An Integrated Multi-Resolution Framework for Jointly Estimating Crash Type and Crash Severity

Shahrior Pervaz*
Graduate Research Assistant
Department of Civil, Environmental and Construction Engineering
University of Central Florida
Tel: 1-407-561-0298; Fax: 1-407-823-3315
Email: shahrior.pervaz@ucf.edu
ORCiD number: 0000-0001-7966-7083

Tanmoy Bhowmik
Assistant Professor
Department of Civil, and Environmental Engineering
Portland State University
Tel: 1-407-927-6574
Email: tbhowmik@pdx.edu
ORCiD number: 0000-0002-0258-1692

Naveen Eluru
Professor
Department of Civil, Environmental and Construction Engineering
University of Central Florida
Tel: 407-823-4815, Fax: 407-823-3315
Email: naveen.eluru@ucf.edu
ORCiD number: 0000-0003-1221-4113

*Corresponding author

ABSTRACT
The current research effort contributes to safety literature by developing an integrated framework that allows for the influence of independent variables from crash type and severity components at the disaggregate level to be incorporated within the aggregate level propensity to estimate crash frequency by crash type and severity. The empirical analysis is based on the crash data drawn from the city of Orlando, Florida for the year 2019. The disaggregate level analysis uses 15,518 crash records of three crash types including rear end, angular and sideswipe. Each crash record contains crash specific factors, driver and vehicle factors, roadway attributes, road environmental and weather information. For aggregate level model analysis, the study aggregates the crash records by crash type over 300 traffic analysis zones. An exhaustive set of independent variables including roadway and traffic characteristics, land-use attributes, built environment and sociodemographic factors are considered in this level. The empirical analysis is further augmented by employing several goodness of fit and predictive measures. A validation exercise is also conducted using a holdout sample to highlight the superiority of the proposed integrated model relative to the non-integrated model system. The proposed framework can also incorporate unobserved heterogeneity in the model system. The findings of the study indicate that the proposed framework is advantageous for capturing the variable effects simultaneously across the aggregate and disaggregate levels.

Keywords: Crash type, Crash severity, Aggregate and disaggregate level analysis, Integrated framework, Unobserved effects.

1. MOTIVATION

In transportation safety literature, the application of statistical and econometric models is an important tool for identifying factors influencing crash occurrence/consequences and devising appropriate countermeasures. Crash frequency models are traditionally applied to study crash occurrence at a facility resolution (such as segment, intersection or traffic analysis zone) while discrete outcome models are employed for crash severity analysis (such as driver injury severity). The econometric model frameworks developed in safety literature have significantly been enhanced in recent years (see Lord and Mannering, 2010; Mannering et al., 2016; Savolainen et al., 2011; Yasmin and Eluru, 2013 for a review). The major advances in literature can be broadly classified along two directions. The ***first direction*** of advances is focused on accommodating the influence of unobserved factors in crash modeling. Several studies following the seminal paper by Mannering et al. (2016) have conducted research along these lines including models with random parameters, heterogeneity in variances, endogeneity models, and multivariate models (Ahmed et al., 2023; Balusu et al., 2018; Bhowmik et al., 2022, 2021a; Yasmin et al., 2018). The different approaches employed for crash frequency analysis include random effect/parameter negative binomial models (Gong et al., 2020; Yan et al., 2020), latent segmentation based-negative binomial models (Bhowmik et al., 2022; Yasmin and Eluru, 2016), random parameter multivariate tobit models (Anastasopoulos, 2016), random parameter zero-inflated/hurdle models (Gu et al., 2020; Yu et al., 2019), negative binomial-fractional split models (Yasmin and Eluru, 2018), panel mixed negative binomial models (Bhowmik et al., 2022, 2019), copula based-crash frequency models (Yasmin et al., 2018), Bayesian Poisson-lognormal multilevel/hierarchical models (Alarifi et al., 2018; Cui and Xie, 2021), geographically weighted regression models (Huang et al., 2018), random parameter negative binomial-Lindley models (Islam et al., 2023a), and integrated crash frequency models (Cai et al., 2019; Pervaz et al., 2022). For crash severity models, advances in this realm include random parameter/mixed logit models (Hou et al., 2022; Islam et al., 2023b), generalized ordered logit models (Marcoux et al., 2018), scaled generalized ordered logit models (Kabli et al., 2023; Marcoux et al., 2018), random parameter latent class clustering and latent segmentation based ordered logit models (Chang et al., 2021; Xiong and Mannering, 2013).

The ***second direction*** of research is geared towards accommodating systematic factors by recognizing different observed attributes typically not considered in the analysis. Studies in this realm building on Mannering (2018) include efforts to carefully incorporate the influence of temporal instability by explicitly recognizing model parameter variation across the time periods (such as years) (Mannering, 2018). Within this realm, another group of studies focus on developing parsimonious models by employing pooled model frameworks. For example, Bhowmik et al. (2019) proposed a pooled univariate crash frequency model (accommodating unobserved heterogeneity) by recasting the multivariate crash frequency modeling approach as repeated measures of crash frequency while recognizing that each repetition represents a different crash type (Bhowmik et al., 2022, 2021b, 2019). The approach simplifies observed parameter inclusion in the models while accommodating for deviations across multiple sections of the data through interaction effects. The heterogeneity in means approach also allows for interactions among observed variables to allow for additional observed impacts to be recognized (Alnawmasi and Mannering, 2022; Huo et al., 2020; Mannering et al., 2016).

To be sure, several advanced frameworks contribute to both directions including (a) latent class models, (b) heterogeneity in means and variances, and (c) integrated multi-resolution approaches. Safety models employing latent class models (or finite mixture models) incorporate the impact of observed and unobserved variables in partitioning the population into different classes with class specific variable impacts. Latent class modeling approaches have been employed for crash frequency analysis (see Bhowmik et al., 2022; Yasmin and Eluru, 2016) and crash severity analysis (see Chang et al., 2021; Xiong and Mannering, 2013; Yasmin et al., 2014). The commonly adopted heterogeneity in means and variances incorporates additional observed heterogeneity through interaction variables and unobserved heterogeneity by allowing unobserved heterogeneity to vary across the data as a function of independent variables. In recent years several studies employed these approaches for crash frequency (see Alnawmasi and Mannering, 2022; Huo et al., 2020) and severity analysis (see Ahmed et al., 2023; Alnawmasi and Mannering, 2022; Xin et al., 2017). While this approach has been very well applied as documented by the burgeoning studies in safety literature, the approach is applied with traditional safety data. Pervaz et al. (2023), a study from the integrated multi-resolution approach, proposed a unified model system to accommodate for observed and unobserved variables by considering safety data from a more nuanced and conceptually enhanced framework (Pervaz et al., 2023). The study developed an integrated framework that recognizes that crash frequency data are generated by aggregating individual crash records. However, the information in these individual crash records is often ignored in developing crash frequency models. In their study, Pervaz et al. (2023) enhanced the observed attributes considered in the model framework by incorporating the influence of independent variables at the crash record level from a crash severity model within the aggregate level crash frequency propensity component. Further, the study also allowed for the influence of unobserved factors across the aggregated and crash record level model components. Thus, the study illustrated how additional observed information present in crash records can be included in an aggregate model framework that traditionally has ignored these crash record data while also controlling for unobserved effects.

2. STUDY CONTEXT

The current study builds on this integrated multi-resolution approach by accommodating for additional information from the crash record data. In Pervaz et al. (2023), crash severity model information was incorporated in the crash frequency models (Pervaz et al., 2023). While this is a significant contribution, it might be beneficial to incorporate the factors contributing to different crash types along with crash severity within the aggregate framework to augment the observed information flow from crash record level data. The approach would involve summing up the crash propensity of each disaggregate level crash record by crash type and injury severity within the aggregate resolution and adding the generated values as new variables in the aggregate level model. The process significantly increases the dimensionality of the terms being considered in the aggregate and disaggregate model components. To summarize, in our current paper, a unified framework that explicitly allows for the information flow of observed and unobserved variables from the crash type and crash severity model components into the aggregate level crash frequency model is proposed and estimated.

In our study, a panel mixed negative binomial-ordered probit fractional split (NB-OPFS) framework is employed at aggregate level to jointly estimate crash frequency by crash type and severity. Specifically, the NB component models the number of crashes by type and the OPFS component determines the proportion of each severity in the pooled dataset for a zone. At disaggregate level, the crash type variable is examined using multinomial logit (MNL) model and crash severity variable is examined using a pooled ordered probit (OP) model. The integrated approach can take two potential forms with these models. In the first structure, the MNL model propensity and the pooled OP model propensity across the crashes in the zone are computed as composite scores and treated as exogenous variables i.e., the crash type and severity model parameters are fixed. In this approach, two additional parameters for the composite variables are estimated for each crash type in the panel mixed NB-OPFS model i.e., the composite score of MNL is included in the count and score for OP is included in the severity proportion components. Alternatively, composite scores can be treated as endogenous and be estimated simultaneously within the panel mixed NB-OPFS model. In this approach, the estimates of the disaggregate models will be allowed to vary while modeling crash frequency. The second approach is computationally more involved as it allows for feedback between aggregate and disaggregate level models. The model selection process can be accomplished using model fit measures such as Bayesian Information Criterion (BIC).

The proposed model system is estimated using data drawn from the City of Orlando, Florida for the year 2019. The study considers three crash types: rear end, angular, and sideswipe for the analysis. These three crash types comprise 15,518 crash records for the disaggregate level model analysis. The records contain crash specific factors, driver and vehicle factors, roadway attributes, road environmental and weather information of each crash record. For the aggregate level analysis, these crash records are aggregated over 300 traffic analysis zones (TAZs). An exhaustive set of independent variables including roadway and traffic characteristics, land-use attributes, built environment factors, and sociodemographic characteristics are considered in this level. The results of the empirical analysis further bolster the importance of developing such an integrated framework for aggregate level crash frequency and disaggregate level discrete crash outcome analysis.

3. METHODOLOGY

In this study, we employ a panel mixed NB-OPFS model and an integrated modeling framework to jointly estimate crash frequency by crash type and severity. However, for the sake of space, we will restrict ourselves in presenting the integrated framework only. Further, within the integrated framework, there are two components: the disaggregate level models (MNL and pooled OP) and the aggregate level model (panel mixed NB-OPFS). For the ease of presentation, we will discuss the methodology by each component.

