2)$. $\phi_i$ is a gamma distributed error term with mean 1 and variance $\alpha_K$. $\psi_{ij}$ captures unobserved factors that simultaneously impact total number of crashes and proportion of crashes by crash types for TAZ $i$.

**Binary Logit Component**

For the binary logit framework, the probability expression is as follows:

$$\Lambda[y_{ij}] = \begin{cases} \pi_{ij} & y_{ij} > 0 \\ 1 - \pi_{ij} & y_{ij} = 0 \end{cases} \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (3)$$

where $\Lambda[y_{ij}]$ represents the probability that TAZ $i$ will have the corresponding crash type $j$ or not (yes/no) and it will be determined based on $\pi_{ij}$ . $y_{ij}$ is the observed fraction of crashes by crash type $j$ $(j = 6)$ in TAZ $i$. With this notation, the equation for $\pi_{ij}$ is as follows:

$$\pi_{ij} = \frac{exp\big(\gamma\boldsymbol{\eta}_{ij}\big)}{1 + exp\big(\gamma\boldsymbol{\eta}_{ij}\big)} \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (4)$$

where, $\boldsymbol{\eta}_{ij}$ is a vector of attributes (including the constant) and $\gamma$ is a conformable parameter vector to be estimated.

**Fractional Split Component**

The modeling of crash proportions by crash types is undertaken using the MNLFS model. We defined the proportion of crash types in TAZs as the dependent variable in MNLFS framework. In estimating the model, we assume that the sum of the proportions across a TAZ is equal to unity and each proportion of crash types in traffic crashes ranges between zero and one. Therefore assuming $y_{ij}$ be the fraction of crashes by crash type $j$ $(j = 6)$ in TAZ.

$$0 \leq y_{ij} \leq 1, \qquad \sum_{j=1}^{J} y_{ij} = 1 \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (5)$$

Let the fraction $y_{ij}$ be a function of a vector $d_{ij}$ of relevant explanatory variables associated with attributes of TAZ $i$.

$$E[y_{ij}|d_{ij}] = G_j(\cdot) \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (6)$$

$$0 < G_j(\cdot) < 1, \sum_{j=1}^{J} G_j(\cdot) = 1 \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (7)$$

where, $G_j(\cdot)$ is a predetermined function. The properties specified in equation 7 for $G_j(\cdot)$ warrant that the predicted fractional crash types will range between 0 and 1 and will add up to 1 for each TAZ. In this study, a MNL functional form for $G_j$ in the fractional split model of equation 7. Then equation 7 is rewritten as:

$$E(y_{ij}|d_{ij}) = G_j(\cdot) = \frac{\exp(v_{ij} * \mathrm{y_{ij}})}{\sum_{j=1}^{J} \exp(v_{ij} * \mathrm{y_{ij}})}, j = 1,2,3, \ldots.., \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (8)$$

where, $v_{ij}$ is the propensity for crash type $j$ in TAZ $i$ and it is specified as the following linear form:

$$v_{ij} = (\boldsymbol{\beta_j} + \boldsymbol{\rho_{ij}})d_{ij} + \xi_{ij} \pm \psi_{ij} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (9)$$

where, $\boldsymbol{d}_{ij}$ is a vector of attributes, $\boldsymbol{\beta}_j$ is the corresponding vector of coefficients to be estimated for crash type $j$. $\boldsymbol{\rho}_{ij}$ is a vector of unobserved factors assumed to be a realization from standard normal distribution: $\boldsymbol{\rho} \sim N(0, \boldsymbol{\nu}_j{}^2)$. $\xi_{ij}$ is the random component assumed to follow a Gumbel type-1 distribution. $\psi_{ij}$ term generates the correlation between equations for total number of crashes and crash proportions by crash types. The $\pm$ sign in front of $\psi_{ij}$ in equation 9 indicates that the correlation in unobserved zonal factors between total crashes and various crash types may be positive or negative. The positive sign implies that TAZ with higher number of total crashes is more likely to experience higher number of crashes for the corresponding crash types, impacted by the same unobserved factor. Alternatively, the negative sign reveals an opposing trend, that is TAZ with higher number of crashes will have lower risk for the corresponding crash types. The appropriate sing (+/-) will be determined based on the model fit. Further, we recognize the correlation can vary across TAZs and hence, we parameterize it as a function of observed attributes:

$$\psi_{ij} = \boldsymbol{\Theta_j} \boldsymbol{t}_{iJ} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (10)$$

where, $\boldsymbol{t}_{ij}$ is a vector of exogenous variables, $\boldsymbol{\Theta_j}$ is a vector of unknown parameters to be estimated (including a constant).

**Model Estimation**

In examining the model structure of total crash count and proportion of crashes by crash types, it is necessary to specify the structure for the unobserved vectors $\boldsymbol{\varsigma}, \boldsymbol{\rho}$ and $\boldsymbol{\Theta}$ represented by $\mho$. In this paper, we assume that these elements are drawn from independent realization from normal population: $\mho \sim N(0, (\boldsymbol{\varsigma}^{\mathbf{2}}, \boldsymbol{\nu}_j{}^2, \aleph_j^2))$. Thus, conditional on $\mho$, the likelihood function for the joint probability can be expressed as:

$$\mathcal{L}_i = \int_{\mho} P(c_{iK}) \times \prod_{j=1}^{J} \left(E(y_{ij}|d_{ij})\right)^{\varpi_i y_{ij}} f(\mho) d\Omega \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (11)$$

where, $\varpi_i$ is a dummy with $\varpi_i = 1$ if TAZ $i$ has at least one crash over the study period and 0 otherwise. $y_{ij}$ is the proportion of crashes in crash type category $j$. Finally, the log-likelihood function is:

$$\mathcal{LL} = \sum_i Ln(L_i) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. (12)$$

All the parameters in the model are estimated by maximizing the logarithmic function $\mathcal{LL}$ presented in equation 12. The parameters to be estimated in the joint model are: $\boldsymbol{\theta}, \alpha_K, \boldsymbol{\beta_j}, \boldsymbol{\nu}_j$ and $\boldsymbol{\aleph_j}$.To estimate the proposed joint models, we apply Quasi-Monte Carlo simulation techniques based on the scrambled Halton sequence to approximate this integral in the likelihood function and maximize the logarithm of the resulting simulated likelihood function across individuals (see (*40–42*) for examples of Quasi-Monte Carlo approaches in literature). The model estimation routine is coded in GAUSS Matrix Programming software (*43*).

## DATA PREPARATION

### Data Source

The data for our analysis was drawn from the Central Florida region for the year 2016. The data were obtained from Florida Department of Transportation (FDOT), Crash Analysis Reporting System and Signal Four Analytics databases. The dataset contains the records for 114,458 motorized and 3,413 non-motorized crashes. Central Florida region consists of 4747 TAZs, and all the crash records were assigned to the corresponding TAZ using the Geographic Information System (GIS). Motorized crashes were classified into 5 categories based on crash type as single vehicle crashes, two vehicle crashes including rear end, angle, sideswipe, single-vehicle, and multiple-vehicle (involving 3 or more vehicles) crash.

