A JOINT ECONOMETRIC APPROACH FOR MODELING CRASH COUNTS BY COLLISION TYPE

Tanmoy Bhowmik
PhD Student
Department of Civil, Environmental & Construction Engineering
University of Central Florida
Tel: 1-407-927-6574; Fax: 1-407-823-3315
Email: tanmoy78@knights.ucf.edu

Shamsunnahar Yasmin*
Postdoctoral Associate and Graduate Faculty Scholar
Department of Civil, Environmental & Construction Engineering
University of Central Florida
Tel: 407-823-4815, Fax: 407-823-3315
Email: shamsunnahar.yasmin@ucf.edu

Naveen Eluru
Associate Professor
Department of Civil, Environmental & Construction Engineering
University of Central Florida
Tel: 407-823-4815, Fax: 407-823-3315
Email: naveen.eluru@ucf.edu

*Corresponding author

ABSTRACT

In recent years, there is growing recognition that common unobserved factors that influence crash frequency by one attribute level are also likely to influence crash frequency by other attribute levels. The most common approach employed to address the potential unobserved heterogeneity in safety literature is the development of multivariate crash frequency models. The current study proposes an alternative joint econometric framework to accommodate for the presence of unobserved heterogeneity – referred to as joint negative binomial-multinomial logit fractional split (NB-MNLFS) model. Furthermore, the study undertakes a first of its kind comparison exercise between the most commonly used multivariate model (multivariate random parameter negative binomial model) and the proposed joint approach by generating an equivalent log-likelihood measure. The empirical analysis is based on the zonal level crash count data for different collision types from the state of Florida for the year 2015. The model results highlight the presence of common unobserved effects affecting the two components of the joint model as well as the presence of parameter heterogeneity. The equivalent log-likelihood and goodness of fit measures clearly highlight the comparable performance offered by the proposed joint model relative to the commonly used multivariate approach. Overall, the model interpretations and fit measures clearly highlight the potential complementary role of the proposed approach for crash frequency analyses.

Keywords: *Joint system; Multinomial fractional split model; Crash types; Equivalent log-likelihood; Comparison exercise; and Unobserved effect.*

1 INTRODUCTION

Road traffic crashes are responsible for nearly 1.25 million fatalities every year and are a leading cause of death among people aged between 15 and 29 years old (World Health Organization, 2015). The extent of societal, emotional and economic impacts of these unfortunate events has warranted coordinated multi-sectoral responses from the fields of transportation, public health, and medicine. A major analytical tool employed for examining the critical factors influencing crash occurrence include the econometric crash prediction/ crash frequency models. These models examine crashes at the micro-level (such as an intersection or roadway segment) or at the macro-level (such as a county or Traffic Analysis Zone (TAZ)). The various crash frequency dimensions frequently explored in existing literature include total crashes, crashes by severity, crashes by collision type and crashes by vehicle type for a spatial unit over a given time period.

A majority of the existing studies in safety literature developed crash frequency models for a single dependent variable; the methods are referred to as univariate modeling approaches (see Lord and Mannering, 2010; Yasmin and Eluru, 2018 for a detailed review of these studies). In recent years, there is growing recognition that univariate approaches, while adequate for analyzing a single dependent variable, fall short in modeling multiple crash frequency variables for a single observational unit. For example, the total number of crashes in a TAZ are a sum of crashes by different collision types (or severity levels) *i.e.* as opposed to analyzing a single total crash variable it is possible to examine crash frequency by different attribute categories. In this case, an extension of univariate approach would be to develop multiple univariate models with frequency by attribute levels considered as multiple dependent variables. Through this approach, the exogenous variables affecting crash counts can exhibit distinct impacts on different attribute levels allowing for a flexible specification. The separate models for crash frequency by attribute level allows us to capture realistic estimates of exogenous variables. Yet, the approach only accommodates for observed factors and inherently neglects the information that the multiple crash frequency variables for a TAZ are potentially correlated. For example, for zonal level crash frequency analysis, it is possible that several characteristics specific to the zone such as driver behavior, geometric design and build quality (possibly of higher or lower quality relative to the other zones) and traffic signal design objectives might influence different crash counts by collision type (such as head-on, rear-end). These factors that influence crash frequency by one attribute level are also likely to influence crash frequency by other attribute levels. Such detailed characteristics are rarely available to analysts for consideration in model development. Ignoring for the presence of such unobserved heterogeneity in model development will result in inaccurate and biased model estimates (see Mannering et al., 2016 for an extensive discussion).

The most common approach employed to address the potential unobserved heterogeneity in safety literature is the development of multivariate crash frequency models. In this approach, the impact of exogenous variables is quantified through the propensity component of count models. The main interaction across different count variables is sought through unobserved effects *i.e.* there is no interaction of observed effects across the multiple count models. These approaches, in general, partition the error components of the dependent variables to accommodate for a common term and an independent term across dependent variables (see Mannering et al., 2016 for a detailed discussion of various methodologies). In our current study, we develop an alternative approach to accommodate for the presence of observed and unobserved heterogeneity. The approach employs a joint crash frequency and multinomial fractional split model to provide an alternative to the multivariate count models in extant literature. The approach builds on recent work by Yasmin and colleagues in multiple studies (Lee et al., 2018; Yasmin et al., 2016; Yasmin and Eluru, 2018). Furthermore, the current study undertakes a first of its kind comparison exercise between the most commonly used multivariate count model (multivariate random parameter negative binomial model) and the proposed approach of the current study. The reader would note that the log-likelihood functions across these models are not directly comparable. Hence, to facilitate a comparison, an equivalent log-likelihood measure is also generated for the proposed joint crash frequency and fractional split model. Finally, an in-sample prediction exercise comparing the two systems is conducted. The models are estimated by using data from Florida at the Statewide Traffic Analysis Zone (STAZ) level for the year 2015.

The rest of the paper is organized as follows: Section 2 provides a brief review of previous relevant research. Section 3 and 4 provide a description of the modelling approach and data, respectively. Model estimation results and comparison results are presented in Section 5. Finally, a summary of model findings and conclusions are presented in Section 6.

2 BACKGROUND AND CURRENT STUDY IN CONTEXT

2.1 Earlier Research

Earlier research efforts in safety literature have focused on univariate model systems for crash frequency analysis. Majority of these studies focus on crash frequency by vehicle involvement (Ivan et al., 2000; Persaud and Mucsi, 1995; Qin et al., 2004; Zhou and Sisiopiku, 1997) or crash type (Chai and Wong, 2014; Li et al., 2016; Wang and Abdel-Aty, 2008a, 2008b, 2006; Yan et al., 2005). It is beyond the scope of our paper to review the vast literature of univariate models (please see Lord and Mannering, 2010; Yasmin and Eluru, 2018 for a literature review).

Recently, research in safety literature has shifted toward modeling multiple dependent variables for each observation unit. The most common approach for modeling multiple dependent variables such as crash frequency by severity or collision type is based on using a multivariate crash frequency model. In these models, every crash frequency variable is associated with its corresponding propensity equation (similar to univariate system). Thus, we allow for the impact of exogenous variables to vary across crash frequency variables. For example, consider the exogenous variable - presence of left guardrail on the roadway. In the presence of a left guardrail, vehicles are prevented from entering the opposite direction thus reducing head-on crashes. On the other hand, vehicles on hitting the guardrail might collide with other vehicles travelling in the same direction. Thus, the overall impact of the guardrail might be an increase in total crashes with distinct effects on head-on and sideswipe crashes. So, considering the guardrail variable in the total crash would yield a positive sign. However, considering the same variable in separate univariate models for head-on collisions and sideswipe collisions offer different results. This is an example of how observed variables exhibit contrasting effects on crash occurrence by collision[1] type. Thus, developing separate models for frequency by collision type allows us to capture realistic estimates of exogenous variables.

In addition to observed factors, the multivariate models inherently account for correlation across multiple crash frequency variables for an observation unit. Ignoring for the presence of such unobserved heterogeneity (associated with missing information or inherently unobservable phenomenon affecting crashes) in model development will result in inaccurate and biased model estimates (see Mannering et al., 2016 for an extensive discussion). In these multivariate models, typically probability computation requires integrating the probability function over the error term distribution. The exact computation is dependent on the distributional assumption and does not have a closed form expression usually[2]. Several studies

[1] We use crash and collision synonymously in the current study context.