3.1 Disaggregate Level Model Structures (MNL and Pooled OP Models)

3.1.1 Multinomial Logit (MNL) Model

Let us consider the probability of a crash record $j (j = 1,2,3, \dots, n)$ ending in a specific crash type $l \ (l = 1,2, \dots, L)$. In this study we consider three crash types: rear end, angular, and sideswipe. The alternative specific latent variables for MNL take the form of:

$$q_{jl} = \boldsymbol{\Psi_l} \mathrm{Y}_{\boldsymbol{jl}} + \mathrm{e}_{jl} \tag{1}$$

where q_{jl} is a function of covariates determining the crash type, $\boldsymbol{\Psi_l}$ is a vector of coefficients to be estimated for outcome of $\boldsymbol{j}$, $\mathrm{Y}_{\boldsymbol{jl}}$ is a vector of exogenous variables, and e_{jl} is the random component assumed to follow a gumbel type 1 distribution. Thus, the MNL probability expression is as follows:

$$P_j(l) = \frac{exp\ [\boldsymbol{\Psi_l} \mathrm{Y}_{\boldsymbol{jl}}]}{\sum_{l=1}^{L} exp\ [\boldsymbol{\Psi_l} \mathrm{Y}_{\boldsymbol{jl}}]} \tag{2}$$

Considering the spatial arrangement of the crash records within the same zone, i.e., the adjacency heterogeneity (dependency), the equation for MNL model propensity can be updated as,

$$q'_{jl} = \boldsymbol{\Psi_l}\mathrm{Y}_{\boldsymbol{jl}} + \boldsymbol{\theta}_{il} + \mathrm{e}_{jl} \tag{3}$$

where, i ($i = 1,2,3, \dots, N$) is the index for traffic analysis zone (TAZ). q'_{jl} is the latent propensity capturing spatial dependency and $\boldsymbol{\theta}_{il}$ is a vector of unobserved effects specific to the zone for the crash records of type l, highlighting the spatial arrangement within the same zone. This $\boldsymbol{\theta}_{il}$ will be same across the crash records of type l if they correspond to same zone (TAZ) and thus the adjacency heterogeneity (dependency) will be captured through the proposed system. The reader would note that, the spatial unobserved heterogeneity can vary across the crash records. Therefore, in the current study, we parameterize the correlation parameter $\boldsymbol{\theta}_{il}$ as a function of observed attributes as follows:

$$\boldsymbol{\theta}_{il} = \boldsymbol{\gamma}_{il}\boldsymbol{s}_{il} \tag{4}$$

where, $\boldsymbol{s}_{il}$ is a vector of exogenous variables at the zonal level i (including a constant) employed for crash records of type l, $\boldsymbol{\gamma}_{il}$ is a vector of parameters to be estimated. Therefore, the updated probability function will be as follows:

$$P'_j(l) = \frac{exp\ [\boldsymbol{\Psi_l}\mathrm{Y}_{\boldsymbol{jl}} + \ \boldsymbol{\theta}_{il}]}{\sum_{l=1}^{L} exp\ [\boldsymbol{\Psi_l}\mathrm{Y}_{\boldsymbol{jl}} + \ \boldsymbol{\theta}_{il}]} \tag{5}$$

3.1.2 Pooled Ordered Probit (OP) Model

In the traditional ordered outcome model, the discrete injury severity levels (v_{jl}) of crash record j and crash type l are assumed to be associated with an underlying continuous latent variable (v_{jl}^*). For the pooled dataset, this latent variable can be specified as the following linear function:

$$v_{jl}^* = X_{jl}\Theta + \varepsilon_{jl} \tag{6}$$

where X_{jl} is a vector of exogenous variables (excluding a constant), Θ is a vector of unknown parameters to be estimated, and ε_{jl} is the random disturbance term assumed to be standard normal distribution. Let us assume k ($k = 1,2,3, \dots, K$) be the index to represent injury severity categories. In this study, k take the values of 'no-injury' ($k = 1$), 'possible injury' ($k = 2$), 'non-incapacitating injury' ($k = 3$) and 'fatal and incapacitating injury' ($k = 4$). The unobservable latent variable v_{jl}^* is related to the observable ordinal variable v_{jl} by the t_{lk} with an outcome mechanism of the following form:

$$v_{jl} = lk,\ if\ t_{l,k-1} < \ v_{jl}^* < t_{lk} \tag{7}$$

where t_{lk} represents the thresholds associated with the severity levels for crash type l. These unknown t_{lk}s are assumed to partition the propensity into $lk - 1$ intervals. In order to ensure the well-defined intervals and natural ordering of observed severity, the thresholds are assumed to be ascending in order, such that $t_{l0} < t_{l1} < \ \dots \dots \dots < t_{lK}$ where $t_{l0} = -\infty$ and $t_{lK} = +\infty$. Given these relationships across the different parameters, the resulting probability expressions π_{jlk}, for record j and alternative k for the ordered probit (OP) take the following form:

$$\pi_{jlk} = Pr\left(v_{jl} = lk \middle| X_{jl}\right) = \Upsilon\left(t_{lk} - X_{jl}\Theta\right) - \Upsilon\left(t_{l,k-1} - X_{jl}\Theta\right) \tag{8}$$

where $\Upsilon(.)$ represents the standard normal distribution function.

Considering the spatial arrangement of the crash records within the same zone, i.e., the adjacency heterogeneity (dependency), the equation for disaggregate level pooled OP model propensity can be updated as,

$$v_{jl}^{*\prime} = X_{jl}\Theta + \boldsymbol{\theta}_{il} + \varepsilon_{jl} \tag{9}$$

where, $v_{jl}^{*\prime}$ is the latent propensity capturing spatial dependency. With the updated propensity, the probability expression is:

$$\pi'_{jlk} = Pr\left(v_{jl}^{*\prime} = lk \middle| X_{jl}\right) = \Upsilon\left(t_{lk} - (X_{jl}\Theta + \boldsymbol{\theta}_{il})\right) - \Upsilon\left(t_{l,k-1} - (X_{jl}\Theta + \boldsymbol{\theta}_{il})\right) \tag{10}$$

3.2 Aggregate Level Model Structure (Panel Mixed NB-OPFS Model)

3.2.1 Count Framework

In our study, the count framework estimates number of crashes by type using a panel mixed univariate negative binomial (NB) modeling framework. We arrange the dataset by taking all three crash types as repeated measures (same TAZ is repeated 3 times) of crash frequency in a univariate NB formulation while recognizing that each repetition represents a different crash type.

Once the disaggregate level MNL propensities are estimated, we adopt two alternative approaches to estimate the aggregate (zonal) level propensities for count component. With the TAZ index i and crash type index l, the two approaches are presented in equation 11 and equation 12 respectively.

$$\mu_{il} = E(c_{il}|\boldsymbol{z}_{il}) = exp\left((\boldsymbol{\delta} + \boldsymbol{\zeta}_i + \boldsymbol{\zeta}'_{il})\boldsymbol{z}_{il} + \rho_{cl} * \ln\left(\sum_{p=1}^{j_{il}} \left(exp\left(q'_{jl}\right)\right)\right) + \varepsilon_{il} + \eta_{il}\right) \tag{11}$$

$$\mu_{il} = E(c_{il}|\boldsymbol{z}_{il}) = exp\left((\boldsymbol{\delta} + \boldsymbol{\zeta}_i + \boldsymbol{\zeta}'_{il})\boldsymbol{z}_{il} + \rho_{cl} * \ln\left(\sum_{p=1}^{j_{il}} \left(exp(\boldsymbol{\Psi_l} \mathrm{Y}_{\boldsymbol{jl}} + \boldsymbol{\theta}_{il} + \mathrm{e}_{jl})\right)\right) + \varepsilon_{il} + \eta_{il}\right) \tag{12}$$

where, $\boldsymbol{z}_{il}$ is a vector of explanatory variables associated with TAZ i and crash type l, $\boldsymbol{\delta}$ is a vector of coefficients to be estimated, $\boldsymbol{\zeta}_i$ is a vector of unobserved factors on crash count propensity for TAZ i, $\boldsymbol{\zeta}'_{il}$ is a vector of unobserved factors specific to the crash type l and ρ_{cl} is scalar associated with the disaggregate level highlighting the share of disaggregate level MNL model propensity to be linked with the aggregate level propensity for count component for each crash type. p is a counter here ranging from 1 to j_{il} represents the crash record j of crash type l in zone i. For example, if 5 rear end crashes occur in the zone i, then we will sum the propensity for these 5 crashes to obtain a value for j_{il} for rear end crash type. The reader would note that the proposed framework allows disaggregate level information flow from propensity for a crash type/severity into the aggregate (zonal) level propensity equation if and only if the zone has that crash type/severity. Therefore, if there is no crash record (zero state) in a zone, no disaggregate level information (no score from the crash type/severity propensity) is carried out to the aggregate level and hence only aggregate level variables information prevail in the equation for that zone. It is also important to note that the aggregate propensity is determined by the overall interaction of aggregate variables and disaggregate composite score. ε_{il} is a gamma distributed error term with mean 1 and variance α. η_{il} captures unobserved factors that simultaneously impact number of crashes and proportion of crashes by severity for each crash type in TAZ i. In estimating the model, it is necessary to specify the structure for the unobserved vectors $\boldsymbol{\zeta}_i$ and $\boldsymbol{\zeta}'_{il}$ represented by $\boldsymbol{\zeta}''$. In this paper, it is assumed that these elements are drawn from independent normal distribution: $\boldsymbol{\zeta}''{\sim}N$ (0, $(\boldsymbol{\sigma}_1^2, \boldsymbol{\sigma}_2^2)$).

The main difference between the two approaches is that the disaggregate level propensity will remain fixed and only the scalar parameters will be estimated in approach 1 (exogenous approach). In the second approach (endogenous approach), we allow the disaggregate level parameters to be jointly influenced by disaggregate and aggregate level model fit.

For the count model, the equation for modeling crash count of type l in the usual NB formulation can be written as:

$$P(c_{il}) = \frac{\Gamma\left(c_{il} + \frac{1}{\alpha}\right)}{\Gamma(c_{il}+1)\Gamma\left(\frac{1}{\alpha}\right)} \left(\frac{1}{1+\alpha\mu_{il}}\right)^{\frac{1}{\alpha}} \left(1 - \frac{1}{1+\alpha\mu_{il}}\right)^{c_{il}} \tag{13}$$

where c_{il} be the index for crashes of type l occurring over a period of time in TAZ i. $P(c_{il})$ is the probability that TAZ i has c_{il} number of crashes for crash type l. $\Gamma(\cdot)$ is the gamma function, α is NB overdispersion parameter and μ_{il} is the expected number of crashes for crash type l occurring in TAZ i over a given time period (as presented in the equation 11 and equation 12).

3.2.2 Fractional Split Framework

The modeling of crash proportions by severity levels for crash type l is undertaken using a panel mixed ordered probit fractional split (OPFS) model. In the ordered outcome framework, the actual injury severity proportions (y_{ilk}) are assumed to be associated with an underlying continuous latent variable (y_{il}^{*}). Following the same approach as presented in the count component, we adopt two alternative approaches to estimate latent propensity equation as follows:

$$y_{il}^{*} = \left((\boldsymbol{\beta} + \boldsymbol{\rho_i} + \boldsymbol{\rho'_{il}})\boldsymbol{x}_{il} + \ \rho_{fl} * \ln\left(\sum_{p=1}^{j_{il}} \left(exp\left(v_{jl}^{*\prime}\right)\right)\right) + \xi_{il} \pm \eta_{il}\right), \quad y_{ilk} = lk \ if \ \tau_{l,k-1} < y_{il}^{*} < \tau_{lk} \tag{14}$$

$$y_{il}^{*} = \left((\boldsymbol{\beta} + \boldsymbol{\rho_i} + \boldsymbol{\rho'_{il}})\boldsymbol{x}_{il} + \ \rho_{fl} * \ln\left(\sum_{p=1}^{j_{il}} \left(exp\left(X_{jl}\Theta + \boldsymbol{\theta}_{il} + \varepsilon_{jl}\right)\right)\right) + \xi_{il} \pm \eta_{il}\right), y_{ilk} = lk \ if \ \tau_{l,k-1} < y_{il}^{*} < \tau_{lk} \tag{15}$$

The latent propensity y_{il}^{*} is mapped to the actual severity proportion categories y_{ilk} by τ thresholds $(\tau_{l0} = -\infty \ and \ \tau_{lK} = +\infty)$ as presented in equation 14 and equation 15. $\boldsymbol{x}_{il}$ is a vector of attributes (not including a constant) that influences the propensity associated with severity proportion categories of the three crash types. $\boldsymbol{\beta}$ is the corresponding vector of mean effects, $\boldsymbol{\rho_i}$ is a vector of unobserved factors on severity proportion propensity for TAZ i, $\boldsymbol{\rho'_{il}}$ is a vector of unobserved factors specific to the crash type l. In estimating the model, it is necessary to specify the structure for the unobserved vectors $\boldsymbol{\rho_i}$ and $\boldsymbol{\rho'_{il}}$ represented by $\boldsymbol{\rho''}$. In this paper, it is assumed that these elements are drawn from independent normal distribution: $\boldsymbol{\rho''} \sim N$ (0, $(\boldsymbol{\sigma}_3^2, \boldsymbol{\sigma}_4^2)$). ρ_{fl} is a scalar associated with the disaggregate level highlighting the share of disaggregate level propensity to be linked with the aggregate level propensity for fractional split component for crash type l. ξ_{il} is an idiosyncratic error term assumed to be identically and independently standard normally distributed across TAZ i. η_{il} term generates the correlation between equations for number of crashes and crash proportions by severity levels of crash type l. The $\pm$ sign in front of η_{il} indicates that the correlation in unobserved individual factors between crash counts and crash proportions by severity levels may be positive or negative. A positive sign implies that TAZs with higher number of crashes with type l are intrinsically more likely to incur higher proportions for severe crashes. On the other hand, negative sign implies that TAZs with higher number of crashes of type l intrinsically incur lower proportions for severe crashes. To determine the appropriate sign one can empirically test the models with both $'+'$ and $'-'$ signs independently. The model structure that offers the superior data fit is considered as the final model.