### Dependent Variable Distribution

A summary of the dependent variable distribution is presented in Figure 1. The figure reports two measures, the number of TAZs with zero crashes by crash type and the mean and standard deviation of crashes by crash type (for all TAZs). From the numbers, it is evident that a large share of TAZs have zero crashes. Among the crash types, non-motorized crashes have the highest share of zero crashes while rear-end crashes have the lowest share of zero crashes. These descriptive statistics clearly indicate the value of a model framework that can accommodate for zero crash state by crash type. Among the crash types, rear end crashes have the highest mean for crashes (10.95) while and non-motorized crashes have the smallest mean (0.72). From the 4,747 TAZs, 3,815 (80%) were used for model estimation, and the rest 932 (20%) TAZs were set aside for validation.

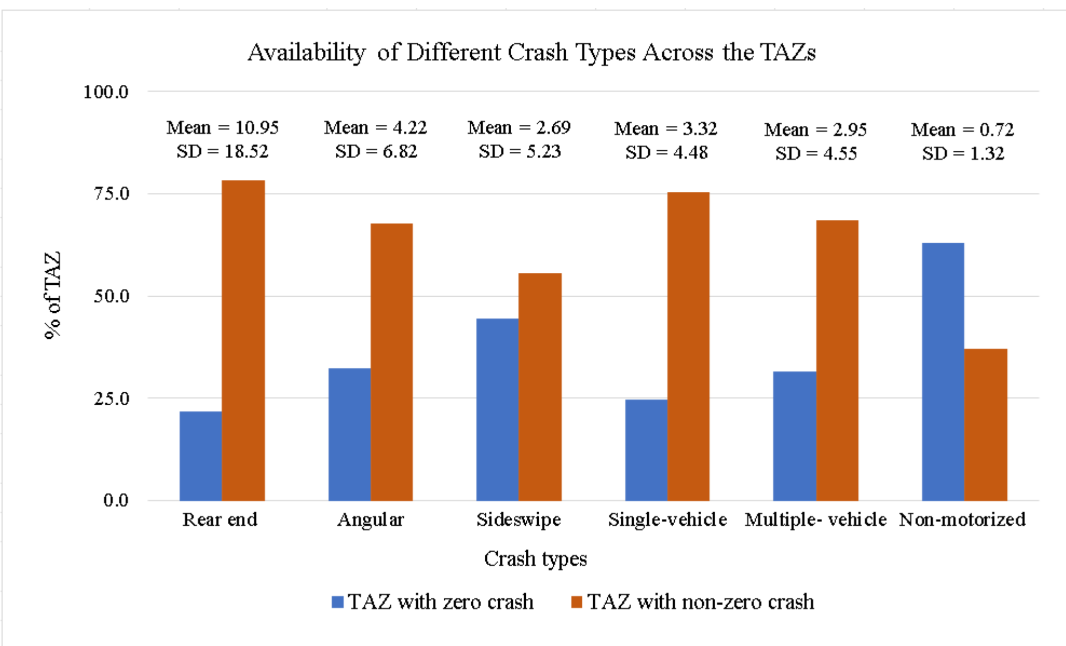


**Figure 1 Distribution of Crashes at Different Crash Types**

**Independent Variables**

The crash data were augmented with a host of independent variables from the following categories: a) roadway characteristics, b) traffic characteristics, c) land-use characteristics, d) built environment characteristics, and e) socio-demographic characteristics. The data were compiled from multiple sources including FDOT Transportation Statistics Division, US Census Bureau, American Community Survey and Florida Geographic Data Library databases. The independent variables were aggregated at corresponding TAZ resolution using GIS. A summary statistic (mean and standard deviation) and variable definitions are presented in Table 1. In the model estimation process several functional forms, such as – Z-scores, and proportion of variable categories were employed, and retained in the final specification based on their statistically significance (at 90% significance level).

**TABLE 1 Variable Summary Statistics (Zonal level)**

| Variables | Variable description | Number of zones = 4,747 | |
|---|---|---|---|
| | | **Mean** | **Standard Deviation** |
| Proportion of rear end crash | Total number of rear end crashes/ Total crashes in the corresponding TAZ | 0.33 | 0.24 |
| Proportion of angle crash | Total number of angle crashes/ Total crashes in the corresponding TAZ | 0.15 | 0.17 |
| Proportion of sideswipe crash | Total number of sideswipe crashes/ Total crashes in the corresponding TAZ | 0.08 | 0.11 |
| Proportion of single-vehicle crash | Total number of single-vehicle crashes/ Total crashes in the corresponding TAZ | 0.19 | 0.23 |
| Proportion of multiple-vehicle (3 or more) crash | Total number of multiple-vehicle crashes/ Total crashes in the corresponding TAZ | 0.13 | 0.17 |
| Proportion of non-motorized crash | Total number of non-motorized crashes/ Total crashes in the corresponding TAZ | 0.03 | 0.08 |
| Proportion of 1-through traffic lane in | Length of 1-through traffic lanes/Total roadway length | 0.16 | 0.23 |
| No of bike lanes | Total number of roadways with bike lane in a TAZ | 0.49 | 1.82 |
| No of intersections | Total number of intersections in a TAZ | 9.82 | 10.09 |
| Proportion of urban road | Length of urban roadways/ Total roadways | 0.81 | 0.38 |
| Proportion of arterial road | Length of arterial roadways/ Total roadways | 0.38 | 0.39 |
| Average inside shoulder width | Average width of inside shoulder in feet | 0.29 | 0.45 |
| Average outside shoulder width | Average width of outside shoulder in feet | 0.96 | 0.58 |

| | | | |
|---|---|---|---|
| Average median width | Average median width in feet | 7.34 | 10.87 |
| Average sidewalk width | Average sidewalk width in feet | 2.09 | 1.79 |
| Proportion of road length with speed limit > 55 mph | Length of roadways with speed limit > 55 mph/ Total roadways | 0.09 | 0.17 |
| Average speed limit | Average speed limit of all roadways in a TAZ | 36.66 | 16.48 |
| Variation of speed limit | Variation in speed limit in a TAZ | 46.30 | 78.86 |
| Truck AADT | Annual average daily truck traffic | 1071.04 | 2091.00 |
| Vehicle miles traveled | Ln (Vehicle miles traveled+1) | 7.91 | 3.37 |
| Proportion of industrial area | Total industrial area/ Total area of the corresponding TAZ | 0.02 | 0.08 |
| Proportion of public area | Total public (government institution) area/ Total area of the corresponding TAZ | 0.14 | 0.20 |
| *Land-use mix* | $Land\ use\ mix = \left[\frac{-\sum_k (p_k * ln(p_k))}{ln\ (N)}\right]$ *where, k is the category of land-use, p is the proportion of the developed land area for specific land-use category, N is the number of land-use categories (13)* | *0.37* | *0.22* |
| No of entertainment establishments | Total number of entertainment establishments in a TAZ | 0.55 | 1.13 |
| No of restaurants | Total number of restaurants in a TAZ | 1.66 | 3.57 |
| No of shopping center | Total number of shopping centers in a TAZ | 2.43 | 5.50 |
| Proportion of individuals aged more than 65 years old | Number of individuals aged more than 65 years old / Total population of the corresponding TAZ | 0.21 | 0.11 |
| Proportion of African Americans | Number of African Americans / Total population of the corresponding TAZ | 0.14 | 0.16 |
| Proportion of commuters use walking as the principal mode | Number of commuters using walking as the main mode/ Total number of commuters at the corresponding TAZ | 0.02 | 0.03 |