[2] In some cases, a parametric multivariate distributional assumption might result in closed form approaches such as the copula based approach (Nashad et al., 2016; Yasmin et al., 2018)

recognizing the importance of unobserved heterogeneity have developed multivariate approaches that account for the potential dependency across count variables. The various model structures developed from multivariate models include multivariate Poisson regression model (Ye et al., 2009), multivariate Poisson lognormal model (Serhiyenko et al., 2016), multinomial-generalized Poisson model (Chiou and Fu, 2013), multivariate Poisson gamma mixture count model (Mothafer et al., 2016), multivariate Poisson lognormal spatial and/or temporal model (Aguero-Valverde et al., 2016; (Amoh-Gyimah et al., 2017; Barua et al., 2016, 2015, 2014; Cheng et al., 2017; Liu and Sharma, 2018, 2017; Ma et al., 2017; Mohammadi et al., 2014; Zeng et al., 2018), grouped random parameter multivariate spatial model (Cai et al., 2018), Integrated Nested Laplace Approximation Multivariate Poisson Lognormal model (Wang et al., 2017), Bayesian latent class flexible mixture multivariate model (Heydari et al., 2017), flexible Bayesian semiparametric approach (Heydari el al., 2016) and multivariate random-parameters zero-inflated negative binomial model (Anastasopoulos, 2016)[3].

An alternative approach - referred to as the fractional split approach - for modeling crash frequency by attribute level is recently being applied in safety literature (Eluru et al., 2013; Papke and Wooldridge, 1996). In a fractional split approach, as opposed to modeling the count events, count proportions by different attributes (such as injury severity, collision type or vehicle type) for a study unit are examined. The fractional split approach directly relates a single exogenous variable to count proportions of all attribute levels simultaneously. Thus, in this model, exogenous variables affect attribute proportions allowing us to obtain a parsimonious specification. This is in contrast to the multivariate crash models where the observed variables in count propensity equations do not interact with other count variables in the model system.

In safety literature, very few studies have employed the fractional split approach. Milton et al. (2008) developed a mixed multinomial fractional split model to study injury-severity distribution of crashes on highway segments by using highway-injury data from Washington State. A number of studies have also examined crash frequency by severity simultaneously by building on multinomial-Poisson transformation (Chiou and Fu, 2013, 2015; Chiou et al., 2014). Geedipally et al. (2010) developed independent crash count (negative binomial model) and crash proportion models (multinomial fractional split) to investigate whether the model system can be used for the estimation of crash counts for each collision type. The study concluded that their approach offered good results. However, the study ignored the influence of common unobserved factors between the crash model and the proportion model. Yasmin et al. (2016) developed an ordered outcome fractional split model that allows the analysis of proportion for variables with multiple alternatives. The approach is applicable only for crash proportions that are ordered. A particularly relevant research effort, Yasmin and Eluru, (2018) extended the ordered proportional framework to incorporate crash frequency (as a negative binomial model) along with crash proportion by injury severity (as an ordered fractional split model).

2.2 Current Study

The literature review clearly highlights the prevalence of multivariate model frameworks in safety literature. An alternative approach – fractional split model is emerging as a promising alternative framework for multivariate counts. However, so far there has not been a comprehensive comparison exercise between these two systems. In this context, the current study makes three methodological contributions. *First*, we develop the first joint system for total crash counts and multinomial fractional split model. Specifically, we propose to estimate

[3] For recent advances in accommodating unobserved heterogeneity in disaggregate level injury severity models the reader is referred to Eluru et al., 2012; Yasmin et al,, 2014; Fountas et al., 2018a; Fountas et al., 2018b.

a joint negative binomial-multinomial logit fractional split (NB-MNLFS) model. The work builds on Yasmin and colleagues' recent work in the ordered and unordered fractional split realm. The approach employed in Yasmin and Eluru (2018) is applicable only for crash proportions that are ordered. In our study, the fractional variable of interest, proportions by crash type, is not an ordered variable and requires an entirely different formulation. Within the joint framework, we also accommodate for random parameters in the count and fractional split components. *Second*, the data fit measures of multivariate count model and the proposed joint system are not directly comparable because of the differences in estimation techniques for the two approaches. In the current study, we propose an equivalent log-likelihood measure for the proposed joint NB-MNLFS system to evaluate comparable data fit metrics. *Third*, we undertake a comprehensive comparison exercise between the most commonly employed multivariate model and its fractional counterpart. Specifically, we examine performance in model estimation and prediction for random parameter multivariate negative binomial model (RPMNB) that accommodates unobserved heterogeneity and the proposed joint model. Empirically, the study develops crash frequency by collision type. The models are estimated at a macro-level using STAZ level crash data for the year 2015 for the state of Florida. The outcomes of macro-level crash count models can be used to devise safety-conscious decision support tools to facilitate proactive approach in assessing medium and long-term policy-based countermeasures. Moreover, the tool plays an important role in safety implications of land use planning initiatives and alternative network-planning initiatives. The proposed model results offer insights on important variables affecting crash frequency, as well as crash proportion by collision type.

3 METHODOLOGY

In this section, we briefly provide details of the model frameworks employed in our study. Let us assume that i $(i = 1,2,3, \ldots, N)$ be the index for STAZ. Let l be the index representing different crash count level, where j $(j = 1,2,3, \ldots, J;\ j \in K\ and\ K = \sum_{j=1}^{J} j)$ be the index to represent different collision types and K represents total crashes at a zonal level. In this empirical study, the index j may take the values of rear-end (j =1), head-on (j =2), angular (J =3), off-road (j =4), other single vehicle (j = 5), other multiple vehicles (j =6), rollover (j =7) and sideswipe (j =8) crashes. Using these notations, the equation system for modeling crash count across different crash count level l, (l can denote either total crashes or crash counts by different collision types) in the usual NB formulation can be written as:

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

where, c_{il} be the index for crash counts specific to level l occurring over a period of time in STAZ i. $P(c_{il})$ is the probability that STAZ i has c_{il} number of crashes for crash count level l. $\Gamma(\cdot)$ is the gamma function, α_l is NB over dispersion parameter and μ_{il} is the expected number of crashes occurring in STAZ i over a given time period for crash count level l. Given this set up, the mathematical formulations of the econometric frameworks considered in the current study context is presented in this section.

3.1 Multivariate NB Model

The focus of multivariate NB model is to examine number of crashes across different collision types jointly. In our current study context, we consider eight different collision types (rear-end, head-on, angular, off-road, others single vehicle, others multiple vehicles, rollover and sideswipe crashes). Thus, in estimating multivariate NB model, we examine eight different NB models for eight different collision types simultaneously. For the multivariate approach, the equation system for modeling crash count across different collision types can be written by replacing the subscript l with j in equation 1. Thus, the probability for crash occurrence for different crash type j can be represented as $P(c_{ij})$, for which we can express μ_{ij} as a function of explanatory variables by using a log-link function as follows:

$$\mu_{ij} = E(c_{ij}|\boldsymbol{z}_{ij}) = exp\big((\boldsymbol{\delta}_j + \boldsymbol{\zeta}_{ij})\boldsymbol{z}_{ij} + \ln(Area_i) + \varepsilon_{ij} + \eta_{ij}\big) \quad (2)$$

where, $\boldsymbol{z}_{ij}$ is a vector of explanatory variables associated with STAZ i and collision type j. $Area_i$ is the STAZ area used as an offset variable in the NB model specification[4]. $\boldsymbol{\delta}_j$ is a vector of coefficients to be estimated. $\boldsymbol{\zeta}_{ij}$ is a vector of unobserved factors on crash count propensity associated with collision type j for STAZ i and its associated zonal characteristics, assumed to be a realization from standard normal distribution: $\boldsymbol{\zeta}_{ij} \sim N(0, \boldsymbol{\pi}_j{}^2)$. ε_{ij} is a gamma distributed error term with mean 1 and variance α_j. η_{ij} captures unobserved factors that simultaneously impact number of crashes across different collision types for STAZ i. Here, it is important to note that the unobserved heterogeneity between total number of crashes across different collision types can vary across STAZs. Therefore, in the current study, the correlation parameter η_i is parameterized as a function of observed attributes as follows:

$$\eta_{ij} = \boldsymbol{\gamma_j} \boldsymbol{s}_{ij} \quad (3)$$

where, $\boldsymbol{s}_{ij}$ is a vector of exogenous variables, $\boldsymbol{\gamma_j}$ is a vector of unknown parameters to be estimated (including a constant). In the current analysis, the multivariate NB model only allows for a positive correlation for total number of crashes across different collision types.