It is important to note here that the unobserved heterogeneity between the number of crashes by type and crash proportions by severity levels can vary across TAZs. Therefore, in the current study, the correlation parameter η_{il} is parameterized as a function of observed attributes as follows:

$$\eta_{il} = \mathrm{G}_{il}\mathrm{Q}_{il} \tag{16}$$

where, Q_{il} is a vector of exogenous variables, G_{il} is a vector of unknown parameters to be estimated (including a constant).

To estimate the model presented in equation 14 and equation 15, we assume that:

$$E(y_{ilk}|\boldsymbol{x}_{il}) = H_{ilk}(\beta, \tau),\ 0 \le H_{ilk} \le 1, \sum_{lk=1}^{LK} H_{ilk} = 1 \tag{17}$$

H_{ilk} in our model takes the ordered probit probability (Λ) form for the crash type l and severity category k.

Given these relationships across different parameters, the resulting probability (Λ) for the panel mixed OPFS model takes the following form:

$$\Lambda(y_{ilk} = lk) = \varphi\{\tau_{lk} - (y_{il}^{*})\} - \varphi\{\tau_{l,k-1} - (y_{il}^{*})\} \tag{18}$$

where, $\varphi(\cdot)$ is the standard normal cumulative distribution function.

3.3 Model Estimation

In examining the model structure of crash count by crash type and severity proportions, we specify the structure for the unobserved vectors $\boldsymbol{\zeta}'', \boldsymbol{\rho}'', \boldsymbol{\gamma}$ and G represented by Ω. In this study, it is assumed that these elements are drawn from independent realization from normal population: $\Omega \sim N(0, (\boldsymbol{\sigma}_1^2, \boldsymbol{\sigma}_2^2, \boldsymbol{\sigma}_3^2, \boldsymbol{\sigma}_4^2, \boldsymbol{\sigma}_5^2, \boldsymbol{\sigma}_6^2))$. Thus, conditional on Ω, the likelihood function for the integrated probability can be expressed as:

$$L_i = \int_{\Omega} \prod_{l=1}^{L} \left[P(c_{il}) \times \prod_{k=1}^{K} \left(\Lambda(y_{ilk} = lk)\right)^{\varpi_{il} d_{ilk}} \times \prod_{p=1}^{j_i} \left(P'_j(l) \times \prod_{k=1}^{K} \pi'_{jlk} \right) d\Omega \right] \tag{19}$$

where, ϖ_{il} is a dummy with $\varpi_{il} = 1$ if TAZ i has at least one crash of type l over the study period and 0 otherwise. d_{ilk} is the proportion of crashes of type l in severity category k. Finally, the log-likelihood function is:

$$LL = \sum_i Ln(L_i) \tag{20}$$

All the parameters in the model are estimated by maximizing the logarithmic function LL presented in equation 20. To estimate the proposed model, 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 (please see Bhat, 2001; Yasmin and Eluru, 2013 for details). In our research, we tested the model specification with several realization levels (such as 50, 100, …, 200). We found that model parameters were stable around 100. We use the GAUSS matrix programming software to run the models (Aptech, 2015).

4. DATA PREPARATION

The data for our analysis is drawn from 2019 Signal Four Analytics (S4A) data for 300 traffic analysis zones (TAZs) in the City of Orlando, Florida. After processing and cleaning the data, a total of 20,204 crash records were obtained in the study area. Among these records, angle crashes, left turn and right turn crashes, rear end and sideswipe crashes comprise a total of 15,518 records.

This study combined angle crashes, left turn and right turn crashes and labelled them as angular crash type. The distribution of the angular (ANG), rear end (RE) and sideswipe (SS) crash types in the dataset is 23.41%, 55.52% and 21.07%, respectively. Each crash record could be further classified into 5 categories by crash severity outcomes such as fatal injury (FI), incapacitating injury (II), non-incapacitating injury (NII), possible injury (PI) and no-injury crashes (NI). This study combines fatal and incapacitating injury crashes as fatal and incapacitating injury (FII) crashes for disaggregate level model estimation. The disaggregate level models consider crash specific variables (such as first harmful event), driver factors (such as driving under influence related, distraction related), vehicle factors (such as presence of passengers), roadway characteristics (such as speed limit, shoulder type), road environmental factors (such as time of the day, light condition) and weather information (such as clear, rain) for the analysis.

For aggregate level model analysis, the study aggregated crash data by crash type over 300 TAZs. In the aggregation, each TAZ is repeated three times for three separate crash types and constitutes a panel data structure. For the count component, the number of crashes by crash type is considered as the dependent variable. For the severity component, four severity levels are considered and the dependent variable for fractional split component (represented as OPFS model) can be represented as proportions (number of specific severity level of type l/total number of all crashes of type l) as follows: (1) proportion of no-injury crashes (2) proportion of possible injury crashes (3) proportion of non-incapacitating injury crashes, and (4) proportion of fatal and incapacitating injury crashes. A comprehensive set of independent variables including roadway and traffic factors, land-use attributes, built environment factors, and sociodemographic characteristics are considered for the analysis of this level. This study selects 255 TAZs randomly for model estimation resulting in a sample of 13,253 crash records. The remaining 45 TAZs with 2,265 crash records are set aside for the validation of the models.

4.1 Variables Considered

The variables for disaggregate and aggregate levels analysis were collected from different data sources including Signal Four Analytics (S4A), Florida Department of Transportation (FDOT) Transportation Statistics Division, US Census Bureau and American Community Survey, and Florida Geographic Data Library databases. These explanatory variables were aggregated at the zonal level using the ArcGIS for aggregate level dataset. Aggregate level analysis uses roadway and traffic characteristics (such as proportion of roads by functional class, average number of lanes, average speed limit, average shoulder width, sidewalk width and median width, intersection density, traffic signal density, AADT, and truck AADT), land-use attributes (such as proportion of residential, commercial, institutional, industrial, recreational and mixed areas), built environment attributes (such as number of restaurants, business centers, commercial centers, educational centers, and shopping centers), and sociodemographic characteristics (such as population density, proportion of males and females, household density, median household income, proportion of car, drive alone, non-motorized means of transport, different population group by age level, household with vehicle availability, and population with different races).

In estimating the models, several functional forms and combination of variables are considered and those variables that provide the best fit are retained in the final specification. The final specifications of the models are based on removing the statistically insignificant variables in a systematic process based on 90% confidence level. Figure 1a and 1b show the sample share of the variables at disaggregate level considered for the final model estimation while the aggregate level variables are presented in Table 1 with the appropriate definitions and summary statistics.

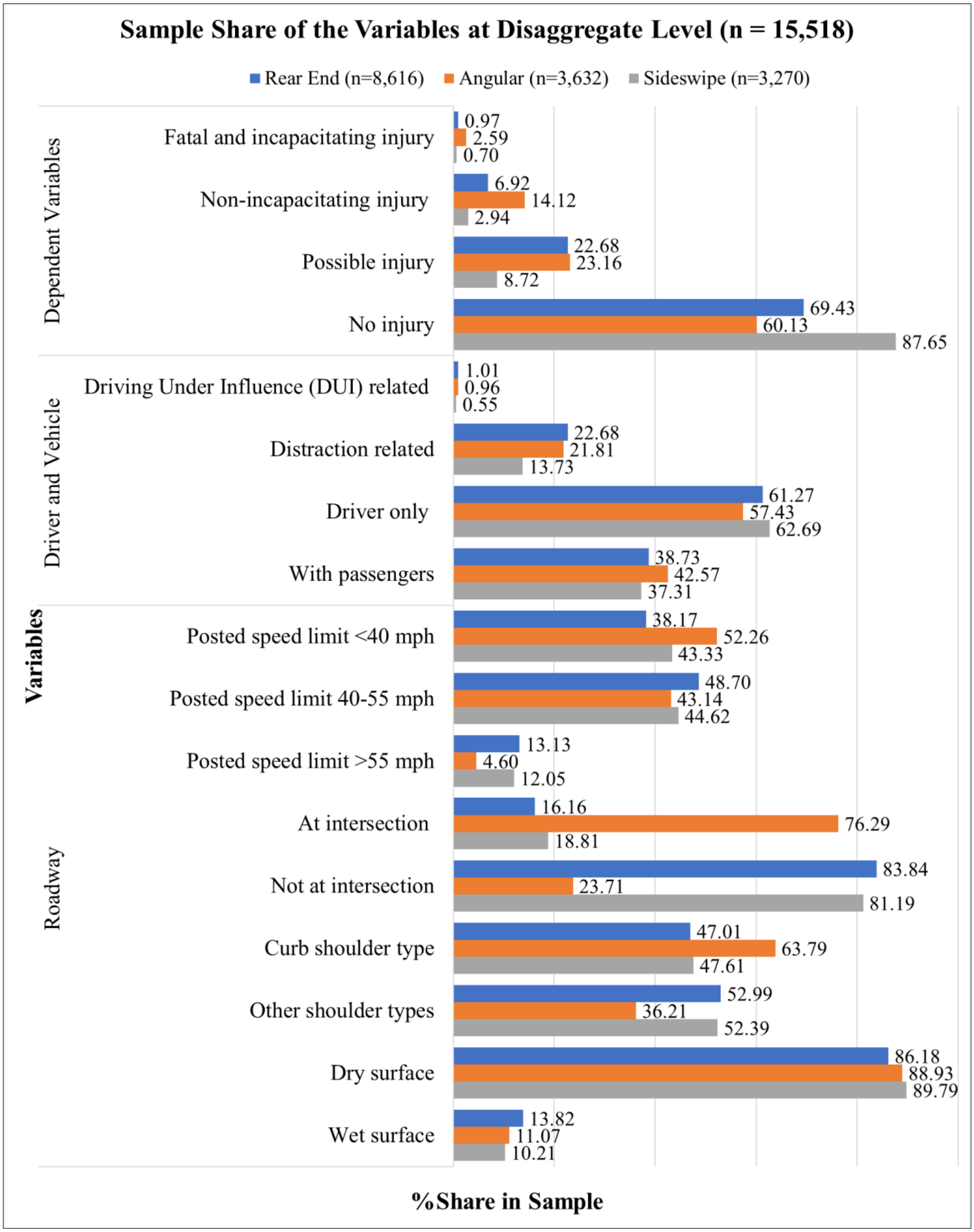

Figure 1a: Sample Share of the Variables at Disaggregate Level (n = 15,518)