## EMPIRICAL ANALYSIS

### Model Specification and Goodness of Fit

This study involves estimation of two model structures: 1) NB-MNLFS model without a component for zero crash state (Model-1), and 2) proposed model system with a binary logit model for considering zero crash state (Model-2). The model estimation process involved the following steps. First, an NB model was developed for the crash count estimation. Second, six binary logit models were developed to recognize the non-zero crash regions for six specific crash types. Third, two different MNLFS models were estimated considering zero crashes and without considering zero crashes. The models were compared based on log-likelihood (LL) and Bayesian Information Criterion (BIC) values. The LL (BIC) values are as follows: (a) Model 1: -20,455.2 (41,355.8) and (b) Model 2: -19,292.2 (39,029.6). From these values, it is clear that the newly proposed Model 2 provides the best fit. The estimation results of the best fitted model are described.

***Joint NB-MNL Fractional Split Model***

Table 2 presents the estimated coefficients of the Joint NB-MNL fractional split model. The second column of the model provides the significant parameters of the NB model and the $3^{rd}$ to $8^{th}$ column provides significant parameters (at 90% significance level) of the MNL fractional split model.

***NB Component***

Crash frequency for different TAZs was estimated using the NB component of the joint framework. A positive (negative) coefficient variable indicates that, an increase in the variable will increase (decrease) crash frequency. The estimation results are discussed by variable categories as follows:

*Roadway Characteristics*

Several roadway characteristics were tested in our model. Among them proportion of arterial roads, proportion of urban roadways, number of intersections, and average inside shoulder width were found to offer significant impact on crash frequency. The positive coefficient of the proportion of arterial roads indicates an increased number of crashes with increasing proportion of arterial roads at the TAZ level (see (*17*, *26*) for similar results). Proportion of urban roadways also provided similar impact(see (*26*) for similar result). Further, as expected, an increased number of intersections are found to increase the total number of crashes. Earlier research also reported similar impacts of the intersection variable (*13*, *24*, *26*). Usually, intersections are likely to experience more crashes due to vehicle maneuvers through several conflict points. Finally, average inside shoulder width was found to have a positive impact on crash frequency (similar result was reported at (*26*).

*Traffic Characteristics*

In terms of traffic characteristics, only Vehicle Miles Traveled (VMT) was found to have significant impact on crash frequency. The positive sign indicates that increased traffic volume is more likely to increase the total number of crashes in TAZs ( see (*44*) for similar result).

*Land-use Characteristics*

Several land-use variables were tested in our model to capture the impact of surrounding land-use characteristics on crash frequency. Among them, the proportion of industrial and public areas are found to provide significant effect. However, they offered contradictory impact on crash frequency. The result indicates that TAZs with larger proportion of industrial area are more likely to experience more crashes, while TAZs with a larger proportion of public areas are found to have a lower number of crashes. Further, the random parameter for the proportion of public area in the crash count NB model was found to be statistically significant. The distributional parameter indicates that the overall impact of the corresponding variable on crash frequency is likely to be negative (83.4%).

*Built-environment Characteristics*

Only one variable - the number of entertainment establishments - was found to be statistically significant affecting crash frequency. The positive sign of the variable indicates that TAZs with a higher number of entertainment establishments are likely to have higher number of crashes.

*Socio-demographic Characteristics*

In terms of socio-demographic attributes, only two variables - proportion of African American individuals and proportion of people over 65 years - were found to have significant impact on crash frequency. The proportion of African American was found to offer positive effect, whereas the proportion of people over 65 years old provided a negative impact (see (*45*) for similar results).

***Binary Logit Component***

In this study, we employed binary logit model of each crash type for identifying the TAZs with non-zero crash regions for that corresponding type. The estimation results with significant coefficients (at 90% significance level) of the six binary logit models were shown in Table 3 and are discussed together by different variable categories.

*Roadway Characteristics*

Various roadway characteristics were found to be significant in each crash type model. Higher proportion of arterial roads, urban roadways, road length with 1-through traffic lane, abrupt variance in posted speed limit, and increased number of intersections in the TAZs were found to increase the likelihood of rear-end crash occurrence in a zone (similar result was found in (*46*)). In terms of angle crashes proportion of arterial roads and signal intensity offered a positive impact; however, proportion of road length with speed limit more than 55 mph provided a negative association. This result is indicative of the availability of well-maintained facilities on the roadways with higher speed limit. Sideswipe crashes are more likely to occur in TAZs with higher proportion of arterial roads, urban roads, increased number of bike lanes and increased signal intensity. The proportion of road length with speed limit > 55mph, and proportion of road length with 1-through traffic lane were found to offer positive impact on single-vehicle crash. This result is intuitive because, 1-through traffic lane provides confined spaces for vehicle mobility resulting in higher number of single-vehicle crashes. On the other hand, TAZ level multiple-vehicle crashes are positively associated with the proportion of arterial roads, and the proportion of road length with 1-through traffic lane; however, proportion of road length with speed limit > 55 mph offered a negative effect. In terms of non-motorized crashes in a TAZ, several variables from this category were found to be significant. Non-motorized crashes are more likely to occur in TAZs with higher proportion of arterial roadways, higher proportion of urban roadways, and increased number of intersections. On the contrary, higher proportion of road length with speed limit > 55 mph, increased number of bike lanes, and larger average sidewalk width decreases the occurrence of non-motorized crashes (similar result was reported in (*47*)). These results indicate that facilities for active transportation (walking, biking) such as the presence of bike lanes, and sidewalks offer safer environment for bicyclists and pedestrians.