In examining the model structure of crash count across different collision types, it is necessary to specify the structure for the unobserved vectors $\boldsymbol{\zeta}$ and $\boldsymbol{\gamma}$ represented by $\boldsymbol{\Omega}$. In this paper, it is assumed that these elements are drawn from independent normal distributions: $\Omega \sim N(0, (\boldsymbol{\pi}_j{}^{\boldsymbol{2}}, \boldsymbol{\sigma}_j^2))$. Thus, conditional on $\boldsymbol{\Omega}$, the likelihood function for the joint probability can be expressed as:

$$L_i = \int_{\boldsymbol{\Omega}} \prod_{j=1}^{J} \big(P(c_{ij})\big) f(\boldsymbol{\Omega}) d\boldsymbol{\Omega} \quad (4)$$

Finally, the log-likelihood function is:

$$LL = \sum_i Ln(L_i) \quad (5)$$

[4] STAZ areas under consideration vary from 10^{-7} mile^2 to 885.321 mile^2 with a mean of 6.472 mile^2. Given the wide range in STAZ areas, we allow the area associated with STAZs as an offset variable to account for different sizes of STAZs in our model specification. The coefficient of the offset variable is restricted to be one in estimating the model to normalize for the number crash events by STAZ area.

All the parameters in the model are estimated by maximizing the logarithmic function LL presented in equation 5. The parameters to be estimated in the multivariate NB model are: $\boldsymbol{\delta}_j$, α_j, $\boldsymbol{\pi}_j$, and $\boldsymbol{\sigma_j}$.

3.2 Joint NB-MNL Fractional Split Model (NB-MNLFS)

The focus of joint NB-MNL fractional split model is to jointly model "total number of crashes" and "proportion of crashes by crash types". Thus, we examine one NB model for total crash count and one MNL fractional split model for crash proportion by crash types simultaneously. For the joint approach, the equation system for modeling total crash count in the usual NB formulation can be written by replacing subscript l by K in equation 1. Thus, the probability for crash occurrence for total crash count K can be represented as $P(c_{iK})$, for which we can express μ_{iK} as a function of explanatory variables by using a log-link function as follows:

$$\mu_{iK} = E(c_{iK}|\boldsymbol{x}_i) = exp\big((\boldsymbol{\theta} + \boldsymbol{\varrho}_i)\boldsymbol{x}_i + \ln(Area_i) + \phi_i + \psi_{ij}\big) \tag{6}$$

where, $\boldsymbol{x}_i$ is a vector of explanatory variables associated with STAZ i. $Area_i$ is the STAZ area used as an offset variable in the NB model specification. $\boldsymbol{\theta}$ is a vector of coefficients to be estimated. $\boldsymbol{\varrho}_i$ is a vector of unobserved factors on crash count propensity for STAZ i and its associated zonal characteristics assumed to be a realization from standard normal distribution: $\boldsymbol{\varrho}_i \sim N(0, \boldsymbol{\varsigma}^2)$. ϕ_i is a gamma distributed error term with mean 1 and variance α_K. ψ_{ij} captures unobserved factors that simultaneously impact total number of crashes and proportion of crashes by crash types for STAZ i.

In the joint model framework, the modeling of crash proportions by crash types is undertaken using the MNL fractional split model. In our current study, the dependent variable in the crash proportion component of the joint model is defined as the proportion of crash type in traffic crashes by STAZ. In estimating the model, we assume that the sum of the proportions across a STAZ is equal to unity and each proportion of crash types in traffic crashes ranges between zero and one. Let y_{ij} be the fraction of crashes by crash type j $(j = 8)$ in STAZ i.

$$0 \leq y_{ij} \leq 1, \qquad \sum_{j=1}^{J} y_{ij} = 1 \tag{7}$$

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

$$\begin{gathered} E[y_{ij}|d_{ij}] = G_j(\cdot) \\ 0 < G_j(\cdot) < 1, \ \sum_{j=1}^{J} G_j(\cdot) = 1 \end{gathered} \tag{8}$$

where $G_j(\cdot)$ is a predetermined function. The properties specified in equation 8 for $G_j(\cdot)$ warrant that the predicted fractional crash types will range between 0 and 1, and will add up to 1 for each STAZ. In this study, a MNL functional form for G_j in the fractional split model of equation 8. Then equation 8 is rewritten as:

$$E(y_{ij}|d_{ij}) = G_j(\cdot) = \frac{\exp(\left(\boldsymbol{\beta_j} + \boldsymbol{\rho_{ij}}\right)d_{ij} + \xi_{ij} \pm \psi_{ij})}{\sum_{j=1}^{J}\exp(\left(\boldsymbol{\beta_j} + \boldsymbol{\rho_{ij}}\right)d_{ij} + \xi_{ij} \pm \psi_{ij})}, j = 1,2,3, \dots, \quad (9)$$

where, $\boldsymbol{d}_{ij}$ is a vector of attributes, $\boldsymbol{\beta}_j$ is the corresponding vector of coefficients to be estimated for crash type j. $\boldsymbol{\rho}_{ij}$ is a vector of unobserved factors assumed to be a realization from standard normal distribution: $\boldsymbol{\rho} \sim N(0, \boldsymbol{\nu}_j{}^2)$. ξ_{ij} is the random component assumed to follow a Gumbel type 1 distribution. ψ_{ij} term generates the correlation between equations for total number of crashes and crash proportions by crash types. The $\pm$ sign in front of ψ_{ij} in equation 9 indicates that the correlation in unobserved zonal factors between total crashes and crash proportions by crash type may be positive or negative. A positive sign implies that STAZs with higher number of crashes are intrinsically more likely to incur higher proportions for the corresponding crash types. On the other hand, negative sign implies that STAZs with higher number of crashes intrinsically incur lower proportions for different crash types. To determine the appropriate sign, we 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 total number of crashes and crash proportions by crash types can vary across STAZs. Therefore, in the current study, the correlation parameter ψ_{ij} is parameterized as a function of observed attributes as follows:

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

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

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

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

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

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

All the parameters in the model are estimated by maximizing the logarithmic function $\mathcal{LL}$ presented in equation 12. The parameters to be estimated in the joint model are: $\boldsymbol{\theta}, \alpha_K, \boldsymbol{\beta_j}, \boldsymbol{\nu}_j$ and $\boldsymbol{\aleph_j}$.

To estimate the proposed joint and multivariate models, we apply Quasi-Monte Carlo simulation techniques based on the scrambled Halton sequence to approximate this integral in the likelihood function and maximize the logarithm of the resulting simulated likelihood function across individuals (see Bhat, 2001; Eluru et al., 2008; Yasmin and Eluru, 2013 for examples of Quasi-Monte Carlo approaches in literature). The model estimation routine is coded in GAUSS Matrix Programming software (Aptech, 2015)

4 DATA PREPARATION

The study draws motorized crash record data from the state of Florida for the year 2015 at STAZ level from Florida Department of Transportation (FDOT), Crash Analysis Reporting System (CARS) and Signal Four Analytics (S4A) databases. The data provides crash information for 8,518 STAZs. The data reports 10 types of collisions: rear-end, head-on, angular, left-turn, right-turn, off-road, rollover, sideswipe, other collision type with one vehicle involved and other collision type with more than one vehicle involved. Based on crash records the angular, left-turn and right-turn collision types are combined as one category; thus 8 collision type categories are considered. Table 1 represents the summary statistics of crash type variables. A total of 487,171 motorized crashes were recorded in Florida during 2015. Of these crashes, rear-end collisions are the most prevalent while rollover crashes are less frequent with 1.06% among all other crash types.

4.1 Variables Considered

Roadway characteristics, land use attributes and traffic characteristics - three broad categories of explanatory variables are considered in our study. The data employed are obtained from FDOT Transportation Statistics Division, and US Census Bureau. The attributes are then aggregated at a STAZ level using geographical information system (GIS). Roadway attributes included are road lengths for different functional class, access and pavement condition, on road, off-road, divided road and roads with different number of lanes (1, 2 and 3 or more), width and variance of median, intersection and signal density, mean and variance of posted speed limit, average width of the sidewalk, inside and outside shoulder mean width. Intersection density denotes the number of intersection per miles of street in a STAZ and signal density is the number of signals per intersection. Land use attributes included area of urban, residential, industrial, institutional, recreational, office, agricultural land use types and land use mix[5]. Further, for traffic characteristics, average annual daily traffic (AADT), average annual daily truck traffic (truck AADT), vehicles miles traveled (VMT), truck vehicles miles traveled (truck VMT) and proportion of heavy traffic are considered.

Table 2 summarizes sample characteristics of the explanatory variables with the definition considered for final model estimation along with the zonal minimum, maximum and mean values. Several functional forms and specifications for different variables are explored.