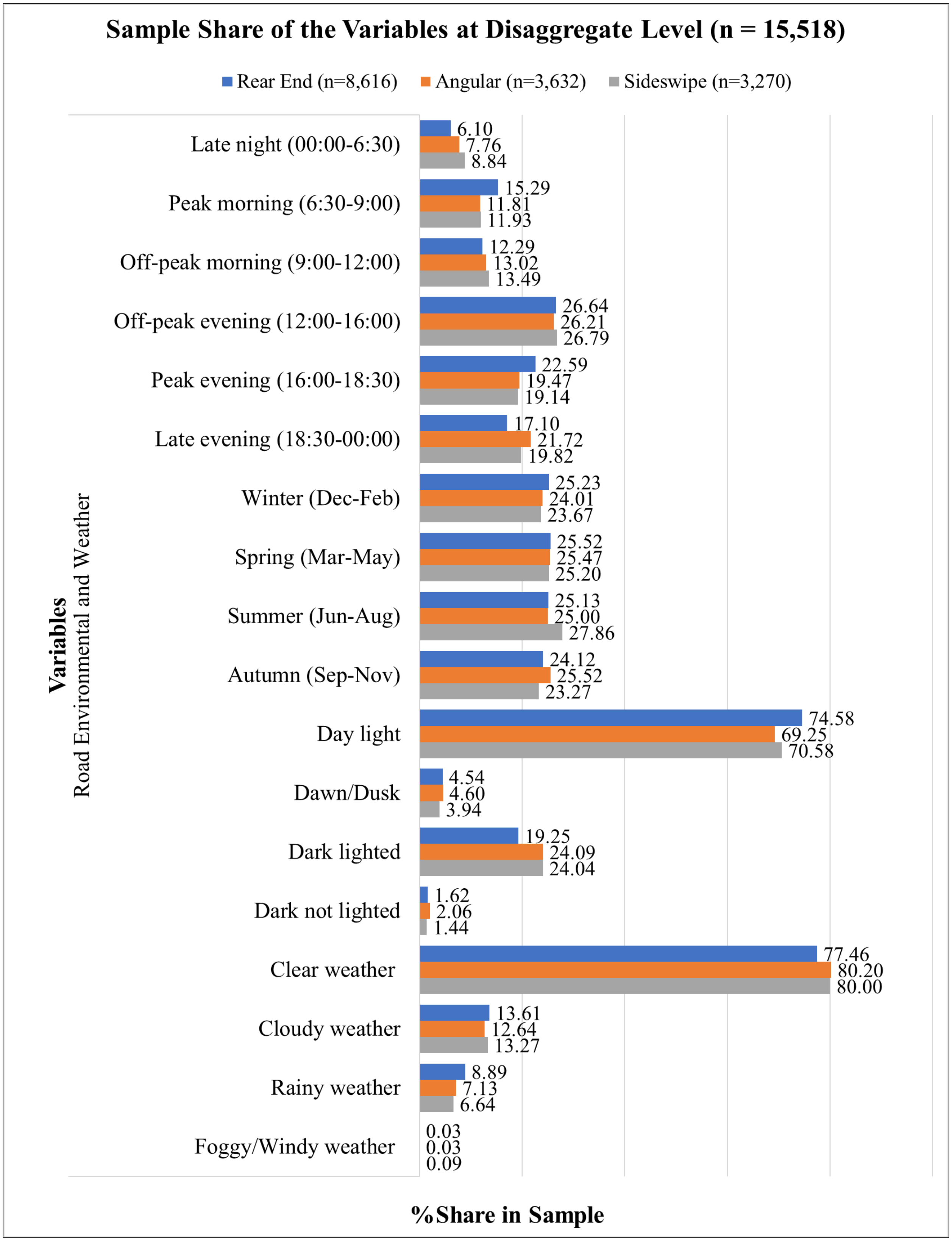

Figure 1b: Sample Share of the Variables at Disaggregate Level (n = 15,518)

Table 1: Summary Statistics of the Variables at Aggregate Level (N = 300 TAZs)

Variables	Definition	Min	Max	Mean	Std. Dev.
Dependent Variables					
Rear End					
Total rear end crashes	Total no. of rear end crashes in a TAZ	0.000	178.000	28.720	27.848
P_FII in rear end	Proportion of FII in rear end crashes	0.000	0.250	0.009	0.025
P_NII in rear end	Proportion of NII in rear end crashes	0.000	0.500	0.070	0.076
P_PI in rear end	Proportion of PI in rear end crashes	0.000	1.000	0.213	0.127
P_NI in rear end	Proportion of NI in rear end crashes	0.000	1.000	0.698	0.157
Angular					
Total angular crashes	Total no. of angular crashes in a TAZ	0.000	78.000	12.107	11.204
P_FII in angular	Proportion of FII in angular crashes	0.000	1.000	0.027	0.085
P_NII in angular	Proportion of NII in angular crashes	0.000	1.000	0.120	0.133
P_PI in angular	Proportion of PI in angular crashes	0.000	1.000	0.233	0.194
P_NI in angular	Proportion of NI in angular crashes	0.000	1.000	0.570	0.258
Sideswipe					
Total sideswipe crashes	Total no. of sideswipe crashes in a TAZ	0.000	110.000	10.900	12.945
P_FII in sideswipe	Proportion of FII in sideswipe crashes	0.000	0.500	0.008	0.040
P_NII in sideswipe	Proportion of NII in sideswipe crashes	0.000	0.667	0.030	0.080
P_PI in sideswipe	Proportion of PI in sideswipe crashes	0.000	1.000	0.084	0.125
P_NI in sideswipe	Proportion of NI in sideswipe crashes	0.000	1.000	0.795	0.280
Roadway Characteristics					
Proportion of interstate-expressways	Interstate-expressways length/total road length	0.000	1.000	0.095	0.200
Proportion of arterial road	Arterial road length/total road length	0.000	1.000	0.451	0.358
Proportion of local road	Local road length/total road length	0.000	0.613	0.027	0.099
Avg. no. of lanes	Average number of lanes in road	1.000	3.500	1.963	0.489
Avg. inside shoulder width	Average inside shoulder width, ft	0.000	18.000	3.008	3.742
Avg. outside shoulder width	Average outside shoulder width, ft	0.000	12.000	4.349	2.013
Proportion of road <40 mph	Length of <40 mph roads/ total road length	0.000	1.000	0.496	0.397
Proportion of road >55 mph	Length of >55 mph roads/ total road length	0.000	1.000	0.105	0.234
Proportion of divided road	Total divided road length/total road length	0.000	1.000	0.610	0.357
Intersection density	Number of intersections/area of TAZ	0.000	0.770	0.085	0.115
Avg. bike lane length	Average bike lane length in TAZ	0.000	3.065	0.177	0.332
Traffic Characteristics					
AADT	Ln (AADT of TAZ)	8.412	13.507	11.189	1.864
Land-use Attributes					
Proportion of institutional area	Total institutional area/TAZ area	0.000	0.991	0.076	0.139
Proportion of commercial area	Total commercial area/TAZ area	0.000	1.000	0.242	0.274
*Land-use mix	Mixed land-use areas/TAZ area	0.000	0.957	0.418	0.242
Built Environment Attributes					
No. of educational centers	**Z score: Number of educational centers	-0.649	3.879	0.000	1.000
No. of restaurants	Z score: Number of restaurants	-0.597	6.690	0.000	1.000
No. of shopping centers	Z score: Number of shopping centers	-0.495	9.698	0.000	1.000
Sociodemographic Factors					

Variables	Definition	Min	Max	Mean	Std. Dev.
NMT transport	Ln (Non-motorized transport means +1)	0.000	4.955	2.128	1.162
Proportion of white American population	Total white American population /total population in TAZ	0.010	0.890	0.410	0.220
Proportion of African American population	Total African American population /total population in TAZ	0.000	0.978	0.223	0.246

**Land-use mix = [$-\frac{\sum(m_h(\ln m_h))}{\ln R}$], where h is the category of land-use, m is the proportion of the developed land area for specific land-use, R is the number of land-use categories; here R= 5 [residential, industrial, institutional, commercial (including office areas) and recreational areas].*
***Z-score represents the standardized form of the actual variable.*

5. EMPIRICAL ANALYSIS

5.1 Model Specification and Overall Measure of Fit

A series of models were estimated for the empirical analysis of the proposed framework. First, we estimated the MNL model for crash type, independent OP models for each crash type and a pooled OP model for disaggregate level crash analysis, and independent NB-OPFS models for each crash type and a panel mixed NB-OPFS model for jointly estimating aggregate level crash count by crash type and severity. The independent model components together provide the benchmark model for comparison. Second, we developed our proposed integrated model system following two approaches: a) exogenous model system: focusing on optimizing the joint log-likelihood of the aggregate and disaggregate level models by only estimating the parameters for propensities aggregated from the disaggregate level models (one parameter per model component i.e., count and severity proportion for each crash type) as shown in equations 11 and 14, and b) endogenous model system: the disaggregate level parameters are estimated based on their contributions to the disaggregate level and the aggregate level models through the disaggregate level propensity components embedded within the aggregate level propensity equation (as shown in equations 12 and 15). Third, we identified the best model by comparing the model performance based on Bayesian Information Criterion (BIC). The BIC for a given empirical model is equal to:

$$BIC = -2LL + Np\ Ln\ (Ob) \tag{21}$$

where *LL* is the log-likelihood value at convergence, *Np* is the number of parameters and *Ob* is the number of observations. The model with the lower BIC is the preferred model.

The corresponding BIC (LL) values of the models are: (1a) independent model system (MNL, OP, and NB-OPFS) (with 108 parameters): 50,020.544 (-24,651.719), (1b) non-integrated model (MNL, pooled OP, and panel mixed NB-OPFS) (with 80 parameters): 49,872.460 (-24,670.635), (2) exogenous model (with 70 parameters): 48,839.107 (-24,187.158), (3) endogenous model (with 67 parameters): 48,851.165 (-24,203.147), and (4) exogenous model with unobserved heterogeneity (with 74 parameters): 48,829.406 (-24,169.028). Based on the BIC values, three specific observations could be drawn. First, the panel modeling approaches provide improved data fit compared to independent models with lower BIC values supporting the findings of previous studies (Bhowmik et al., 2021b, 2019). Second, all the integrated systems provide improved data fit as evidenced by the lower BIC values in comparison to the non-integrated model. Third, within the integrated systems, our proposed exogenous model provides the lowest BIC indicating the best data fit in comparison to the proposed endogenous model. Finally, we accommodate unobserved heterogeneity in our exogenous model (the best model in terms of data fit) and find that the model provides further improved BIC (lower).

5.2 Model Estimation Results

This section provides a detailed discussion of the exogenous variables affecting the crash count by crash type and severity at aggregate level and the factors influencing the crash outcome variables at the disaggregate level models. Table 2 presents the model estimation results for exogenous model with unobserved heterogeneity. The reader would note that the different model components (except for crash type) were estimated using a pooled estimation process i.e., a common parameter was estimated across the three components and interactions were used to identify deviations across the three crash types. If the deviations were significant, they were retained or else dropped. The number of parameters in Table 2 describes the number of unique statistical parameters for the model component from a possible set of three (one effect for each crash type). The reader would also note that a positive (negative) sign for a variable in Table 2 indicates that an increase in the variable is likely to result in more (less) crashes as well as exhibit a higher (lower) impact on crash type and severity. The results of the non-integrated model system (MNL, pooled OP and panel mixed NB-OPFS models) with the net variable effects are presented in Table A.1 and Table A.2 of the Appendix.

5.3 Disaggregate Level Attributes

5.3.1 Crash Type Component

The constant variables in the crash type model do not have substantive interpretation after including independent variables.

Among driver attributes, driving under the influence reduces the risk of sideswipe crash type and increases risk for other crash types (see Neyens and Boyle, 2007; Razi-Ardakani et al., 2018 for similar results). The crashes that are distraction related are more likely to result in angular crashes and less likely to result in sideswipe crashes (relative to rear end crashes). In the presence of passengers, the findings indicate that angular crashes are more likely to occur.

The posted speed limit variable offers interesting findings. The likelihood of angular and sideswipe crash types is higher for road sections with posted speed limit <40mph (relative to rear end crash type). On the other hand, for road sections with posted speed limit >55mph the likelihood of angular crash type is lower (relative to other crash types). The results indicate how posted speed limit potentially affects crash types. The model results offer interesting results for intersection location. We find that the propensity of angular crashes is higher relative to rear end and sideswipe crashes at intersections. The finding might be reflective of the turning movements occurring at intersections potentially resulting in angular crashes. The presence of a curb shoulder also increases the propensity for angular crashes. In the presence of curb shoulder, drivers might consider riskier maneuvers for turning potentially resulting in angular crashes. The pavement condition being wet increases the propensity for rear end crashes. On a wet surface, vehicles are likely to slip and lose traction resulting in higher incidence of rear end crashes. All these findings are consistent with many previous studies (Bhowmik et al., 2019; Razi-Ardakani et al., 2018).