**TABLE 2 Estimation Result of Joint NB-MNLFS Model**

| **Parameters** | **Crash frequency model** | **Rear end crash** | **Angle crash** | **Sideswipe** | **Single vehicle crash** | **Multiple vehicle crash** | **Non-motorized crash** |
|---|---|---|---|---|---|---|---|
| | *Coefficient (T- value)* | *Coefficient (T- value)* | *Coefficient (T- value)* | *Coefficient (T- value)* | *Coefficient (T- value)* | *Coefficient (T- value)* | *Coefficient (T- value)* |
| Intercept | 1.23 (10.68) | ---- | -0.10 (-1.67) | -0.59 (-8.69) | 0.14 (2.23) | 0.16 (2.03) | -0.42 (-4.73) |
| ***Roadway Characteristics*** | | | | | | | |
| Proportion of arterial road | 0.16 (3.17) | ---- | ---- | ---- | -0.35 (-8.05) | -0.11 (-2.55) | ---- |
| Number of intersections | 0.01 (6.57) | ---- | 0.01 (6.30) | -0.01 (-3.43) | ---- | 0.01 (3.38) | 0.01 (2.56) |
| Proportion of roadway with 1 through traffic lane | ---- | 0.32 (5.53) | ---- | ---- | ---- | ---- | ---- |
| Proportion of road length with speed limit > 55 mph | ---- | ---- | -0.75 (-5.70) | ---- | 0.32 (3.12) | -0.61 (-5.43) | -0.65 (-3.52) |
| ln (Average inside shoulder width) | 0.39 (8.90) | ---- | -0.23 (-6.21) | ---- | -0.18 (-5.22) | -0.21 (-5.56) | -0.28 (-5.11) |
| ln (Average outside shoulder width) | ---- | -0.19 (-7.42) | ---- | ---- | ---- | ---- | ---- |
| Proportion of urban roadway | 0.32 (4.80) | ---- | ---- | -0.20 (-3.89) | -0.79 (-16.07) | -0.35 (-6.52) | ---- |
| ***Traffic Characteristics*** | | | | | | | |
| ln (Vehicle miles travelled) | 0.15 (14.55) | ---- | -0.08 (-9.91) | -0.04 (-5.26) | ---- | -0.09 (-12.52) | -0.13 (-12.89) |
| ***Land-use Characteristics*** | | | | | | | |
| Land use mix | ---- | ---- | ---- | -0.15 (-2.06) | -0.28 (-3.93) | ---- | -0.39 (-4.06) |
| Proportion of industrial area | 0.52 (1.92) | ---- | ---- | ---- | ---- | ---- | ---- |
| Proportion of public area | -0.58 (-5.90) | ---- | ---- | ---- | ---- | ---- | ---- |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Standard deviation | 0.60 (3.42) | ---- | ---- | ---- | ---- | ---- | ---- |
| ***Built Environment Characteristics*** | | | | | | | |
| Z-score: Number of entertainment establishments | 0.14 (6.40) | ---- | ---- | ---- | ---- | ---- | ---- |
| Z-score: Number of shopping center | ---- | 0.03 (2.46) | ---- | ---- | -0.06 (-1.80) | ---- | ---- |
| Z-score: Number of restaurants | ---- | ---- | -0.05 (-3.70) | -0.02 (-1.86) | -0.20 (-7.97) | ---- | -0.20 (-9.75) |
| ***Socio-demographic Characteristics*** | | | | | | | |
| Proportion of African American individuals | 0.77 (6.61) | ---- | ---- | ---- | ---- | ---- | ---- |
| Proportion of people over 65 years old | -1.00 (-4.42) | ---- | ---- | ---- | ---- | ---- | ---- |
| Proportion of commuters use walking as principal mode of transportation | ---- | ---- | ---- | ---- | ---- | 0.91 (2.80) | 1.01 (2.77) |
| Dispersion parameter | 0.95 (21.39) | ---- | ---- | ---- | ---- | ---- | ---- |
| ***Correlations*** | | | | | | | |
| Correlation between total crash count and crash proportions of rear end, angle, and sideswipe | -0.38 (-10.69) | -0.38 (-10.69) | -0.38 (-10.69) | -0.38 (-10.69) | ---- | ---- | ---- |
| Correlation between total crash count and crash proportions of single vehicle and non-motorized crashes | 0.03 (1.92) | ---- | ---- | ---- | 0.03 (1.92) | ---- | 0.03 (1.92) |

**TABLE 3 Estimation of Binary Logit Model**

| **Parameters** | **Rear end** | **Angle** | **Sideswipe** | **Single vehicle crash** | **Multiple vehicle crash** | **Non-motorized crash** |
|---|---|---|---|---|---|---|
| | *Coefficient (T- value)* | *Coefficient (T- value)* | *Coefficient (T- value)* | *Coefficient (T- value)* | *Coefficient (T- value)* | *Coefficient (T- value)* |
| Intercept | -0.45 (-3.19) | 1.16 (12.12) | -0.86 (-6.93) | 0.36 (4.43) | 0.66 (7.25) | -2.67 (-16.83) |
| ***Roadway Characteristics*** | | | | | | |
| Proportion of arterial road | 0.22 (1.76) | 0.32 (3.13) | 0.42 (4.24) | ---- | 0.39 (4.00) | 0.26 (2.62) |
| Proportion of urban roadway | 0.54 (4.60) | ---- | 0.87 (8.41) | ---- | ---- | 0.92 (6.93) |
| Proportion of road length with speed limit > 55 mph | ---- | -1.19 (-5.53) | ---- | 1.58 (5.65) | -0.62 (-2.96) | -1.46 (-5.03) |
| Ln (variance in speed limit) | 0.13 (4.57) | ---- | ---- | ---- | ---- | ---- |
| Proportion of road length with 1 through traffic lane | 0.71 (3.67) | ---- | ---- | 0.32 (1.87) | 0.67 (4.10) | ---- |
| Number of intersections | 0.03 (3.60) | ---- | ---- | ---- | ---- | 0.04 (8.64) |
| Signal intensity | ---- | 1.40 (2.72) | 0.74 (1.65) | ---- | ---- | |
| Number of bike lanes | ---- | ---- | 0.06 (2.31) | ---- | ---- | -0.03 (-1.77) |
| Average sidewalk width | ---- | ---- | --- | ---- | ---- | -0.11 (-4.80) |
| ***Traffic Characteristics*** | | | | | | |
| Ln (VMT+1) | 0.17 (8.43) | ---- | ---- | ---- | ---- | 0.14 (7.34) |
| Proportion of heavy vehicle | ---- | 3.90 (4.74) | 6.94 (8.59) | 0.13 (11.10) | 6.25 (7.20) | ---- |
| ***Built Environment Characteristics*** | | | | | | |
| Z score: Number of shopping center | ---- | 3.13 (16.73) | 1.16 (8.70) | 0.71 (5.13) | 1.30 (8.02) | ---- |
| Z score: Number of restaurants | 1.84 (9.31) | ---- | 0.81 (7.33) | 0.55 (4.86) | 0.98 (6.98) | ---- |

*Traffic Characteristics*
Among traffic characteristics, VMT and proportion of heavy vehicle were found to be significant. VMT is associated with higher likelihood of rear-end and non-motorized crashes. This result is intuitive because increased traffic volume increases pedestrian and bicycle mobility. An increase in non-motorized modes of transportation causes disruption in driving which leads to rear end and non-motorized crashes (see (*46*, *47*) for similar result). On the other hand, higher proportion of heavy vehicle movement in a TAZ increases the occurrence of angle, sideswipe, single vehicle, and multiple vehicle crashes.