---

[5] Land use mix is defined as: $\left[\frac{-\sum_k(p_k(lnp_k))}{lnN}\right]$, where k is the category of land-use, p_k is the proportion of the developed land area devoted to a specific land-use k, N is the number of land-use categories in a STAZ. In our study, six land use types were considered including residential, park facilities, industrial, institutional, agricultural and office areas. Institutional land use refers to land uses that cater to community's social and educational needs (schools, town hall, police station) while park facilities refer to land used for recreational or entertainment purposes. The value of this index ranges from zero to one - zero (no mix) corresponds to a homogenous area characterized by single land use type and one to a perfectly heterogeneous mix).

The final specification of the model development was based on removing the statistically insignificant variables in a systematic process based on 90% significance level.

5 EMPIRICAL ANALYSIS

5.1 Model Specification and Overall Measure of Fit

The empirical analysis involves estimation of four different models: 1) Independent NB-MNLFS model[6], 2) Joint NB-MNLFS model with correlation, 3) Independent Multivariate NB model[7], 4) Multivariate NB model with correlation. The log-likelihood values at convergence are estimated as follows: (1) Independent NB-MNLFS model (52 parameters) is -53858.29[8], (2) Joint NB-MNLFS model with correlation (55 parameters) is -53843.04, (3) Independent Multivariate NB model (116 parameters) is -163958.22 and (4) Multivariate NB model with correlation (119 parameters) is-160941.40. From the log-likelihood values we can see that the joint and multivariate models performed better than their respective independent models. The estimation results of the joint NB-MNLFS model results are discussed in detail. However, the estimation results of multivariate NB model are not discussed for the sake of brevity (the parameter estimates are presented in Table 4).

5.2 Behavioural Comparison

Prior to discussing the variable impacts for the NB-MNLFS model, we provide a behavioral comparison of the proposed approach with RPMNB model. For this comparison, let us consider the "proportion of arterial road" variable. In the NB-MNLFS model (Table 3), the presence of higher proportion of arterial roads contributes to higher propensity for total crashes. At the same time, the variable also increases the proportion of rear-end and sideswipe crashes while reducing off-road crashes. Now, it might appear on first glance that the proportion of arterial road does not affect other crash types. This is not true. The increase in the proportion of rear-end and side-swipe crash types are realized from reductions to off-road crashes as well as all other crash types that serve as base. So, if we were to impose a crash hierarchy across crash types for the proportion of arterial roads variable based on the NB-MNLFS model it would be: Side-swipe > Rear-end > Head-on = Angle = Other single vehicle = Other multiple vehicle = Rollover > Off-road. Based on the RPMNB model (Table 4), examining the coefficients for the arterial proportion variable, the hierarchy is: Side-swipe > Rear-end > Other multiple vehicle > Head-on > Angle > Other single vehicle > Off-road > Rollover. Given the coefficients in the RPMNB model are close for some crash types, there might not be any statistically significant difference. The comparison clearly highlights very little difference in hierarchy across the two systems. The estimation of this relationship required only 4 parameters in the NB-MNLFS model while 8 parameters were estimated in the RPMNB model[9]. The comparison illustrates how the different mathematical structure of the proposed model offers a parsimonious specification while not reducing the information resolution.

---

[6] The independent model for NB-MNLFS is estimated by restricting $\psi_{ij}=0$ in equation 9. Further, in the independent model, we do not capture any random terms in the propensity function (i.e. $\nu_j=0$).

[7] The independent multivariate model is estimated by restricting $\eta_i=0$ and $\zeta_{ij}=0$ in equation 2.

[8] The reader would note that the log-likelihood values for the fractional split models refers to the quasi log-likelihood values.

[9] To be sure, the RPMNB model could be estimated with restrictions on variable impacts. The process would require additional programming and is typically not possible with existing software.

5.3 NB-MNL Fractional Split Joint Model

Table 3 presents the model estimation results of the joint NB-MNL fractional split model with NB for the total crash component and MNL fractional split for the proportion of crashes by collision type. The second column provides the results of the NB component while columns 3 through 10 present the results of the MNL fractional split model. The model results are discussed separately for total crash component and proportion by collision type.

5.3.1 NB Component (Total Crash)

A positive (negative) sign for a variable in the crash count component of Table 3 indicates that an increase in the variable is likely to result in more (less) motor vehicle crashes. The reader would note that in crash frequency models, area of the STAZ is used as an offset.

Roadway Characteristics: The parameter estimates for proportion of arterial roads indicate that risk of motor vehicle crashes increases with increasing proportion of arterial roads in the STAZ. A similar result is observed for the proportion of urban roads (see Abdel-Aty and Radwan, 2000 for a similar result). Intersection density variable exhibits a positive impact on motorized crashes; an expected result because intersections are likely to increase potential vehicle conflicts due to the high number of turning movements (see Abdel-Aty et al., 2013, 2005). An increased proportion of roads with three or more lanes at the zonal level is found to be positively associated with motor vehicle crash incidence. The result is intuitive because higher number of lanes result in lane changing movements that could potentially lead to vehicle conflicts and crashes. The coefficient for mean posted speed limit in the zone has a negative coefficient indicating a reduction in crash incidence. The result is indicative of better roadway facility condition and design for high speed facilities (see Milton and Mannering, 1998 for a similar result). An increase in the length of divided roads in a STAZ reduces crash incidence. Divided roads reduce vehicle conflicts and are likely to reduce crash risk. The variables average inside and outside shoulder width offer contrasting effects. While an increase in average inside shoulder width is associated with higher crash risk, an increase in average outside shoulder width is likely to improve safety for motor vehicles. An increase in sidewalk width is associated with increased crash risk. The presence of increased sidewalk width is likely to contribute to more vehicle crashes. The result is potentially an indication of the presence of increased conflicts in the presence of sidewalks. But, the result warrants further investigation.

Land-Use Attributes: As expected, large proportion of urban area is associated with increased incidence of traffic crashes. The urban area proportion serves as a surrogate for exposure – urban areas attract larger amount of traffic and thus increase crash risk. On the other hand, land use mix variable highlights how zones exhibiting high mixture of residential, industrial, institutional and other areas are likely to reduce driving speeds and reduce motor vehicle crash risk. No significant impact was found for other land use attributes in the analysis.

Traffic Characteristics: The traffic volume variables representing AADT and Truck AADT offer expected results. With increase in AADT and truck AADT in the STAZ, the incidence of traffic crashes is likely to increase.

5.3.2 MNL Fractional Split Component

In the MNL fractional split model, one of the outcomes must be the base for every variable for the sake of identification. In our analysis, fraction of rear-end crashes is considered as the base alternative for model estimation. Therefore, a positive (negative) sign for a variable indicates that an increase in the variable is likely to result in higher proportion of crashes for the corresponding alternative relative to the rear-end fraction. For rear-end crash proportions, impact of some exogenous variables is estimated and, in those cases, other alternatives are considered as the base alternatives.

Roadway Characteristics: With higher proportion of arterial roads, the proportion of rear-end and sideswipe crashes increases while the proportion of off-road crashes decreases. The result is indicative of higher traffic and associated likelihood of traffic conflicts with vehicles in the same direction leading to rear-end and sideswipe crashes. With increasing proportion of urban roads, the results indicate a reduction in the proportion of off-road and rollover crashes. A positive association is observed for the intersection density variable in the proportion of angular, sideswipe and other multiple vehicle crash categories while a negative association is observed for off-road and rollover crash proportions. At intersections, there are complicated turning movements that result in more angular and sideswipe crashes rather than off-road and roll over crashes.

The parameter for proportion of 3 or more lane roads reveals a positive association with sideswipe crash proportion. As expected, average posted speed limit in a zone is positively associated with head-on, off-road and angular crash proportions (see Ye et al., 2009 for similar result for head-on crashes). The result is contrary to the impact of the variable in the total crash component. The length of divided road in a STAZ is found to negatively influence off-road and sideswipe crash proportions. The average width of inside shoulder variable indicates a negative impact on proportion of head-on crashes. The result is expected because increasing width of inside shoulder reduces the potential for head-on collisions. In terms of outside shoulder width variable, the influence is positive for off-road, other single vehicle and rollover crash proportions. Average width of sidewalk variable negatively affects the proportion of off-road and rollover crashes. As the average sidewalk width in a zone increases, it is indicative of the roadway infrastructure with clear demarcation of roadway facilities. In the presence of such clear demarcation, it is not surprising that the proportion of rollover and off-road crashes are likely to be lower.

Land-Use Attributes: In terms of land use attributes, only proportion of urban area variable has significant impact on crash proportions. The likelihood of rear-end crashes increases for a high percentage of urbanized area in a STAZ while off-road, rollover and other single vehicle crash proportion reduces. The result seems reasonable because in urbanized area, there is high density of slow moving vehicles with reduced gap and as a result more rear-end crashes are likely to occur.