Table 2: Results of the Integrated (Exogenous Model) with Unobserved Heterogeneity with Net Variable Effects

Disaggregate Level										
Variables	MNL Model Propensity (Base: Rear End)				OP Model Propensity					
	Angular		Sideswipe		Rear End		Angular		Sideswipe	
	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat
Constant (2, --) *	-2.448	-36.557	-0.733	-13.750	--	--	--	--	--	--
Threshold Parameters										
Threshold between NI-PI (--, 3)	--	--	--	--	0.671	29.019	0.372	9.125	1.299	36.120
Threshold between PI-NII (--, 3)	--	--	--	--	1.604	3.130	1.103	9.482	1.937	7.281
Threshold between NII-FII (--,1)	--	--	--	--	2.548	1.827	2.047	1.827	2.882	1.827
Driver and Vehicle Factors										
DUI related (if yes 1, otherwise 0) (1, 1)	--	--	-0.689	-2.512	0.525	4.419	0.525	4.419	--	--
Distraction related (if yes 1, otherwise 0) (2, 1)	0.190	3.050	-0.544	-8.880	0.238	8.953	0.238	8.953	0.238	8.953
With passengers (if yes 1, otherwise 0) (1, 1)	0.133	2.648	--	--	0.336	14.847	0.336	14.847	0.336	14.847
Roadway Characteristics										
Posted speed limit (Base: 40-55mph)										
Posted speed limit<40 (2, 1)	0.199	3.750	0.181	3.982	-0.082	-3.574	-0.082	-3.574	-0.082	-3.574
Posted speed limit>55 (1, --)	-0.314	-2.865	--	--	--	--	--	--	--	--
At intersection (if yes 1, otherwise 0) (1, --)	2.717	53.507	--	--	--	--	--	--	--	--
Curb shoulder (if yes 1, otherwise 0) (1, 1)	0.504	9.985	--	--	--	--	-0.133	-3.063	--	--
Wet surface (if yes 1, otherwise 0) (2, 1)	-0.345	-4.348	-0.341	-4.791	--	--	--	--	0.307	2.237
Road Environmental and Weather Factors										
Time (Base: Off-peak evening, late evening)										
Late night (12:00-6:30) (1, 1)	--	--	0.211	2.410	0.232	4.629	0.232	4.629	--	--
Peak morning (6:30-9:00) (1, --)	--	--	-0.337	-4.832	--	--	--	--	--	--
Off-peak morning (9:00-12:00) (--, 1)	--	--	--	--	--	--	0.139	2.228	--	--
Peak evening (16:00-18:30) (1, 1)	--	--	-0.220	-3.785	-0.086	-2.460	--	--	--	--
Peak morning and peak evening (1, --)	-0.163	-3.009	--	--	--	--	--	--	--	--
Season (Base: Summer)										
Winter (1, --)	--	--	-0.163	-2.666	--	--	--	--	--	--
Spring (1, --)	--	--	-0.120	-1.990	--	--	--	--	--	--
Autumn (1, --)	--	--	-0.182	-2.953	--	--	--	--	--	--
Light condition (Base: Day light, dawn/dusk)										
Dark lighted (1, --)	--	--	0.160	2.761	--	--	--	--	--	--
Dark not lighted (1, --)	0.505	2.789	--	--	--	--	--	--	--	--
Weather condition (Base: Clear)										
Rainy (--, 1)	--	--	--	--	--	--	--	--	-0.455	-2.421
Fog and wind (--, 1)	--	--	--	--	--	--	-2.946	-9.928	--	--

Aggregate Level												
Variables	Count Component						Severity Proportion Component					
	Rear End		Angular		Sideswipe		Rear End		Angular		Sideswipe	
	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat
Constant (3, --)	-0.765	-6.027	-1.436	-11.275	-0.977	-8.207	--	--	--	--	--	--
Threshold Parameter												
Threshold between P_NI-P_PI (--, 2)	--	--	--	--	--	--	0.262	2.673	0.262	2.673	1.404	9.495
Threshold between P_PI-P_NII (--, 3)	--	--	--	--	--	--	1.146	3.372	1.048	4.827	2.045	4.801
Threshold between P_NII-P_FII (--, 1)	--	--	--	--	--	--	2.013	2.386	1.916	2.386	2.913	2.386
Roadway and Traffic Characteristics												
AADT (1, --)	0.021	2.320	0.021	2.320	0.021	2.320	--	--	--	--	--	--
Proportion of interstate-expressways (--, 1)	--	--	--	--	--	--	0.210	2.001	--	--	--	--
Proportion of local road (--, 1)	--	--	--	--	--	--	--	--	--	--	-0.927	-2.590
Avg. no. of lanes (--, 1)	--	--	--	--	--	--	-0.257	-9.396	--	--	--	--
Avg. inside shoulder width (1, 1)	0.009	1.944	--	--	--	--	--	--	-0.026	-3.651	-0.026	-3.651
Avg. outside shoulder width (--, 1)	--	--	--	--	--	--	--	--	--	--	0.052	2.068
Proportion of road <40 mph (1, --)	-0.183	-4.013	--	--	--	--	--	--	--	--	--	--
Proportion of divided road length (1, --)	--	--	0.137	1.874	--	--	--	--	--	--	--	--
Intersection density (--, 1)	--	--	--	--	--	--	--	--	-1.251	-4.440	--	--
Avg. bike lane length (1, --)	--	--	-0.029	-2.877	--	--	--	--	--	--	--	--
Land-use and Built Environment Attributes												
Proportion of institutional area (1, --)	--	--	--	--	0.384	2.063	--	--	--	--	--	--
No. of restaurants (--, 1)	--	--	--	--	--	--	--	--	-0.113	-4.129	--	--
No. of shopping centers (1, 1)	0.048	1.972	--	--	0.048	1.972	--	--	--	--	-0.098	-2.250
Over dispersion (3, --)	0.106	1.728	0.058	4.709	0.082	4.187	--	--	--	--	--	--
Parameter for Disaggregate Level Propensity Sum (2, 1)	0.993	46.791	1.007	38.112	1.007	38.112	0.076	2.645	0.076	2.645	0.076	2.645
Unobserved Heterogeneity												
Constant ($\boldsymbol{\theta}_{il}$) (3)	0.077	2.136	0.159	5.721	0.158	3.490	0.077	2.136	0.159	5.721	0.158	3.490
Proportion of commercial area (1)	0.188	3.605	0.188	3.605	0.188	3.605	0.188	3.605	0.188	3.605	0.188	3.605
LL = -24,169.028												
BIC = 48,829.406												
Number of parameters = 74												

*Note: "--" denotes variables are not significant at 90% confidence interval; *Numbers in the parenthesis denote the number of parameters estimated; FII=Fatal and incapacitating injury, NII=Non-incapacitating injury, PI= Possible injury, and NI= No injury.*

The model results show that late nighttime period increases risk for sideswipe crashes while peak morning and peak evening time periods reduce angular and sideswipe crashes. Among the seasonal effects, winter, spring, and autumn reduce the possibility of sideswipe crash type compared to other crash types.

Our model estimates indicate that under dark-lighted conditions, the propensity of sideswipe crashes is higher than for other crash types. On the other hand, dark-unlighted condition is associated with high risk of angular crash type (as found in Neyens and Boyle, 2007). The results indicate that under dark conditions there is an overall increase in angular and sideswipe crashes. The result might be highlighting how vehicles are less likely to be visible in turning movements and thus might possibly result in higher angular and sideswipe crashes.

5.3.2 Crash Severity Component

The threshold parameters demarcate the various severity categories and do not have any substantive interpretation.

In the severity model, among driver and vehicle attributes, as expected, driving under influence, distracted conditions and presence of passengers in the vehicle contribute to higher severity likelihood across all crash types (see Das et al., 2009; Marcoux et al., 2018; Paleti et al., 2010; Weiss et al., 2014; Yasmin and Eluru, 2013 for similar findings).

With regards to roadway attributes, as expected, lower posted speed limit is associated with lower severity across all the crash types. The presence of a curb shoulder reduces injury severity risk for angular crashes possibly by additional reduction of vehicle maneuvering speed (Jiang et al., 2013). The results also show that wet pavement surface condition increases the severity propensity for sideswipe crashes.

The impact of time of the day variables offers different trends by crash type. Rear end and angular crashes occurring in the late nighttime period are likely to result in severe injury. The vehicle operating speeds are likely to be higher in these time periods and are thus likely to result in severe crashes (see Marcoux et al., 2018 for similar findings). Angular crashes in the off-peak morning time period are also likely to result in severe injuries. However, rear end crashes during peak evening period are likely to result in less severe crashes (as found in Behnood and Mannering, 2019). The higher traffic volume and lower speeds during peak hours are likely to contribute to less severe angular crashes.

Among several weather attributes considered, rainy conditions and fog and wind conditions exhibit a discernible impact on crash type-based severity. In rainy weather conditions, sideswipe crashes are likely to result in reduced severity. In foggy and windy weather, we observe a reduced severity risk for angular crashes. These findings perhaps reflect the increased driver caution while driving in adverse weather conditions (Abrari Vajari et al., 2020; Uddin and Huynh, 2020).

5.4 Aggregate Level Attributes

5.4.1 Count Component

The constant terms in the crash frequency by crash type do not have substantive interpretation after including independent variables.

Among the several roadway and traffic factors, AADT variable offers expected results. TAZs with higher AADT are more likely to experience increased risk across all crash types (see Alarifi et al., 2017; Alhomaidat et al., 2020; Cai et al., 2019; Huang et al., 2016; Ivan et al., 2023 for similar results). The parameters for average inside shoulder width show positive association with crash frequency for rear end crash type. The presence of inside shoulder may provide shelter for emergency stops. However, stopped vehicles may be hazardous and contribute to rear end crashes (Stamatiadis et al., 2009). The variable associated with proportion of divided road length shows positive effect on angular crash frequency. This finding is consistent with several previous studies (Bhowmik et al., 2019, 2018). The impact of posted speed limit proportions across the TAZ offers intuitive results. An increase in the proportion of <40 mph roads in a TAZ reduces the risk of rear end crashes. The average bike lane length in the zone is negatively associated with angular crash frequency. The result is quite interesting and is supportive of addition of bike infrastructure in the zones (Pervaz et al., 2022).

Among the land-use characteristics, the results found that a higher proportion of institutional area in the zone increases the propensity for sideswipe crashes.

With regards to the built environment attributes explored, the number of shopping centers was found to be strongly associated with increased rear end and sideswipe crashes. These results serve as additional surrogate variables for traffic intensity (Bhowmik et al., 2022, 2021b).

5.4.2 Severity Proportion Component

The threshold parameters demarcate the various severity proportion categories and do not have any substantive interpretation.

Among roadway and traffic factors, it is interesting to note that AADT has no impact on severity proportion. The zones with a higher proportion of interstate-expressways indicate a higher risk of severe rear end crashes. While the result might be counter-intuitive, it is potentially reflecting the presence of ramps close to interstates and signalized intersections near these ramps where rear end crashes are likely to occur as drivers transition from high-speed facilities to low-speed roadways. On the other hand, a higher proportion of local roads is associated with lower severity proportions for sideswipe crashes possibly due to the lower vehicle operating speed (Pervaz et al., 2023; Yasmin and Eluru, 2018). As the average number of lanes increase in a zone, the propensity for rear end crash severity proportion is likely to reduce. The inner and outer shoulder width variables offer contrasting results. The inner shoulder width is associated with lower severity risk for angular and side swipe crash types. The outer shoulder width is associated with higher severity risk for sideswipe crashes. These results while interesting might require further analysis. In terms of intersection density, the findings indicate that severity proportions for angular crash type are likely to tend toward less severe crashes. The increased presence of intersections offers increased protection for turning movements and thus reduce injury severity in the event of a crash (Bhowmik et al., 2021b).

Among built environment variables, an increased presence of restaurants is likely to reduce the severity proportion for angular crashes. In a similar manner, increased presence of shopping centers is associated with lower severity risk for sideswipe crashes. These findings, similar to crash occurrence, reflect increased traffic density and lower vehicle operating speed (Bhowmik et al., 2021b; Yasmin and Eluru, 2018).