*Land-use Characteristics*
Several land-use attributes were tested in the binary logit models, but no significant variable was found.

*Built-environment Characteristics*
Two built-environment attributes, number of shopping centers, and number of restaurants, were found to offer significant impact in the binary logit model. It is noticeable that rear end crashes are more likely to occur in TAZs with increased number of restaurants. On the other hand, increased number of shopping centers increases the likelihood of angle crash.es Both variables - number of shopping centers and number of restaurants - are found to have a positive effect on sideswipe, single-vehicle, and multiple-vehicle crashes.

*Socio-demographic Characteristics*
Several socio-demographic attributes were tested in our models, but no significant variable was found.

***MNL Fractional Split Component***
The proportion of crashes at different crash types were estimated using MNL fractional split model. In our analysis proportion of rear-end crashes was generally considered as the base category for this model. A positive (negative) sign of any variable for any crash type indicates a higher (lower) likelihood of occurrence of that crash type compared to rear-end crash for a unit increase of the corresponding variable. The MNLFS model estimation results are discussed below by different variable categories.

*Roadway Characteristics*
Various roadway variables were found to be significant in the fractional split model. The proportion of arterial roads was found to be negatively associated with single-vehicle and multiple-vehicle crashes (relative to rear-end crashes). The insignificant impact for other crash categories indicates that the proportions of rear-end, angle, sideswipe, and non-motorized crashes are not differentially affected by the proportion of arterial roads. An increased number of intersections are likely to increase the proportion of angle, multiple-vehicle, and non-motorized crashes (see (*26*) for similar result). At the same time, this variable was found to decrease the proportion of sideswipe crashes. Sideswipe, single vehicle, and multiple vehicle crashes are less likely to occur in a TAZ with higher proportion of urban roadways. The proportion of roadways with a speed limit more than 55 mph were found to be positively associated with single vehicle crashes and negatively associated with angle, multiple, and non-motorized crashes (similar result was reported in (*13*)). A higher average inside shoulder width in a zone decreases the proportion of angle, single

vehicle, multiple-vehicle, and non-motorized crashes. In terms of outside shoulder, a higher average outside shoulder width decreases the proportion of rear-end crashes (see (*13*) for similar result). Finally, the proportion of roadways with 1-through traffic lane is positively associated with proportion of rear-end crashes.

*Traffic Characteristics*

Only one traffic characteristic – Vehicle Miles Traveled (VMT) – was found to be significant in our model. VMT offers a negative association with angle, sideswipe, multiple-vehicle, and non-motorized crashes. This result is intuitive because increased traffic volume on the roadways increases the proportion of rear-end crashes. However, the impact corresponding to the single vehicle crashes seems counter intuitive and requires further assessment in future research.

*Land-use Characteristics*

Several land-use characteristics were tested in this study. Among them only land-use mix was found to offer a significant effect. It is negatively associated with sideswipe, single-vehicle, and non-motorized crashes.

*Built-environment Characteristics*

Among the built-environment attributes, number of shopping centers and number of restaurants in the TAZ were found to offer a significant effect. Number of shopping centers offered a positive effect on rear-end crashes; however, it is negatively associated with single-vehicle crashes. The number of restaurants variable offered a negative association with angle, sideswipe, single vehicle, and non-motorized crashes.

*Socio-demographic Characteristics*

Several socio-demographic variables were tested in our model; however only one variable – proportion of commuters using walking as the principal mode of transportation – was found to be significant. This variable is associated with higher share of multiple-vehicle and non-motorized crashes. The result is quite interesting and needs further investigation.

***Unobserved Heterogeneity***

The final set of variables in Table 3 represents the unobserved heterogeneity across TAZs. In the current study, two common unobserved components were found to be significant. The first parameter represents the common correlation between total crash count and crash proportions of rear end, angle, and sideswipe. The second unobserved component corresponds to the correlation between total crash count and crash proportions of single vehicle and non-motorized crashes. These significant parameters lend support to the presence of unobserved factors between total crash count and proportions by crash types.

## MODEL PREDICTIVE PERFORMANCE

To illustrate the improved performance of our proposed model, we conducted a prediction exercise by total crash count and crash counts by crash type using validation dataset. Specifically, we compared the model performance of Model 1 (model that does not explicitly control for zero state) and Model 2 (model that explicitly controls for zero state). The predicted log-likelihood values for the two model systems are presented in Figure 2.  Figure 2 provides a comparison for total crashes and 6 crash types.