Traffic Characteristics: The estimated AADT variable implies a positive effect on rear-end crash proportions and a negative effect on other single vehicle crash proportions. The result is intuitive as with higher number of vehicles, the likelihood of rear-end crashes increase. Truck AADT is negatively associated with off-road and rollover crash proportions. The result suggests that off-road and rollover crashes goes down with the increasing portion of trucks on the road.

5.3.3 Common Unobserved and Random Parameter

Second last row panel of Table 3 represent the significance of unobserved effects in the joint NB-MNLFS model. Two sets of effects are tested in our analysis: (1) the common unobserved factors jointly affecting total crashes and the fractional split model (corresponding to ψ_{ij} in the model system in Section 3.2) and (2) testing for the presence of parameter heterogeneity for different variables – also referred to as random parameters – (corresponding to $\boldsymbol{\zeta}_{ij}$ and $\boldsymbol{\varrho}_{i}$ in Section 3.2). Two common unobserved factors were found to be significant. The first parameter represents the common correlation between total crash and crash proportions of off-road, other-single vehicle and rollover collision types. The second parameter presents common factors affecting total crash and proportions of rear-end, head-on, sideswipe, angle and other-multiple vehicle collision types. As shown in Equation 9 of Section 3.2, the correlation between the two components could be either positive or negative. In our analysis, we found the negative sign to offer better fit for common correlation between total crash and crash proportions of off-road, other-single vehicle and rollover collision types while positive component offered better fit in terms of common effects affecting total crash and proportion of rear-end, head-on, sideswipe, angle and other-multiple vehicle collision types. Overall, the results clearly support our hypothesis that common unobserved factors influence the two components.

In terms of random parameters, one statistically significant parameter was obtained; proportion of urban area specific to total crash count component. The result highlights the variation of the influence of proportion of urban area for total crashes. The reader would note that the distributional parameters indicate that the overall impact is most likely positive for this variable i.e. increase total number of crashes (99.99% likelihood). The results illustrate the need to consider parameter heterogeneity in the model frameworks.

5.4 Comparison Exercise

5.4.1 Equivalent Log-Likelihood Measure

The estimated multivariate NB model and the joint NB-MNLFS model fit measures (in term of log-likelihood or Information criterion) are not directly comparable. Therefore, in the current study context, we develop an equivalent approach for comparing the data fit measures of these two different frameworks. The objective of the proposed measure is to compare two model systems arising from different behavioral frameworks. The measure examines data fit for the two components in the proposed model and translates it into a log-likelihood measure. For zero crash data records, the data fit contribution comes directly from total crash count (i.e. proportion component does not exist). For the records with non-zero crashes, we compute the predicted proportions for each crash type and then use that to identify the expected number of total crashes based on the observed number of crashes by crash type. By following this procedure, we ensure that the computed predicted probabilities accounts for total crash prediction errors from crash proportion predictions across all crash types considered. Then, the log-likelihood for the expected number of crashes is generated from the total crash model. For example, if the number of observed crashes for rear-end is 30 and we obtain a predicted proportion probability of 0.3, the expected total number of crashes given 30 observed rear-end crashes = 30/0.3 = 100. Then we compute the log-likelihood for 100 from the total crash model. We repeat this exercise for all crash types. However, as is intuitive, the LL for total crashes will be in a much different scale relative to LL for crashes by crash type. To account for this difference in scales, we compute the ratio of LL for univariate crash type model and total crash model and employ this weight in our equivalent log-likelihood.

The exact equation for the computation of equivalent log-likelihood takes the following form:

$$EL = \sum_i \left[\left(P(c_{iK}) \right)^{1-\varpi_i} + \left(\mathcal{F}_j * ln\left(\prod_{j=1}^{J} \{ P(\mathbb{C}_{ij}) \} \right) \right)^{\varpi_i} \right], with\ \mathbb{C}_{ij} = \frac{c_{ij}}{E(y_{ij}|d_{ij})} \text{ and } \mathcal{F}_j = \frac{\ell_{Sj}}{\ell_{SK}} \quad (13)$$

where i $(i = 1,2,3, \dots, N)$ be the index for STAZ and j $(j = 1,2,3, \dots, J)$ is the index for different collision types. ϖ_i is a dummy with $\varpi_i = 1$ if STAZ i has at least one crash over the study period and 0 otherwise. c_{iK} is total number of observed crashes in STAZ i. $P(c_{iK})$ is the probability for crash occurrence for total crash count K (in Equation 1). $\mathcal{F}_j$ represents the weighting factor of log-likelihoods from independent NB model for different crash type (ℓ_{Sj}) and for total crash (ℓ_{SK}).

The computed equivalent log-likelihood for the joint NB-MNLFS model is-151913.40. On the other hand, the computed log-likelihood at convergence for the multivariate NB model is -160941.40. We can observe that the joint NB-MNLFS model offers better data fit with better likelihood values relative to multivariate NB model. Thus, we can argue that the joint NB-MNLFS model outperforms the multivariate NB model in the current study context with substantially fewer parameters. The reader would note that once an equivalent log-likelihood measure is generated, traditional information criterion measures such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) can be computed using the equivalent log-likelihood for the proposed model. These measures are not computed in our paper because the proposed model provides an improved log-likelihood with fewer parameters. So, it is obvious that the proposed model will offer better fit for these IC measures as well.

5.4.2 Predictive Performance Evaluation

We undertake an in-sample comparison exercise between the multivariate NB model and the joint NB-MNLFS model in terms of predictive performance by employing three different fit of measures: mean prediction bias (MPB), mean absolute deviation (MAD), mean squared prediction error (MSPE), mean percentage error (MPE) and mean absolute percentage error (MAPE). MPB represents the magnitude and direction of average bias in model prediction. The model with the lower MPB provides better prediction of the observed data and is computed as:

$$\text{MPB} = \ mean\ (\hat{y}_i - y_i) \tag{14}$$

where, $\hat{y}_i$ and y_i are the predicted and observed, number of crashes occurring over a period of time in a STAZ i (i be the index for STAZ, $i = 1,2,3,\ldots,8518$). On the other hand, MAD describes average misprediction of the estimated models. The model with lower MAD value closer to zero provides better average predictions of observed data. MAD is defined as:

$$\text{MAD} = \ mean\ |\hat{y}_i - y_i| \tag{15}$$

MSPE quantifies the error associated with model predictions and is defined as:

$$\text{MSPE} = mean\ (\hat{y}_i - y_i)^2 \tag{16}$$

The smaller the MSPE, the better the model predicts observed data. MPE measures the prediction accuracy and is defined as:

$$\text{MPE} = (\frac{\hat{y}_i - y_i}{y_i}) \tag{17}$$

The smaller the MPE, the better the model predicts observed data. Finally, MAPE also measure the size of the error in terms of percentage and is defined as:

$$\text{MAPE} = \left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{18}$$

The smaller the MAPE, the better the model predicts observed data. These measures of fit are generated at disaggregate level: across all crash types and across all observations

Table 5 presents the values of these measures for multivariate NB and joint NB-MNLFS models at the disaggregate level. The results highlight that the joint NB-MNLFS model performance is very similar to the multivariate model performance across the various measures computed. Specifically, multivariate negative binomial model performs marginally better than the proposed model with respect to MPB, MAE and MAPE while the proposed model performs better for equivalent log-likelihood, MAD and MSPE. The comparable performance of the proposed model is particularly significant given the large difference in the number of parameters between the two specifications (55 vs 119). The proposed joint model is substantially parsimonious and yet offers comparable data fit as indicated by the equivalent log-likelihood measure and the in-sample comparison.

6 CONCLUSIONS

In recent years, there is growing recognition that common unobserved factors that influence crash frequency by one attribute level are also likely to influence crash frequency by other attribute levels for the same observation unit. The most common approach employed to address the potential unobserved heterogeneity in existing safety literature is the development of multivariate crash frequency models. In the current study, we formulated and estimated an alternative joint econometric framework to accommodate for the presence of unobserved heterogeneity – referred to as joint negative binomial-multinomial logit fractional split (NB-MNLFS) model. Furthermore, a first of its kind comparison exercise between the most commonly used multivariate model (multivariate random parameter negative binomial model) and the proposed joint NB-MNLFS model was performed by generating equivalent log-likelihood measure.