5.5 Parameter for Disaggregate Level Propensity Sum

The coefficients for the fixed propensity from the disaggregate level models in the count component and severity component are presented in the lower row panel of Table 2. The positive sign of the parameters for examined crash types in both count and severity proportion component indicates that a higher value of disaggregate level model propensity is likely to increase the number of crashes and the crash severity for rear end, angular and sideswipe crashes. The results clearly highlight that the higher propensities for crash type and severity at the disaggregate level are significantly associated with the increase of the crash counts by crash type and severity at the aggregate level. It is also interesting to note that the impact of disaggregate propensity does not vary across the three crash types in the severity component.

5.6 Unobserved Heterogeneity

As described in the methodology section, the proposed model system can accommodate unobserved heterogeneity while estimating the models. In our estimation, we considered spatial correlation among crash records, and several other correlations such as correlations within crash count, within crash severity proportion, and between crash count and severity proportion for all the crash types explored in the analysis, and random parameter effects of the variables. The spatial correlation is estimated in the form of spatial variations for crash records in a zone through a common spatial correlation between all crash records (for similar crash type) from a zone. The unobserved heterogeneity variable constant presented in Table 2 corresponds to this common zone spatial correlation. The significant effect of this correlation parameter ($\boldsymbol{\theta}_{il}$) clearly highlights the presence of common unobserved factors across crash records of similar crash types in the same zone. We also parameterized the correlation parameter and tested for several independent variables. In our testing, we found a zonal variable – proportion of commercial area - exhibit significant unobserved correlation across crash records within the same zone. On the other hand, no statistically significant effect was recovered for unobserved correlations between crash counts and crash proportions by types and severity levels *(*η), and random parameter effects *(*$\boldsymbol{\zeta}$ and $\boldsymbol{\rho}$*)* in our dataset for both non-integrated and proposed integrated model systems.

5.7 Predictive Performance of the Model

To demonstrate the applicability of the proposed model, we undertake a comparison exercise between the proposed integrated exogenous model with unobserved heterogeneity and the non-integrated panel mixed NB-OPFS model by testing model performance on estimation and holdout samples. We compare the models by employing two measures of fit: mean absolute deviation (MAD) and mean squared prediction error (MSPE) (please see Bhowmik et al., 2018; Pervaz et al., 2023 for a detailed definition of these measures). The model with the lower values of MAD and MSPE provides better predictions for the observed data.

Figure 2 presents the values of these measures for the proposed integrated exogenous model and the non-integrated panel mixed NB-OPFS model. From the figure, we observe that our proposed integrated model performs better than non-integrated model across most of the measures for all crash types. For rear end crash type, our proposed model performs better for all measures (for both estimation and validation samples). For angular and sideswipe crash types, our model shows better performance across most of the MAD and MSPE values. Though non-integrated system shows equal or lower deviation values across a few measures for angular (6 values out of 20) and sideswipe (4 values out of 20) crash types, the differences are very marginal. In summary, the resulting goodness of fit measures clearly show the comparable performance of our proposed integrated model that considers both aggregate and disaggregate level crash attributes than traditional non-integrated model for estimating crashes by type and severity.

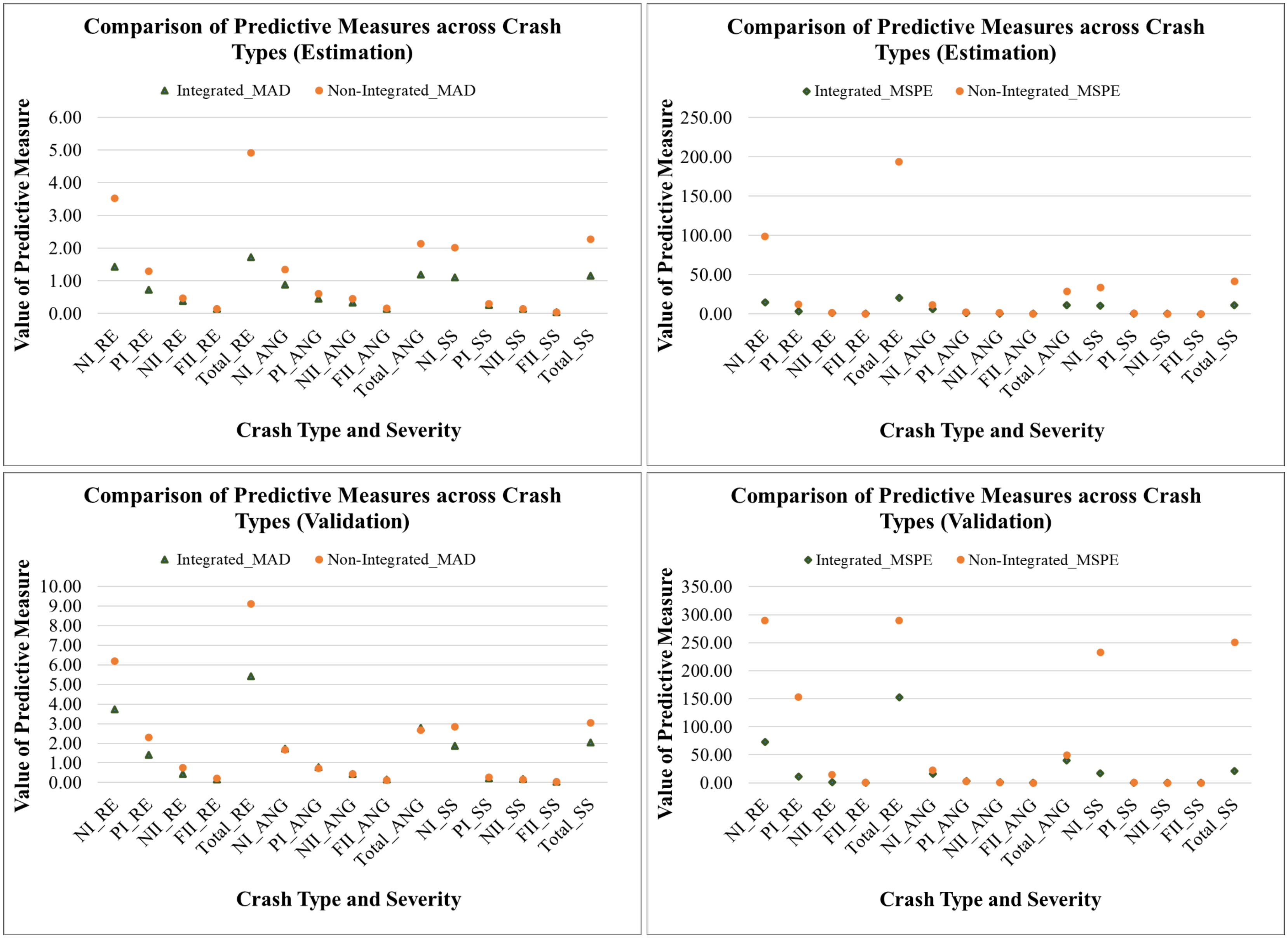

Figure 2: Predictive Performance of the Models across Crash Types for Estimation and Validation Datasets

5.8 Elasticity Analysis

The model estimation results presented in Table 2 represent a joint interaction of aggregate and disaggregate level variables and do not directly provide the magnitude of the effects of the variables on crash count by crash type and severity. The actual magnitudes of the effect of the variables can be obtained by computing elasticity effects of the variables. The proposed model offers a unique framework to capture the information flow from disaggregate level into the estimation of aggregate level crash count by crash type and severity. Therefore, for elasticity computation, we intend to focus on showing the elasticity effects of disaggregate level variables by following the procedure demonstrated in Eluru and Bhat (2007). Following this procedure, the percentage change in the expected total zonal crash counts by crash type and severity caused by the change in the disaggregate level exogenous variable were computed. As all the exogenous variables in the disaggregate level are indicator variables, we obtain these changes by changing the value of the variable to one for the subsample of observations for which the variable takes a value of zero and to zero for the subsample of observations for which the variable takes a value of one. Specifically, the elasticity effects computed in this procedure are aggregated percentage elasticity based on the aggregated change and the overall shares of the sample.

The computed elasticities are presented in Table 3 and Table 4. The reader would note that we considered the rear end crash type as the base category while allowing for the information flow from the disaggregate level MNL model. Therefore, the elasticity analysis in this section provides the results for angular and sideswipe crash types relative to rear end crash type. The results presented in Table 3 and Table 4 show the percentage change in the number of angular and sideswipe crashes respectively by different severity level due to the changes in the disaggregate level exogenous variable of interest.

Table 3: Elasticity Effects for Angular Crash Type

Variables	%NI Crashes	%PI Crashes	%NII Crashes	%FII Crashes	%Total Crashes
Driver and Vehicle Factors					
DUI related	-2.51	2.48	5.64	9.50	0.00
Distraction related	19.05	21.56	23.09	24.90	20.30
With passengers	11.93	15.19	17.17	19.52	13.55
Roadway Characteristics					
Posted speed limit (Base: 40-55mph)					
Posted speed limit<40 mph	20.24	19.54	19.12	18.61	19.89
Posted speed limit>55 mph	-27.45	-27.47	-27.48	-27.49	-27.46
At intersection	252.11	249.22	247.28	245.04	250.64
Curb shoulder	48.32	47.15	46.44	45.59	47.74
Wet surface	-30.43	-30.43	-30.42	-30.42	-30.43
Road Environmental and Weather Factors					
Time (Base: Off-peak evening, late evening)					
Late night (12:00-6:30)	-1.10	1.11	2.47	4.10	0.00
Peak morning (6:30-9:00)	-12.81	-12.80	-12.81	-12.82	-12.81
Off-peak morning (9:00-12:00)	-0.66	0.66	1.47	2.43	0.00
Peak evening (16:00-18:30)	-14.07	-14.02	-14.00	-13.97	-14.04
Light condition (Base: Day light, dawn/dusk)					
Dark lighted	--	--	--	--	--
Dark not lighted	65.70	65.64	65.59	65.54	65.67
Weather condition (Base: Clear)					
Rainy	--	--	--	--	--
Fog and wind	13.46	-15.25	-28.51	-41.46	0.00

For example, in Table 3, the elasticity estimate for the DUI related variable indicates that driving under influence increases the fatal and incapacitating angular crash count by 9.50%. For this variable, we also find that the change in the crash counts for higher severity categories is greater than the change in the lower severity categories while demonstrating insignificant changes in the total angular crash count. To be specific, for angular crash type, the DUI related variable provides significant information flow for crash severity (from OP model component) while providing insignificant effect on total angular crash count (from MNL model component). These findings are consistent with the results presented in Table 2. The effects of all the variables presented in Table 3 and Table 4 can be interpreted in a similar manner for angular and sideswipe crash type respectively.

Table 4: Elasticity Effects for Sideswipe Crash Type

Variables	%NI Crashes	%PI Crashes	%NII Crashes	%FII Crashes	%Total Crashes
Driver and Vehicle Factors					
DUI related	-50.26	-50.28	-50.28	-50.29	-50.27
Distraction related	-46.42	-44.52	-43.69	-42.59	-46.12
With passengers	-0.64	3.45	5.17	7.41	0.00
Roadway Characteristics					
Posted speed limit (Base: 40-55mph)					
Posted speed limit<40 mph	18.49	17.48	17.06	16.52	18.33
Posted speed limit>55 mph	--	--	--	--	--
Wet surface	-30.66	-27.77	-26.53	-24.91	-30.20
Road Environmental and Weather Factors					
Time (Base: Off-peak evening, late evening)					
Late night (12:00-6:30)	23.19	23.17	23.17	23.16	23.19
Peak morning (6:30-9:00)	-29.99	-30.01	-30.01	-30.02	-30.00
Off-peak morning (9:00-12:00)	--	--	--	--	--
Peak evening (16:00-18:30)	-20.75	-20.75	-20.75	-20.75	-20.75
Season (Base: Summer)					
Winter	-15.75	-15.75	-15.75	-15.75	-15.75
Spring	-11.74	-11.74	-11.75	-11.75	-11.74
Autumn	-17.53	-17.54	-17.54	-17.54	-17.54
Light condition (Base: Day light, dawn/dusk)					
Dark lighted	16.78	16.79	16.79	16.78	16.78
Dark not lighted	--	--	--	--	--
Weather condition (Base: Clear)					
Rainy	0.86	-4.63	-6.83	-9.62	0.00
Fog and wind	--	--	--	--	--

By analyzing the elasticity results presented in Table 3 and Table 4, several important observations can be drawn. First, there are differences in the elasticity effects across the expected number of crashes for different crash types and severities. Second, the most significant variables affecting the expected number of total angular crashes are intersection location, dark unlighted condition, curb shoulder type, distracted driving, speed limit <40 mph road, driving with passengers as shown in Table 3. On the other hand, the most significant variables affecting the number of total sideswipe crashes are late night, lower speed limit road, dark lighted condition as shown in Table 4. Third, the most significant variables affecting the number of angular crash counts in the higher severity categories are distracted driving, driving with passengers, DUI driving, late night and off-peak morning (see Table 3). On the other hand, driving with passenger increases the expected number of <u>sideswipe crash counts in the higher severity categories</u> (see Table 4).