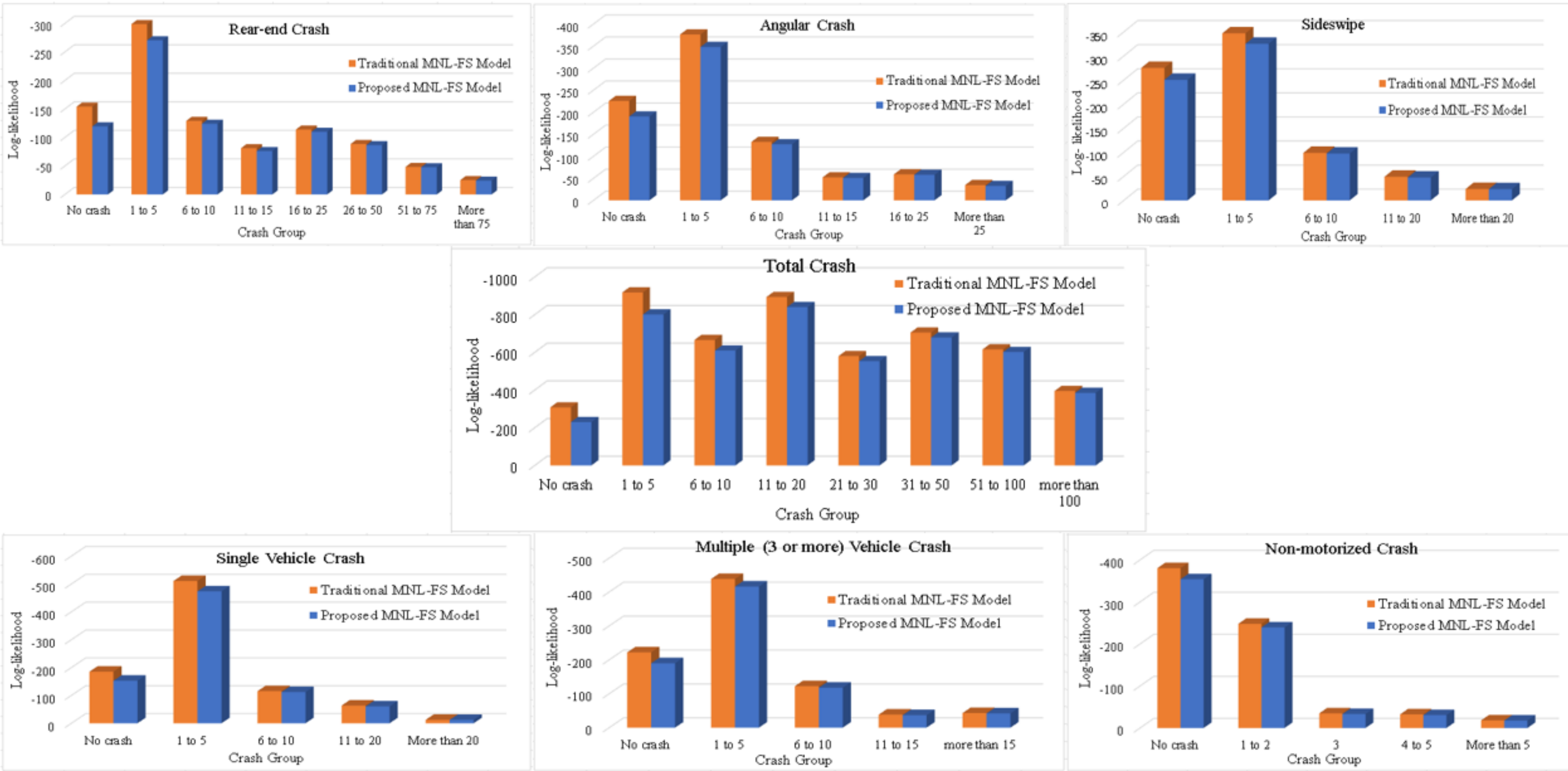


**Figure 2 Log-likelihood Values of Proposed MNL-FS Model Considering Zero Crash Regions and Its Counterpart across Different Crash Groups**

From the figure, we can observe that the proposed model outperforms Model 1 across all model systems including the Total crash and crash type frequency models. A closer examination reveals a higher improvement in log-likelihood for low crash categories. The differences across the model systems are smaller in favor of our proposed model for higher crash categories. Overall, the results indicate that accommodating for zero state in crash type frequency considerations significantly improves model prediction accuracy for low crash categories while not deteriorating performance in the higher crash frequency categories across all crash types. This finding is quite encouraging and supports the consideration of the proposed framework for future crash frequency analysis.

## CONCLUSION

Earlier research has established that fractional split model systems perform as well (if not better) than the traditional multivariate approaches with a parsimonious specification. However, it is possible that across spatial units, several crash configurations (crash type/ crash severity) might have large share of zero crashes. In the traditional multivariate context, the presence of such zeros was accommodated by employing different zero-inflated or hurdle variants. The current research effort furthers safety methodology by improving the fractional split based multivariate model systems with an explicit consideration for the potential presence of zeros by crash configuration. The modified framework with an additional component for zero crashes results in three components: (1) total crash NB component, (2) binary logit (BL) component by crash configuration and (3) multinomial logit fractional split (MNLFS) crash proportion component. The binary component can be employed to identify safer (or riskier) zones by crash configuration. The MNLFS model was developed to determine the probability of a particular crash type only if it was present in the TAZ. The framework also accommodates for unobserved heterogeneity across the three components of the model system. The study contributes to empirical literature on safety by estimating the proposed model system for 6 crash types including rear-end, angle, sideswipe, single-vehicle, multi-vehicle (3 or more), and non-motorized crashes. The model estimation is conducted using a host of independent variables including roadway-, traffic-, land use, built environment, and socio-demographic characteristics, which will contribute to evaluate their impact on crash frequency for different collision types.

The data for our analysis was drawn from the 4,747 TAZs in Central Florida region for the year 2016. The dataset contains the records for 114,458 motorized and 3,413 non-motorized crashes. All the crash records were assigned to the corresponding TAZ using the Geographic Information System (GIS). The proposed model system with a binary logit model for considering zero crash state (Model 2) was compared with the traditional MNL-FS model (Model 1) based on log-likelihood (LL) and Bayesian Information Criterion (BIC) values. The LL (BIC) values are as follows: (a) Model 1: -20,455.2 (41,355.8) and (b) Model 2: -19,292.2 (39,029.6). From these values, it is clear that the newly proposed Model 2 provided the best fit. To illustrate the improved performance of our proposed model, we conducted a prediction exercise by total crash count and crash counts by crash type using validation dataset. The results indicate that accommodating for zero state in crash type frequency considerations significantly improves model prediction accuracy for low crash categories while not deteriorating performance in the higher crash frequency categories across all crash types. The reader would note that the degrees of freedom of the proposed model and the traditional model were the same in this study. The binary logit component development was not included in the joint NB-MNLFS framework. They were developed for future prediction of the occurrence of various crash types in each zone. It would be worthwhile to explore a joint model system in future research efforts.

## AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: Naveen Eluru, Tanmoy Bhowmik; data collection: Tanmoy Bhowmik; analysis and interpretation of results: Md Istiak Jahan, Tanmoy Bhowmik, Naveen Eluru; draft manuscript preparation: Md Istiak Jahan, Naveen Eluru, Tanmoy Bhowmik. All authors reviewed the results and approved the final version of the manuscript.

## REFERENCES

1. NHTSA. NHTSA's 2021 Estimate of Traffic Deaths Shows 16-Year High. https://www.nhtsa.gov/press-releases/early-estimate-2021-traffic-fatalities. Accessed Jun. 21, 2022.
2. Li, Z., X. Chen, Y. Ci, C. Chen, and G. Zhang. A Hierarchical Bayesian Spatiotemporal Random Parameters Approach for Alcohol/Drug Impaired-Driving Crash Frequency Analysis. *Analytic Methods in Accident Research*, Vol. 21, 2019, pp. 44–61. https://doi.org/10.1016/J.AMAR.2019.01.002.
3. Han, C., H. Huang, J. Lee, and J. Wang. Investigating Varying Effect of Road-Level Factors on Crash Frequency across Regions: A Bayesian Hierarchical Random Parameter Modeling Approach. *Analytic Methods in Accident Research*, Vol. 20, 2018, pp. 81–91. https://doi.org/10.1016/J.AMAR.2018.10.002.
4. Kamrani, M., R. Arvin, and A. J. Khattak. Extracting Useful Information from Basic Safety Message Data: An Empirical Study of Driving Volatility Measures and Crash Frequency at Intersections: *https://doi.org/10.1177/0361198118773869*, Vol. 2672, No. 38, 2018, pp. 290–301. https://doi.org/10.1177/0361198118773869.
5. Pu, Z., A. M. Asce, Z. Li, R. Ke, X. Hua, Y. Wang, and F. Asce. Evaluating the Nonlinear Correlation between Vertical Curve Features and Crash Frequency on Highways Using Random Forests. *Journal of Transportation En- gineering, Part A: Systemsgineering, Part A: Systems*, Vol. 146, No. 10, 2020, pp. 04020115-(1-13). https://doi.org/10.1061/JTEPBS.0000410.
6. Zhang, X., S. T. Waller, and P. Jiang. An Ensemble Machine Learning-Based Modeling Framework for Analysis of Traffic Crash Frequency. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 35, No. 3, 2020, pp. 258–276. https://doi.org/10.1111/MICE.12485.
7. Rashidi, M. H., S. Keshavarz, P. Pazari, N. Safahieh, and A. Samimi. Modeling the Accuracy of Traffic Crash Prediction Models. *IATSS Research*, 2022. https://doi.org/10.1016/J.IATSSR.2022.03.004.
8. Almasi, S. A., and H. R. Behnood. Exposure Based Geographic Analysis Mode for Estimating the Expected Pedestrian Crash Frequency in Urban Traffic Zones; Case Study of Tehran. *Accident Analysis & Prevention*, Vol. 168, 2022, p. 106576. https://doi.org/10.1016/J.AAP.2022.106576.
9. Alrejjal, A., M. Moomen, and K. Ksaibati. Evaluating the Effectiveness of Law Enforcement in Reducing Truck Crashes for a Rural Mountainous Freeway in Wyoming. *The International Journal of Transportation Research*, 2021, pp. 1–11. https://doi.org/10.1080/19427867.2021.1949089.