In our current research effort, a joint NB-MNLFS approach was proposed to employ a crash frequency model for total crashes in conjunction with a fractional split model for proportion of crashes by different collision types. The study was conducted by using data from Florida at the Statewide Traffic Analysis Zone (STAZ) level for the year 2015 considering a host of exogenous variables including roadway characteristics, land use attributes and traffic characteristic for the model estimation. The findings highlighted the presence of common unobserved factors influencing total crash frequency and proportions by crash types. In terms of random parameters, proportion of urban area and average posted speed limit revealed significant variability in the total crash count component of the proposed joint model.

A comprehensive comparison of the proposed model with the most commonly used multivariate negative binomial (NB) model was conducted. The joint and the multivariate models are not directly comparable because the joint model is estimated based on a quasi-likelihood system. To address this, an equivalent log-likelihood measure was generated for the proposed joint model. The equivalent log-likelihood value for the proposed approach was better than the log-likelihood estimate for the multivariate NB model with a substantially fewer number of parameters. To investigate the comparison further, different fit measures were generated to compare the in-sample predictive performance of the two models. The equivalent log-likelihood and goodness of fit measures clearly highlight the comparable performance offered by the proposed joint model relative to the commonly used multivariate approach.

Overall, the model interpretations and fit measures clearly highlight the potential complementary role of the proposed approach for crash frequency analyses.

The paper is not without limitations. In our study, we considered left-turn and right-turn crashes in the same category due to sample size restrictions. In future research efforts, it might be useful to consider them separately given that the crash mechanisms for these crash types could be potentially different. For the comparison analysis, the study used in-sample prediction measures whereas other measures that represent effective parameter numbers accurately might offer additional insights on the comparison between the two frameworks (for example see measures proposed in Spiegelhalter et al. (2002)). Further, given the inherent data aggregation involved in macro-level models considering a framework that accounts for spatial correlation would be beneficial to improve the proposed model formulation. The current methodology while accommodating for frequency by crash type does not consider frequency by crash severity. In future research efforts, considering a combination of frequency by crash type crash severity would be a fruitful exercise. Moreover, it might be interesting to explore the transferability of models developed for crash count and crash type simultaneously by estimating similar models for multiple spatial units across several years.

ACKNOWLEDEGMENTS

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

REFERENCES

Abdel-Aty, M.A., and Radwan, A.E., 2000. Modeling traffic accident occurrence and involvement. Accident Analysis and Prevention 32(5), 633–642.

Abdel-Aty, M., Keller, J., Brady, P., 2005. Analysis of types of crashes at signalized intersections by using complete crash data and tree-based regression. Transportation Research Record: Journal of the Transportation Research Board 1908, 37–45.

Abdel-Aty, M., Lee, J., Siddiqui, C., Choi, K., 2013. Geographical unit based analysis in the context of transportation safety planning. Transportation Research Part A 49, 62–75.

Aguero-Valverde, J., Wu, K.F.K., Donnell, E.T., 2016. A multivariate spatial crash frequency model for identifying sites with promise based on crash types. Accident Analysis and Prevention 87, 8–16.

Amoh-Gyimah, R., Saberi, M., Sarvi, M., 2017. The effect of variations in spatial units on unobserved heterogeneity in macroscopic crash models. Analytic Methods in Accident Research 13, 28–51.

Anastasopoulos, P.C., 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, URL http://www.aptech.com/ (accessed 9.19.15).

Barua, S., El-Basyouny, K., Islam, M.T., 2016. Multivariate random parameters collision count data models with spatial heterogeneity. Analytic Methods in Accident Research 9, 1–15.

Barua, S., El-Basyouny, K., Islam, M.T., 2015. Effects of spatial correlation in random parameters collision count-data models. Analytic Methods in Accident Research 5, 28–42.

Barua, S., El-Basyouny, K., Islam, M.T., 2014. A Full Bayesian multivariate count data model of collision severity with spatial correlation. Analytic Methods in Accident Research 3, 28–43.

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

Cai, Q., Abdel-Aty, M., Lee, J., Wang, L., Wang, X., 2018. Developing a grouped random parameters multivariate spatial model to explore zonal effects for segment and intersection crash modeling. Analytic Methods in Accident Research 19, 1–15.
Chai, C., Wong, Y.D., 2014. Micro-simulation of vehicle conflicts involving right-turn vehicles at signalized intersections based on cellular automata. Accident Analysis and Prevention 63, 94–103.
Cheng, W., Gill, G.S., Dasu, R., Xie, M., Jia, X., Zhou, J., 2017. Comparison of Multivariate Poisson lognormal spatial and temporal crash models to identify hot spots of intersections based on crash types. Accident Analysis and Prevention 99, 330–341.
Chiou, Y.C., and Fu, C., 2015. Modeling crash frequency and severity with spatiotemporal dependence. Analytic Methods in Accident Research 5, 43–58.
Chiou, Y.C., and Fu, C., 2013. Modeling crash frequency and severity using multinomial-generalized Poisson model with error components. Accident Analysis and Prevention 50, 73–82.
Chiou, Y.C., Fu, C., Chih-Wei, H., 2014. Incorporating spatial dependence in simultaneously modeling crash frequency and severity. Analytic Methods in Accident Research 2, 1–11.
Eluru, N., Bhat, C.R., Hensher, D.A., 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accident Analysis and Prevention 40, 1033–1054.
Eluru, N., Bagheri, M., Miranda-Moreno, L.F. and Fu, L., 2012. A latent class modeling approach for identifying vehicle driver injury severity factors at highway-railway crossings. Accident Analysis & Prevention 47, 119-127.
Eluru, N., Chakour, V., Chamberlain, M., Miranda-Moreno, L.F., 2013. Modeling vehicle operating speed on urban roads in Montreal: a panel mixed ordered probit fractional split model. Accident Analysis and Prevention 59, 125–134.
Fountas, G., Anastasopoulos, P.C., Abdel-Aty, M., 2018a. Analysis of accident injury-severities using a correlated random parameters ordered probit approach with time variant covariates. Analytic Methods in Accident Research 18, 57–68.
Fountas, G., Anastasopoulos, P.C., Mannering, F.L., 2018b. Analysis of vehicle accident-injury severities: A comparison of segment- versus accident-based latent class ordered probit models with class-probability functions. Analytic Methods in Accident Research 18, 15–32.
Geedipally, S., Patil, S., Lord, D., 2010. Examination of methods to estimate crash counts by collision type. Transportation Research Record: Journal of the Transportation Research Board 2165, 12–20.
Heydari, S., Fu, L., Lord, D., Mallick, B.K., 2016. Multilevel Dirichlet process mixture analysis of railway grade crossing crash data. Analytic Methods in Accident Research 9, 27–43.
Heydari, S., Fu, L., Miranda-Moreno, L.F., Joseph, L., 2017. Using a flexible multivariate latent class approach to model correlated outcomes: a joint analysis of pedestrian and cyclist injuries. Analytic Methods in Accident Research 13, 16–27.
Ivan, J.N., Wang, C., Bernardo, N.R., 2000. Explaining two-lane highway crash rates using land use and hourly exposure. Accident Analysis and Prevention 32(6), 787–795.
Lee, J., Yasmin, S., Eluru, N., Abdel-Aty, M., Cai, Q., 2018. Macroscopic Analysis of Crash Proportion by Mode: Fractional Split Multinomial Logit Modeling Approach, Accident Analysis and Prevention 111, 12–22.
Li, X., Yan, X., Wu, J., Radwan, E., Zhang, Y., 2016. A rear-end collision risk assessment model based on drivers' collision avoidance process under influences of cell phone use and gender—A driving simulator based study. Accident Analysis and Prevention 97, 1–18.
Liu, C., and Sharma, A., 2018. Using the multivariate spatio-temporal Bayesian model to analyze traffic crashes by severity. Analytic Methods in Accident Research 17, 14–31.