The results of the elasticity effect analysis highlight the influence of disaggregate level variables on overall crash frequency by type and severity. Traditional approaches that develop separate models for aggregate and disaggregate resolutions cannot identify any impact of disaggregate variables on crash frequency by type and severity. For instance, in our results, we highlight the impact of disaggregate level information such as driver behavior and vehicle features (such as DUI, and distraction related crashes), weather and environmental factors (such as crash timing, lighting and Fog/Wind) in estimating crash frequency. The multifaceted information available from the integrated approach can contribute to the implications of the road safety policy by providing traffic engineers, planners, vehicle manufacturers, psychologists, environmentalists, and law enforcement agencies to adopt integrated and coordinated decisions. For example, results indicate that intersection locations contribute to a higher number of crashes of the angular type. Therefore, intersection improvement policies such as providing dedicated/exclusive turning lanes, signal and signage improvement scheme, improvement of driving behavior for yielding to the signals and signages in the zone might help to mitigate this type of crash. Furthermore, policies such as targeted enforcement, road safety awareness campaigns and large-scale traffic safety education programs can be adopted in the zones with a higher shared transport use, higher DUI and distraction driving rates. This policy initiative would not be possible from a non-integrated model system. Similarly, roadway improvement and maintenance programs such as lighting improvement can be accelerated in the zones to improve visibility for safer traffic movement, particularly during dark conditions. Overall, a better understanding of the potential crash contributing factors from different analysis levels and their possible interactions can lead to better policy decisions and increase the effectiveness of various road safety interventions.

6. CONCLUSIONS

The recent development of integrated multi-resolution approaches has enriched the transportation safety literature by allowing the safety analysts to capture the impacts of the variables from disaggregate crash record data within aggregate crash models. These approaches augment the traditional aggregate crash model systems with rich observed information available in the crash records while also accommodating for unobserved effects. A recent study illustrated this framework incorporating disaggregate severity data with crash frequency models by severity. The current study builds on the previous effort through a unified framework that allows for the information flow of observed and unobserved variables from the disaggregate level crash type and crash severity model components into the aggregate level crash frequency model. The approach involves summing up the crash propensities of disaggregate level crash type and crash severity models within the aggregate resolution and adding the generated values as new variables in the aggregate level propensity estimation. In this study, we employed a panel mixed NB-OPFS framework at the aggregate level model to jointly examine crash frequency by crash type and severity and the MNL and pooled OP models at the disaggregate level to analyze the crash type and crash severity, respectively. In the panel mixed NB-OPFS framework, the NB component models the number of crashes by type and the OPFS component determines the proportion of each severity in the pooled dataset for a zone. The introduction of the disaggregate measures as a composite score in aggregate models can be accommodated exogenously or endogenously. In the exogenous approach, the disaggregate level parameters are fixed and only the parameters on the composite scores are estimated. In the endogenous approach, the disaggregate model parameters are also allowed to vary. The variation that offers the superior data fit is preferred. The empirical analysis was conducted using 2019 crash data drawn from the City of Orlando, Florida. We considered three crash types: rear end, angular and sideswipe over 300 traffic analysis zones (TAZs) for the analysis. The disaggregate model component considered crash specific factors, vehicle and driver factors, roadway attributes, road environmental and weather information. For aggregate level analysis, independent variables including roadway and traffic characteristics, land-use attributes, built environment factors, and sociodemographic characteristics were considered.

For the empirical assessment of the proposed integrated framework, the exogenous and endogenous model systems were compared with the non-integrated model system (composed of MNL, and a pooled OP model for three crash types at disaggregate level, and a panel mixed NB-OPFS model for jointly estimating aggregate level crash count by crash type and severity). The study results clearly highlighted the improved performance of the proposed integrated models over non-integrated model system. Within the integrated model approaches, exogenous model outperformed the endogenous model in terms of BIC value. Finally, the exogenous model was further improved by accommodating unobserved heterogeneity. We compared the performance of the proposed integrated model with the non-integrated model system using several predictive performance measures. The measures clearly highlighted the improved performance of our proposed integrated model in estimation and holdout samples. Further, an elasticity exercise was conducted to illustrate how the influence of disaggregate level crash attributes can be examined on the aggregate level crash count by crash type and severity analysis.

This study is not without limitations. The proposed integrated approach requires substantial effort for data compilation in the region. The compilation can also be cumbersome as data from various sources are needed leading to the handling of large datasets and substantial data processing and coding resources. In addition, the model framework requires systematic analysis i.e., the approach should start from disaggregate level analysis and then integrated with the aggregate level analysis. Further, the current study considered one year crash data for the empirical analysis of the proposed framework as we obtained a good number of crash records from one year data for the study area. It would be useful to consider the data from multiple years while also accounting for potential temporal heterogeneity of the parameter estimates within the proposed integrated framework in future research efforts

ACKNOWLEDGMENTS

The authors would like to gratefully acknowledge the Signal Four Analytics (S4A), Florida Department of Transportation (FDOT) and other data sources for providing access to Florida crash and geospatial data.

AUTHOR CONTRIBUTION STATEMENT

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

REFERENCES

Abrari Vajari, M., Aghabayk, K., Sadeghian, M., Shiwakoti, N., 2020. A multinomial logit model of motorcycle crash severity at Australian intersections. Journal of Safety Research 73, 17–24.

Ahmed, S.S., Corman, F., Anastasopoulos, P.Ch., 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. Analytic Methods in Accident Research 37, 100250.

Alarifi, S.A., Abdel-Aty, M., Lee, J., 2018. A Bayesian multivariate hierarchical spatial joint model for predicting crash counts by crash type at intersections and segments along corridors. Accident Analysis and Prevention 119, 263–273.

Alarifi, S.A., Abdel-Aty, M.A., Lee, J., Park, J., 2017. Crash modeling for intersections and segments along corridors: A Bayesian multilevel joint model with random parameters. Analytic Methods in Accident Research 16, 48–59.

Alhomaidat, F., Kwigizile, V., Oh, J.-S., Houten, R. Van, 2020. How does an increased freeway speed limit influence the frequency of crashes on adjacent roads? Accident Analysis and Prevention 136, 105433.

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. Analytic Methods in Accident Research 34, 100205.

Anastasopoulos, P.Ch., 2016. Random parameters multivariate tobit and zero-inflated count data models: Addressing unobserved and zero-state heterogeneity in accident injury-severity rate and frequency analysis. Analytic Methods in Accident Research 11, 17–32.

Aptech, 2015. Aptech Systems Inc.

Balusu, S.K., Pinjari, A.R., Mannering, F.L., Eluru, N., 2018. Non-decreasing threshold variances in mixed generalized ordered response models: A negative correlations approach to variance reduction. Analytic Methods in Accident Research 20, 46–67.

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

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

Bhowmik, T., Rahman, M., Yasmin, S., Eluru, N., 2021a. Exploring analytical, simulation-based, and hybrid model structures for multivariate crash frequency modeling. Analytic Methods in Accident Research 31, 100167.

Bhowmik, T., Yasmin, S., Eluru, N., 2022. Accommodating for systematic and unobserved heterogeneity in panel data: Application to macro-level crash modeling. Analytic Methods in Accident Research 33, 100202.

Bhowmik, T., Yasmin, S., Eluru, N., 2021b. A New Econometric Approach for Modeling Several Count Variables: A Case Study of Crash Frequency Analysis by Crash Type and Severity. Transportation Research Part B 153, 172–203.

Bhowmik, T., Yasmin, S., Eluru, N., 2019. 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 24, 100107.

Bhowmik, T., Yasmin, S., Eluru, N., 2018. A joint econometric approach for modeling crash counts by collision type. Analytic Methods in Accident Research 19, 16–32.

Cai, Q., Abdel-Aty, M., Lee, J., Huang, H., 2019. Integrating macro- and micro-level safety analyses: a Bayesian approach incorporating spatial interaction. Transportmetrica A 15(2), 285–306.

Chang, F., Yasmin, S., Huang, H., Chan, A.H.S., Haque, Md.M., 2021. Injury severity analysis of motorcycle crashes: A comparison of latent class clustering and latent segmentation based models with unobserved heterogeneity. Analytic Methods in Accident Research 32, 100188.

Cui, H., Xie, K., 2021. An accelerated hierarchical Bayesian crash frequency model with accommodation of spatiotemporal interactions. Accident Analysis and Prevention 153, 106018.

Das, A., Abdel-Aty, M., Pande, A., 2009. Using conditional inference forests to identify the factors affecting crash severity on arterial corridors. Journal of Safety Research 40(4), 317–327.

Eluru, N., Bhat, C.R., 2007. A joint econometric analysis of seat belt use and crash-related injury severity. Accident Analysis and Prevention 39(5), 1037–1049.

Gong, H., Wang, F., Zhou, B. (Brenda), Dent, S., 2020. Application of random effects negative binomial model with clustered dataset for vehicle crash frequency analysis. International Journal of Transportation Science and Technology 9(3), 183–194.

Gu, X., Yan, X., Ma, L., Liu, X., 2020. 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 and Prevention 144, 105674.

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. Analytic Methods in Accident Research 33, 100191.

Huang, H., Song, B., Xu, P., Zeng, Q., Lee, J., Abdel-Aty, M., 2016. Macro and micro models for zonal crash prediction with application in hot zones identification. Journal of Transport Geography 54, 248–256.

Huang, Y., Wang, X., Patton, D., 2018. Examining spatial relationships between crashes and the built environment: A geographically weighted regression approach. Journal of Transport Geography 69, 221–233.

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. Accident Analysis and Prevention 147, 105759.

Islam, A.S.M.M., Shirazi, M., Lord, D., 2023a. Grouped Random Parameters Negative Binomial-Lindley for accounting unobserved heterogeneity in crash data with preponderant zero observations. Analytic Methods in Accident Research 37, 100255.

Islam, S.M., Washington, S., Kim, J., Haque, M.M., 2023b. A hierarchical multinomial logit model to examine the effects of signal strategies on right-turn crash injury severity at signalised intersections. Accident Analysis and Prevention 188, 107091.

Ivan, J.N., Zhao, S., Wang, K., Hossain, J., Eluru, N., Abdel-Aty, M., Bhowmik, T., Parvez, D., Hoover, L., 2023. Development and Application of Crash Severity Models for Highway Safety: User Guidelines. NCHRP Report 1047, Transportation Research Board, Washington, D.C.

Jiang, X., Huang, B., Zaretzki, R.L., Richards, S., Yan, X., Zhang, H., 2013. Investigating the influence of curbs on single-vehicle crash injury severity utilizing zero-inflated ordered probit models. Accident Analysis and Prevention 57, 55–66.

Kabli, A., Bhowmik, T., Eluru, N., 2023. Exploring the temporal variability of the factors affecting driver injury severity by body region employing a hybrid econometric approach. Analytic Methods in Accident Research 37, 100246.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transportation Research Part A 44(5), 291–305.