10. Azimian, A., and J. Jiao. Modeling Factors Contributing to Dockless E-Scooter Injury Accidents in Austin, Texas. *Traffic Injury Prevention*, Vol. 23, No. 2, 2022, pp. 107–111. https://doi.org/10.1080/15389588.2022.2030057/SUPPL_FILE/GCPI_A_2030057_SM3695.DOCX.
11. Khedher, M. B. Ben, and D. Yun. Generalized Linear Models to Identify the Impact of Road Geometric Design Features on Crash Frequency in Rural Roads. *KSCE Journal of Civil Engineering*, Vol. 26, No. 3, 2022, pp. 1388–1395. https://doi.org/10.1007/s12205-021-1317-x.
12. Raihan, A., P. Alluri, W. Wu, and A. Gan. Estimation of Bicycle Crash Modification Factors (CMFs) on Urban Facilities Using Zero Inflated Negative Binomial Models. 2018. https://doi.org/10.1016/j.aap.2018.12.009.
13. Bhowmik, T., S. Yasmin, and N. Eluru. Do We Need Multivariate Modeling Approaches to Model Crash Frequency by Crash Types? A Panel Mixed Approach to Modeling Crash Frequency by Crash Types. *Analytic Methods in Accident Research*, Vol. 24, 2019, p. 100107. https://doi.org/10.1016/j.amar.2019.100107.
14. Mannering, F. L., V. Shankar, and C. R. Bhat. Unobserved Heterogeneity and the Statistical Analysis of Highway Accident Data. *Analytic Methods in Accident Research*, Vol. 11, 2016, pp. 1–16. https://doi.org/10.1016/j.amar.2016.04.001.
15. Wang, K., T. Bhowmik, S. Zhao, N. Eluru, and E. Jackson. Highway Safety Assessment and Improvement through Crash Prediction by Injury Severity and Vehicle Damage Using Multivariate Poisson-Lognormal Model and Joint Negative Binomial-Generalized Ordered Probit Fractional Split Model. *Journal of Safety Research*, Vol. 76, 2021, pp. 44–55. https://doi.org/10.1016/J.JSR.2020.11.005.
16. Ding, H., and N. N. Sze. Effects of Road Network Characteristics on Bicycle Safety: A Multivariate Poisson-Lognormal Model. *Multimodal Transportation*, Vol. 1, No. 2, 2022, p. 100020. https://doi.org/10.1016/J.MULTRA.2022.100020.
17. Bhowmik, T., M. Rahman, S. Yasmin, and N. Eluru. Exploring Analytical, Simulation-Based, and Hybrid Model Structures for Multivariate Crash Frequency Modeling. *Analytic Methods in Accident Research*, Vol. 31, 2021, p. 100167. https://doi.org/10.1016/j.amar.2021.100167.
18. Yasmin, S., S. U. Momtaz, T. Nashad, and N. Eluru. A Multivariate Copula-Based Macro-Level Crash Count Model: *https://doi.org/10.1177/0361198118801348*, Vol. 2672, No. 30, 2018, pp. 64–75. https://doi.org/10.1177/0361198118801348.
19. Rahman Shaon, M. R., X. Qin, A. P. Afghari, S. Washington, and M. M. Haque. Incorporating Behavioral Variables into Crash Count Prediction by Severity: A Multivariate Multiple Risk Source Approach. *Accident Analysis & Prevention*, Vol. 129, 2019, pp. 277–288. https://doi.org/10.1016/J.AAP.2019.05.010.
20. Alarifi, S. A., M. Abdel-Aty, and J. Lee. A Bayesian Multivariate Hierarchical Spatial Joint Model for Predicting Crash Counts by Crash Type at Intersections and Segments along Corridors. *Accident Analysis & Prevention*, Vol. 119, 2018, pp. 263–273. https://doi.org/10.1016/J.AAP.2018.07.026.
21. Yasmin, S., N. Eluru, J. Lee, and M. Abdel-Aty. Ordered Fractional Split Approach for Aggregate Injury Severity Modeling. *Transportation Research Record: Journal of the Transportation Research*, Vol. 2583, 2016, pp. 119–126. https://doi.org/10.3141/2583-15.
22. Yasmin, S., T. Bhowmik, M. Rahman, and N. Eluru. Enhancing Non-Motorist Safety by Simulating Trip Exposure Using a Transportation Planning Approach. *Accident; analysis*