Liu, C., and Sharma, A., 2017. Exploring spatio-temporal effects in traffic crash trend analysis. Analytic Methods in Accident Research 16, 104–116.
Lord, D., and 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.
Ma, X., Chen, S., Chen, F., 2017. Multivariate space-time modeling of crash frequencies by injury severity levels. Analytic Methods in Accident Research 15, 29–40.
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.
Milton, J., and Mannering, F., 1998. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. Transportation 25(4), 395–413.
Milton, J.C., Shankar, V.N., Mannering, F.L., 2008. Highway accident severities and the mixed logit model: an exploratory empirical analysis. Accident Analysis and Prevention 40(1), 260–266.
Mohammadi, M.A., Samaranayake, V.A., Bham, G.H., 2014. Crash frequency modeling using negative binomial models: An application of generalized estimating equation to longitudinal data. Analytic Methods in Accident Research 2, 52–69.
Mothafer, G.I., Yamamoto, T., Shankar, V.N., 2016. Evaluating crash type covariances and roadway geometric marginal effects using the multivariate Poisson gamma mixture model. Analytic Methods in Accident Research 9, 16–26.
Nashad, T., Yasmin, S., Eluru, N., Lee, J., Abdel-Aty, M.A., 2016. Joint modeling of pedestrian and bicycle crashes: copula-based approach. Transportation Research Record 2601, 119–127.
Papke, L.E., and Wooldridge, J.M., 1996. Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. Journal of Applied Econometrics 11(6), 619–632.
Persaud, B.N., and Mucsi, K., 1995. Microscopic accident potential models for two-lane rural roads. Transportation Research Record 1485, 134–139.
Qin, X., Ivan, J.N., Ravishanker, N., 2004. Selecting exposure measures in crash rate prediction for two-lane highway segments. Accident Analysis and Prevention 36(2), 183–191.
Serhiyenko, V., Mamun, S.A., Ivan, J.N., Ravishanker, N., 2016. Fast Bayesian inference for modeling multivariate crash counts. Analytic Methods in Accident Research 9, 44–53.
Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64(4), 583-639.
Wang, K., Ivan, J.N., Ravishanker, N., Jackson, E., 2017. Multivariate poisson lognormal modeling of crashes by type and severity on rural two lane highways. Accident Analysis and Prevention 99, 6–19.
Wang, X., and Abdel-Aty, M., 2008a. Modeling left-turn crash occurrence at signalized intersections by conflicting patterns. Accident Analysis and Prevention 40(1), 76–88.
Wang, X., and Abdel-Aty, M., 2008b. Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models. Accident Analysis and Prevention 40(5), 1674–1682.
Wang, X., and Abdel-Aty, M., 2006. Temporal and spatial analyses of rear-end crashes at signalized intersections. Accident Analysis and Prevention 38(6), 1137–1150.
World Health Organization, 2015. Global status report on road safety 2015. World Health Organization
Yan, X., Radwan, E., Abdel-Aty, M., 2005. Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model. Accident Analysis and Prevention 37(6), 983–995.

Yasmin, S., Eluru, N., Bhat, C.R. and 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., and Eluru, N., 2018. A Joint Econometric Framework for Modeling Crash Counts by Severity, Transportmetrica A: Transport Science 14(3), 230-255.
Yasmin, S., Momtaz, S.U., Nashad, T. and Eluru, N., 2018. A Multivariate Copula Based Macro-level Crash Count Model. Forthcoming Transportation Research Record.
Yasmin, S., and Eluru, N., 2013. Evaluating alternate discrete outcome frameworks for modeling crash injury severity. Accident Analysis and Prevention 59, 506–521.
Yasmin, S., Eluru, N., Lee, J., Abdel-Aty, M., 2016. Ordered fractional split approach for aggregate injury severity modeling. Transportation Research Record: Journal of the Transportation Research Board 2583, 119–126.
Ye, X., Pendyala, R.M., Washington, S.P., Konduri, K., Oh, J., 2009. A simultaneous equations model of crash frequency by collision type for rural intersections. Safety Science 47(3), 443–452.
Zeng, Q., Wen, H., Huang, H., Pei, X., Wong, S.C., 2018. Incorporating temporal correlation into a multivariate random parameters Tobit model for modeling crash rate by injury severity. Transportmetrica A: Transport Science 14, 177–191.
Zhou, M., and Sisiopiku, V., 1997. Relationship between volume-to-capacity ratios and accident rates. Transportation Research Record: Journal of the Transportation Research Board 1581, 47–52.

LIST OF TABLES

TABLE 1: Descriptive statistics of Dependent variables

TABLE 2: Summary statistics of exogenous variables

TABLE 3: Joint NB-MNLFS model estimation results

TABLE 4: The Multivariate NB model results

TABLE 5: In-sample predictive performance measures of two Models

TABLE 1 Descriptive Statistics of Dependent Variables

Variable Names	Definition	Zonal (N=8518)				
		Minimum	**Maximum**	**Average**	**Standard Deviation**	**% of STAZs with zero crash record**
Count variables						
Total Crash	Total number of crashes in STAZ	0.000	877.000	57.193	75.999	5.100
Rear-end Crash	Total number of rear-end crashes in STAZ	0.000	315.000	20.285	29.665	17.363
Head-on Crash	Total number of head-on crashes in STAZ	0.000	76.000	1.185	3.382	59.779
Angular Crash	Total number of angular crashes in STAZ	0.000	180.000	10.907	16.623	21.930
Off-road Crash	Total number of off-road crashes in STAZ	0.000	65.000	4.485	5.615	21.965
Other Single Vehicle Crash	Total number of other single vehicle crashes in STAZ	0.000	99.000	2.410	3.632	35.724
Other Multiple Vehicle Crash	Total number of other multiple vehicle crashes in STAZ	0.000	419.000	11.433	19.142	20.756
Rollover Crash	Total number of rollover crashes in STAZ	0.000	24.000	0.605	1.286	67.739
Sideswipe Crash	Total number of sideswipe crashes in STAZ	0.000	123.000	5.883	9.091	28.575
Fraction variables						
Rear-end crash fraction	Proportion of rear-end crashes (total number of rear-end crashes / total number of motorized vehicle crashes in STAZ)	0.000	1.000	0.284	0.198	17.363
Head-on crash fraction	Proportion of head-on crashes (total number of head-on crashes / total number of motorized vehicle crashes in STAZ)	0.000	1.000	0.022	0.064	59.779
Angular crash fraction	Proportion of angular crashes (total number of angular crashes / total number of motorized vehicle crashes in STAZ)	0.000	1.000	0.166	0.148	21.930
Off-road crash fraction	Proportion of off-road crashes (total number of off-road crashes / total number of motorized vehicle crashes in STAZ)	0.000	1.000	0.132	0.179	21.965

Other single vehicle crash fraction	Proportion of other single vehicle crashes (total number of other single vehicle crashes / total number of motorized vehicle crashes in STAZ)	0.000	1.000	0.061	0.110	35.724
Other multiple vehicle crash fraction	Proportion of other multiple vehicle crashes (total number of other multiple vehicle crashes / total number of motorized vehicle crashes in STAZ)	0.000	1.000	0.172	0.160	20.756
Rollover crash fraction	Proportion of rollover crashes (total number of rollover crashes / total number of motorized vehicle crashes in STAZ)	0.000	1.000	0.023	0.076	67.739
Sideswipe crash fraction	Proportion of sideswipe crashes (total number of sideswipe crashes / total number of motorized vehicle crashes in STAZ)	0.000	1.000	0.09	0.105	28.575

TABLE 2 Summary Statistics of Exogenous Variables

Variable Names	Definition	Zonal (N=8518)			
		Minimum	**Maximum**	**Average**	**Standard Deviation**
Roadway Characteristics					
Proportion of arterial road	Total length of arterial road/ Total road length in STAZ	0.000	1.000	0.477	0.363
Proportion of collector road	Total length of collector road/ Total road length in STAZ	0.000	1.000	0.410	0..353
Proportion of local road	Total length of local road/ Total road length in STAZ	0.000	1.000	0.088	1.952
Proportion of urban road	Total length of urban road/ Total road length in STAZ	0.000	1.000	0.756	0.411
Proportion of rural road	Total length of rural road/ Total road length in STAZ	0.000	1.000	0.219	0.394
Proportion of no control road	Total length of no access control road/ Total road length in STAZ	0.000	1.000	0.912	0.245
Proportion of major road	Total length of major road/ Total road length in STAZ	0.000	1.000	0.594	0.355
Proportion of minor road	Total length of minor road/ Total road length in STAZ	0.000	1.000	0.331	0.334
Signal intensity	Total number of signal/ Total number of intersection in STAZ	0.000	1.667	0.048	0.112
Divided road length	Ln (total length of divided road in STAZ in meter)	-1.394	12.075	6.705	3.463
Proportion of 1 lane road length	Total length of road with 1 lane / Total road length in STAZ	0.000	1.000	0.109	0.159
Proportion of 2 lane road length	Total length of road with 2 lane / Total road length in STAZ	0.000	1.000	0.629	0.346
Proportion of 3 or more lane road length	Total length of road with 3 or more lane / Total road length in STAZ	0.000	1.000	0.241	0.331
Average median width	Ln (average median width in meter in STAZ)	-0.089	5.527	1.501	0.821
Variance of median width	Ln (variance of median width in meter in STAZ)	-4.279	11.724	1.815	2.228