Mannering, F., 2018. Temporal instability and the analysis of highway accident data. Analytic Methods in Accident Research 17, 1–13.

Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Analytic Methods in Accident Research 11, 1–16.

Marcoux, R., Yasmin, S., Eluru, N., Rahman, M., 2018. Evaluating temporal variability of exogenous variable impacts over 25 years: An application of scaled generalized ordered logit model for driver injury severity. Analytic Methods in Accident Research 20, 15–29.

Neyens, D.M., Boyle, L.N., 2007. The effect of distractions on the crash types of teenage drivers. Accident Analysis and Prevention 39(1), 206–212.

Paleti, R., Eluru, N., Bhat, C.R., 2010. Examining the influence of aggressive driving behavior on driver injury severity in traffic crashes. Accident Analysis and Prevention 42(6), 1839–1854.

Pervaz, S., Bhowmik, T., Eluru, N., 2023. An econometric framework for integrating aggregate and disaggregate level crash analysis. Analytic Methods in Accident Research 39, 100280.

Pervaz, S., Bhowmik, T., Eluru, N., 2022. Integrating macro and micro level crash frequency models considering spatial heterogeneity and random effects. Analytic Methods in Accident Research 36, 100238.

Razi-Ardakani, H., Mahmoudzadeh, A., Kermanshah, M., 2018. A Nested Logit analysis of the influence of distraction on types of vehicle crashes. European Transport Research Review 10(2), 1–14.

Savolainen, P.T., Mannering, F.L., Lord, D., Quddus, M.A., 2011. The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. Accident Analysis and Prevention 43(5), 1666–1676.

Stamatiadis, N., Pigman, J.G., Sacksteder, J., Ruff, W., Lord, D., 2009. Impact of Shoulder Width and Median Width on Safety. NCHRP Report 633, Transportation Research Board, Washington, D.C.

Uddin, M., Huynh, N., 2020. Injury severity analysis of truck-involved crashes under different weather conditions. Accident Analysis and Prevention 141, 105529.

Weiss, H.B., Kaplan, S., Prato, C.G., 2014. Analysis of factors associated with injury severity in crashes involving young New Zealand drivers. Accident Analysis and Prevention 65, 142–155.

Xin, C., Guo, R., Wang, Z., Lu, Q., Lin, P.-S., 2017. The effects of neighborhood characteristics and the built environment on pedestrian injury severity: A random parameters generalized ordered probability model with heterogeneity in means and variances. Analytic Methods in Accident Research 16, 117–132.

Xiong, Y., Mannering, F.L., 2013. The heterogeneous effects of guardian supervision on adolescent driver-injury severities: A finite-mixture random-parameters approach. Transportation Research Part B 49, 39–54.

Yan, Y., Zhang, Y., Yang, X., Hu, J., Tang, J., Guo, Z., 2020. Crash prediction based on random effect negative binomial model considering data heterogeneity. Physica A 547, 123858.

Yasmin, S., Eluru, N., 2018. A joint econometric framework for modeling crash counts by severity. Transportmetrica A 14(3), 230–255.
Yasmin, S., Eluru, N., 2016. Latent segmentation based count models: Analysis of bicycle safety in Montreal and Toronto. Accident Analysis and Prevention 95, 157–171.
Yasmin, S., Eluru, N., 2013. Evaluating alternate discrete outcome frameworks for modeling crash injury severity. Accident Analysis and Prevention 59, 506–521.
Yasmin, S., Eluru, N., Bhat, C.R., Tay, R., 2014. A latent segmentation based generalized ordered logit model to examine factors influencing driver injury severity. Analytic Methods in Accident Research 1, 23–38.
Yasmin, S., Momtaz, S.U., Nashad, T., Eluru, N., 2018. A Multivariate Copula-Based Macro-Level Crash Count Model. Transportation Research Record 2672(30), 64–75.
Yu, R., Wang, Y., Quddus, M., Li, J., 2019. A marginalized random effects hurdle negative binomial model for analyzing refined-scale crash frequency data. Analytic Methods in Accident Research 22, 100092.

APPENDIX

Table A.1: Results of the Disaggregate Level Models (MNL and Pooled OP) with Net Variable Effects

Variables	MNL Model Propensity (Base: Rear End)				OP Model Propensity					
	Angular		Sideswipe		Rear End		Angular		Sideswipe	
	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat
Constant (2, --) *	-2.448	-36.557	-0.733	-13.750	--	--	--	--	--	--
Threshold Parameters										
Threshold between NI-PI (--, 3)	--	--	--	--	0.671	29.019	0.372	9.125	1.299	36.120
Threshold between PI-NII (--, 3)	--	--	--	--	1.604	3.130	1.103	9.482	1.937	7.281
Threshold between NII-FII (--,1)	--	--	--	--	2.548	1.827	2.047	1.827	2.882	1.827
Driver and Vehicle Factors										
DUI related (if yes 1, otherwise 0) (1, 1)	--	--	-0.689	-2.512	0.525	4.419	0.525	4.419	--	--
Distraction related (if yes 1, otherwise 0) (2, 1)	0.190	3.050	-0.544	-8.880	0.238	8.953	0.238	8.953	0.238	8.953
With passengers (if yes 1, otherwise 0) (1, 1)	0.133	2.648	--	--	0.336	14.847	0.336	14.847	0.336	14.847
Roadway Characteristics										
Posted speed limit (Base: 40-55mph)										
Posted speed limit<40 (2, 1)	0.199	3.750	0.181	3.982	-0.082	-3.574	-0.082	-3.574	-0.082	-3.574
Posted speed limit>55 (1, --)	-0.314	-2.865	--	--	--	--	--	--	--	--
At intersection (if yes 1, otherwise 0) (1, --)	2.717	53.507	--	--	--	--	--	--	--	--
Curb shoulder (if yes 1, otherwise 0) (1, 1)	0.504	9.985	--	--	--	--	-0.133	-3.063	--	--
Wet surface (if yes 1, otherwise 0) (2, 1)	-0.345	-4.348	-0.341	-4.791	--	--	--	--	0.307	2.237
Road Environmental and Weather Factors										
Time (Base: Off-peak evening, late evening)										
Late night (12:00-6:30) (1, 1)	--	--	0.211	2.410	0.232	4.629	0.232	4.629	--	--
Peak morning (6:30-9:00) (1, --)	--	--	-0.337	-4.832	--	--	--	--	--	--
Off-peak morning (9:00-12:00) (--, 1)	--	--	--	--	--	--	0.139	2.228	--	--
Peak evening (16:00-18:30) (1, 1)	--	--	-0.220	-3.785	-0.086	-2.460	--	--	--	--
Peak morning and peak evening (1, --)	-0.163	-3.009	--	--	--	--	--	--	--	--
Season (Base: Summer)										
Winter (1, --)	--	--	-0.163	-2.666	--	--	--	--	--	--
Spring (1, --)	--	--	-0.120	-1.990	--	--	--	--	--	--
Autumn (1, --)	--	--	-0.182	-2.953	--	--	--	--	--	--
Light condition (Base: Day light, dawn/dusk)										
Dark lighted (1, --)	--	--	0.160	2.761	--	--	--	--	--	--
Dark not lighted (1, --)	0.505	2.789	--	--	--	--	--	--	--	--
Weather condition (Base: Clear)										

Variables	MNL Model Propensity (Base: Rear End)				OP Model Propensity					
	Angular		Sideswipe		Rear End		Angular		Sideswipe	
	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat
Rainy (--, 1)	--	--	--	--	--	--	--	--	-0.455	-2.421
Fog and wind (--, 1)	--	--	--	--	--	--	-2.946	-9.928	--	--
LL	-11,142.009				-10,289.086					
BIC	22,492.842				20,749.027					
Number of parameters	22				18					

*Note: "--" denotes variables are not significant at 90% confidence interval; *Numbers in the parenthesis denote the number of parameters estimated; FII=Fatal and incapacitating injury, NII=Non-incapacitating injury, PI= Possible injury, and NI= No injury.*

Table A.2: Result of the Panel Mixed NB-OPFS Model with Net Variable Effects

Variables	Count Component						Severity Proportion Component					
	Rear End		Angular		Sideswipe		Rear End		Angular		Sideswipe	
	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat
Constant (3, --) *	-0.097	-0.192	-1.094	-2.142	-1.457	-2.909	--	--	--	--	--	--
Threshold Parameter												
Threshold between P_NI-P_PI (--, 2)	--	--	--	--	--	--	0.152	2.646	0.152	2.646	1.166	8.365
Threshold between P_PI-P_NII (--, 3)	--	--	--	--	--	--	1.035	3.409	0.942	4.716	1.811	4.733
Threshold between P_NII-P_FII (--, 1)	--	--	--	--	--	--	1.903	2.433	1.811	2.433	2.679	2.433
Roadway and Traffic Characteristics												
AADT (1, --)	0.244	5.101	0.244	5.101	0.244	5.101	--	--	--	--	--	--
Proportion of interstate-expressways (--, 1)	--	--	--	--	--	--	0.279	2.597	--	--	--	--
Proportion of arterial road (--, 1)	--	--	--	--	--	--	0.162	1.980	--	--	--	--
Proportion of local road (--, 1)	--	--	--	--	--	--	--	--	--	--	-0.770	-2.031
Avg. no. of lanes (--, 1)	--	--	--	--	--	--	-0.139	-3.213	--	--	--	--
Avg. inside shoulder width (1, 1)	0.055	5.110	--	--	0.055	5.110	--	--	-0.027	-3.832	-0.027	-3.832
Avg. outside shoulder width (1, 1)	-0.066	-2.155	--	--	--	--	--	--	--	--	0.044	1.763
Proportion of road <40 mph (1, 1)	-0.275	-1.955	--	--	--	--	--	--	--	--	-0.242	-1.944
Proportion of road >55mph (1, --)	--	--	-0.434	-2.000	--	--	--	--	--	--	--	--
Proportion of divided road (1, --)	0.408	3.990	0.408	3.990	0.408	3.990	--	--	--	--	--	--
Intersection density (1, 1)	--	--	1.012	2.518	1.012	2.518	--	--	-1.224	-4.201	--	--
Avg. bike lane length (1, --)	--	--	-0.035	-2.165	--	--	--	--	--	--	--	--
Land-use Attributes												
Proportion of institutional area (1, --)	--	--	--	--	1.080	1.997	--	--	--	--	--	--
Proportion of commercial area (--, 1)	--	--	--	--	--	--	-0.157	-1.903	--	--	--	--
Land-use mix (1, --)	0.550	2.992	--	--	--	--	--	--	--	--	--	--
Built Environment Attributes												
No. of education centers (1, --)	0.083	2.147	--	--	--	--	--	--	--	--	--	--
No. of restaurants (1, 1)	0.256	7.420	0.256	7.420	0.256	7.420	--	--	-0.076	-2.841	--	--
No. of shopping centers (1, 1)	0.132	2.914	--	--	0.132	2.914	--	--	--	--	-0.070	-1.823
Sociodemographic Factors												
NMT transport (1, --)	0.133	5.510	0.133	5.510	0.133	5.510	--	--	--	--	--	--
Proportion of White-American population (--, 1)	--	--	--	--	--	--	-0.395	-3.493	--	--	--	--
Proportion of African American population (1, 1)	--	--	0.698	3.958	--	--	--	--	0.330	3.418	0.330	3.418
Over dispersion (3, --)	0.464	6.102	0.412	8.674	0.563	8.566	--	--	--	--	--	--

Variables	Count Component						Severity Proportion Component					
	Rear End		Angular		Sideswipe		Rear End		Angular		Sideswipe	
	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat
LL = -3,239.540												
BIC = 6,744.675												
Number of parameters = 40												
Non-Integrated Models												
Total LL = -24,670.635												
Total BIC = 49,872.460												
Total number of parameters = 80												

*Note: "--" denotes variables are not significant at 90% confidence interval; *Numbers in the parenthesis denote the number of parameters estimated; FII=Fatal and incapacitating injury, NII=Non-incapacitating injury, PI= Possible injury, and NI= No injury.*