*and prevention*, Vol. 156, No. April, 2021, p. 106128. https://doi.org/10.1016/j.aap.2021.106128.
23. Bhowmik, T., S. Yasmin, and N. Eluru. A Multilevel Generalized Ordered Probit Fractional Split Model for Analyzing Vehicle Speed. *Analytic Methods in Accident Research*, Vol. 21, 2019, pp. 13–31. https://doi.org/10.1016/J.AMAR.2018.12.001.
24. Yasmin, S., and N. Eluru. A Joint Econometric Framework for Modeling Crash Counts by Severity. *Transportmetrica A: Transport Science*, Vol. 14, No. 3, 2018, pp. 230–255. https://doi.org/10.1080/23249935.2017.1369469.
25. Lee, J., S. Yasmin, N. Eluru, M. Abdel-Aty, and Q. Cai. Analysis of Crash Proportion by Vehicle Type at Traffic Analysis Zone Level: A Mixed Fractional Split Multinomial Logit Modeling Approach with Spatial Effects. *Accident Analysis & Prevention*, Vol. 111, 2018, pp. 12–22. https://doi.org/10.1016/J.AAP.2017.11.017.
26. Bhowmik, T., S. Yasmin, and N. Eluru. A Joint Econometric Approach for Modeling Crash Counts by Collision Type. *Analytic Methods in Accident Research*, Vol. 19, 2018, pp. 16–32. https://doi.org/10.1016/J.AMAR.2018.06.001.
27. Dey, B. K., S. D. Tirtha, N. Eluru, and K. C. Konduri. Transformation of Ridehailing in New York City: A Quantitative Assessment. *Transportation Research Part C: Emerging Technologies*, Vol. 129, 2021, p. 103235. https://doi.org/10.1016/J.TRC.2021.103235.
28. Briz-Redón, Á., J. Mateu, and F. Montes. Modeling Accident Risk at the Road Level through Zero-Inflated Negative Binomial Models: A Case Study of Multiple Road Networks. *Spatial Statistics*, Vol. 43, 2021, p. 100503. https://doi.org/10.1016/J.SPASTA.2021.100503.
29. Chai, T., D.-Q. Xiong, and J. Weng. A Zero-Inflated Negative Binomial Regression Model to Evaluate Ship Sinking Accident Mortalities. *Article Transportation Research Record*, Vol. 2672, No. 11, 2018, pp. 65–72. https://doi.org/10.1177/0361198118787388.
30. Mathew, J., and R. F. Benekohal. Highway-Rail Grade Crossings Accident Prediction Using Zero Inflated Negative Binomial and Empirical Bayes Method. *Journal of Safety Research*, Vol. 79, 2021, pp. 211–236. https://doi.org/10.1016/J.JSR.2021.09.003.
31. Raihan, M. A., P. Alluri, W. Wu, and A. Gan. Estimation of Bicycle Crash Modification Factors (CMFs) on Urban Facilities Using Zero Inflated Negative Binomial Models. *Accident Analysis and Prevention*, Vol. 123, 2019, pp. 303–313. https://doi.org/10.1016/j.aap.2018.12.009.
32. Sun, M., X. Sun, M. A. Rahman, M. Akter, and S. Das. Modeling Two-Way Stop-Controlled Intersection Crashes with Zero-Inflated Models on Louisiana Rural Two-Lane Highways. *IATSS Research*, Vol. 45, No. 3, 2021, pp. 303–309. https://doi.org/10.1016/J.IATSSR.2020.12.007.
33. Gu, X., X. Yan, L. Ma, and X. Liu. Modeling the Service-Route-Based Crash Frequency by a Spatiotemporal-Random-Effect Zero-Inflated Negative Binomial Model: An Empirical Analysis for Bus-Involved Crashes. *Accident Analysis & Prevention*, Vol. 144, 2020, p. 105674. https://doi.org/10.1016/J.AAP.2020.105674.
34. Champahom, T., S. Jomnonkwao, A. Karoonsoontawong, and V. Ratanavaraha. Spatial Zero-Inflated Negative Binomial Regression Models: Application for Estimating Frequencies of Rear-End Crashes on Thai Highways. *Journal of Transportation Safety & Security*, Vol. 14, No. 3, 2020, pp. 523–540. https://doi.org/10.1080/19439962.2020.1812786.
35. Fountas, G., and P. C. Anastasopoulos. Analysis of Accident Injury-Severity Outcomes:

The Zero-Inflated Hierarchical Ordered Probit Model with Correlated Disturbances. *Analytic Methods in Accident Research*, Vol. 20, 2018, pp. 30–45. https://doi.org/10.1016/J.AMAR.2018.09.002.

36. Gamaleldin, G., H. Al-Deek, A. Sandt, J. McCombs, A. El-Urfali, and N. Uddin. Developing Context-Specific Safety Performance Functions for Florida Intersections to More Accurately Predict Intersection Crashes. *https://doi.org/10.1080/19439962.2020.1796865*, Vol. 14, No. 4, 2020, pp. 607–629. https://doi.org/10.1080/19439962.2020.1796865.
37. Hong, J., R. Tamakloe, and D. Park. A Comprehensive Analysis of Multi-Vehicle Crashes on Expressways: A Double Hurdle Approach. *Sustainability 2019, Vol. 11, Page 2782*, Vol. 11, No. 10, 2019, p. 2782. https://doi.org/10.3390/SU11102782.
38. Qiu, J., D. B. Logan, J. Oxley, and C. Lowe. TRR Application of a Hurdle Model with Random Effects to Explore the Relationship between Operational Characteristics and Safety Performance. *Transportation Research Record*, Vol. 2020, No. 8, 2020, pp. 327–337. https://doi.org/10.1177/0361198120928074.
39. Li, J. Q., W. Liu, X. Yang, P. Lu, and K. C. P. Wang. Statistical Safety Performance Models Considering Pavement and Roadway Characteristics. *Journal of Advanced Transportation*, Vol. 2022, 2022, pp. 1–12. https://doi.org/10.1155/2022/5871601.
40. Bhat, C. R. Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model. *Transportation Research Part B: Methodological*, Vol. 35, No. 7, 2001, pp. 677–693. https://doi.org/10.1016/S0191-2615(00)00014-X.
41. Eluru, N., C. R. Bhat, and D. A. Hensher. A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes. *Accident Analysis and Prevention*, Vol. 40, No. 3, 2008, pp. 1033–1054. https://doi.org/10.1016/J.AAP.2007.11.010.
42. Yasmin, S., and N. Eluru. Evaluating Alternate Discrete Outcome Frameworks for Modeling Crash Injury Severity. *Accident Analysis and Prevention*, Vol. 59, 2013, pp. 506–521. https://doi.org/10.1016/J.AAP.2013.06.040.
43. Aptech. Aptech Systems, Inc.
44. Mahmud, A., and V. V. Gayah. Estimation of Crash Type Frequencies on Individual Collector Roadway Segments. *Accident Analysis & Prevention*, Vol. 161, 2021, p. 106345. https://doi.org/10.1016/J.AAP.2021.106345.
45. Regev, S., J. J. Rolison, and S. Moutari. Crash Risk by Driver Age, Gender, and Time of Day Using a New Exposure Methodology. *Journal of Safety Research*, Vol. 66, 2018, pp. 131–140. https://doi.org/10.1016/J.JSR.2018.07.002.
46. Lao, Y., G. Zhang, Y. Wang, and J. Milton. Generalized Nonlinear Models for Rear-End Crash Risk Analysis. *Accident Analysis & Prevention*, Vol. 62, 2014, pp. 9–16. https://doi.org/10.1016/J.AAP.2013.09.004.
47. Lee, J., M. Abdel-Aty, and X. Jiang. Multivariate Crash Modeling for Motor Vehicle and Non-Motorized Modes at the Macroscopic Level. *Accident Analysis & Prevention*, Vol. 78, 2015, pp. 146–154. https://doi.org/10.1016/J.AAP.2015.03.003.