Average inside shoulder width	Ln (average width of inside shoulder in feet in STAZ)	0.000	2.996	0.535	0.586
Average outside shoulder width	Ln (average width of outside shoulder in feet in STAZ)	0.000	3.066	1.588	0.483
Average sidewalk width	Ln (average width of sidewalk in feet in STAZ)	0.000	3.497	1.259	0.795
Intersection density	Ln of total number of intersection per square miles of street in STAZ (Total number of intersection/ Total length of street in STAZ in miles)	-2.608	7.948	1.751	1.003
Average posted speed limit	Ln (average posted speed limit in mile per hour in STAZ)	0.000	4.248	3.390	1.089
Variance of posted speed limit	Ln (variance of posted speed limit in mile per hour in STAZ)	0.000	6.920	2.415	1.951
Built Environment					
Proportion of residential area	Residential area / Total area of STAZ	0.000	0.777	0.024	0.090
Proportion of agricultural area	Agricultural area / Total area of STAZ	0.000	0.987	0.022	0.114
Proportion of industrial area	Industrial area / Total area of STAZ	0.000	0.871	0.002	0.022
Proportion of institutional area	Institutional area / Total area of STAZ	0.000	0.585	0.019	0.134
Proportion of office area	Office and retail area / Total area of STAZ	0.000	0.786	0.008	0.044
Proportion of recreational area	Recreational area / Total area of STAZ	0.000	0.965	0.004	0.037
Land use mix	Land use mix = $\left[\frac{-\sum_k (p_k(lnp_k))}{lnN}\right]$, where k is the category of land-use, p is the proportion of the developed land area devoted to a specific land-use, N is the number of land-use categories in STAZ	0.000	0.859	0.046	0.145
Proportion of urban area	Urban area / Total area of STAZ	0.000	1.000	0.731	0.425
LTZM	Ln of area of STAZ in meter square (used as the exposure variable in the model)	-3.749	21.553	14.596	2.271
Traffic Characteristics					

AADT	Ln of average annual daily traffic in STAZ	0.000	13.312	10.362	2.093
Truck AADT	Ln of average annual daily truck traffic in STAZ	0.000	11.020	4.901	3.870
VMT	Ln of vehicles miles traveled in STAZ	0.000	13.524	9.442	2.192
Truck VMT	Ln of truck vehicles miles traveled in a STAZ	-10.185	11.243	4.133	3.584
Proportion of heavy traffic	Total number of truck traffic/ Total number of vehicles in STAZ	0.000	0.848	0.037	0.051

TABLE 3 Joint NB-MNLFS Model Estimation Results

Joint Component	**NB Model (Counts)**	**MNLFS Model (Proportions)**							
Crash Type	**Total Crash**	**Rear-end**	**Head-on**	**Angle**	**Off-Road**	**Other Single Vehicle**	**Other Multiple Vehicle**	**Rollover**	**Sideswipe**
Variable Name	**Estimate (S.E)***	**Estimate (S.E)**	**Estimate (S.E)**	**Estimate (S.E)**	**Estimate (S.E)**	**Estimate (S.E)**	**Estimate (S.E)**	**Estimate (S.E)**	**Estimate (S.E)**
Constant	-13.622 (0.123)	---	-1.283 (0.366)	0.305 (0.234)	1.417 (0.26)	0.802 (0.328)	0.861 (0.215)	-0.237 (0.47)	0.286 (0.224)
Roadway Characteristic									
Proportion of arterial road	0.96 (0.051)	0.335 (0.078)	---	---	-0.187 (0.109)	---	---	---	0.455 (0.115)
Proportion of urban road	0.631 (0.091)	---	---	---	-0.46 (0.17)	---	---	-0.735 (0.365)	---
Intersection density	0.539 (0.022)	---	---	0.299 (0.04)	-0.096 (0.049)	---	0.199 (0.039)	-0.183 (0.105)	0.105 (0.047)
Proportion of road length with 3 or more lanes	0.656 (0.061)	---	---	---	---	---	---	---	0.539 (0.113)
Average posted speed limit	-0.184 (0.022)	---	0.179 (0.087)	0.094 (0.034)	0.112 (0.038)	---	---	---	---
Divided road length	-0.114 (0.007)	---	---	---	-0.022 (0.011)	---	---	---	-0.04 (0.012)
Average inside shoulder width	0.157 (0.035)	---	-0.306 (0.141)	---	---	---	---	---	---
Average outside shoulder width	-0.738 (0.044)	---	---	---	0.323 (0.092)	0.406 (0.119)	---	0.672 (0.228)	---
Average sidewalk width	0.167 (0.027)	---	---	---	-0.115 (0.05)	---	---	-0.214 (0.103)	---
Land Use Attribute									
Proportion of urban area	2.472 (0.085)	0.178 (0.083)	---	---	-0.59 (0.163)	-0.883 (0.121)	---	-1.219 (0.363)	---
Standard Deviation	0.348 (0.037)								

Land use mix	-1.467 (0.116)	---	---	---	---	---	---	---	---
Traffic Characteristic									
AADT	0.107 (0.015)	0.133 (0.021)	---	---	---	-0.067 (0.027)	---	---	---
Truck AADT	0.023 (0.005)	---	---	---	-0.04 (0.01)	---	---	-0.04 (0.022)	---
Dispersion parameter	1.931 (0.029)	---	---	---	---	---	---	---	---
Correlation 1	0.228 (0.042)	---	---	---	0.228 (0.042)	0.228 (0.042)	---	0.228 (0.042)	---
Correlation 2	0.116 (0.066)	0.116 (0.066)	0.116 (0.066)	0.116 (0.066)	---	---	0.116 (0.066)	---	0.116 (0.066)

Note: *S.E. = Standard Error

TABLE 4 Multivariate NB Model Results

Crash Type	Rear-end	Head-on	Angular	Off-Road	Other Single Vehicle	Other Multiple Vehicle	Rollover	Sideswipe
Variable Name	**Estimate (S.E)***	**Estimate (S.E)**	**Estimate (S.E)**	**Estimate (S.E)**	**Estimate (S.E)**	**Estimate (S.E)**	**Estimate (S.E)**	**Estimate (S.E)**
Constant	-16.520 (0.137)	-19.846 (0.259)	-17.464 (0.142)	-16.573 (0.128)	-17.149 (0.156)	-16.396 (0.133)	-19.282 (0.278)	-17.543 (0.159)
Roadway Characteristic								
Proportion of arterial road	0.968 (0.052)	0.558 (0.074)	0.553 (0.050)	0.325 (0.042)	0.555 (0.052)	0.642 (0.052)	0.317 (0.076)	0.990 (0.058)
Standard deviation	--	--	--	--	--	--	--	0.825 (0.057)
Proportion of urban road	0.880 (0.101)	0.901 (0.163)	0.982 (0.101)	0.578 (0.080)	0.635 (0.102)	0.982 (0.100)	--	0.966 (0.114)
Intersection density	0.507 (0.024)	0.742 (0.041)	0.857 (0.024)	0.411 (0.020)	0.606 (0.026)	0.719 (0.024)	0.398 (0.038)	0.587 (0.027)
Standard deviation	--	0.221 (0.038)	--	--	--	--	--	--
Proportion of road length with 3 or more lanes	0.751 (0.059)	0.758 (0.080)	0.643 (0.055)	0.346 (0.047)	0.506 (0.057)	0.914 (0.058)	0.402 (0.083)	1.164 (0.061)
Average posted speed limit	-0.210 (0.021)	-0.139 (0.031)	-0.147 (0.021)	-0.083 (0.017)	-0.204 (0.021)	-0.258 (0.021)	-0.144 (0.029)	-0.245 (0.023)
Divided road length	-0.037 (0.008)	-0.089 (0.012)	-0.086 (0.007)	-0.061 (0.006)	-0.075 (0.008)	-0.077 (0.008)	-0.026 (0.011)	-0.131 (0.008)
Average inside shoulder width	0.166 (0.036)	-0.262 (0.052)	--	0.193 (0.028)	0.163 (0.035)	0.070 (0.035)	0.177 (0.049)	0.200 (0.038)
Average outside shoulder width	-0.869 (0.044)	-0.556 (0.062)	-0.785 (0.042)	-0.561 (0.036)	-0.579 (0.045)	-0.784 (0.043)	-0.235 (0.067)	-0.935 (0.047)
Average sidewalk width	0.177 (0.028)	0.259 (0.043)	0.210 (0.027)	--	0.099 (0.028)	0.275 (0.028)	-0.167 (0.038)	0.197 (0.031)
Land Use Attribute								
Proportion of urban area	2.770 (0.090)	2.061 (0.138)	2.421 (0.089)	1.682 (0.071)	1.540 (0.090)	2.470 (0.089)	1.174 (0.078)	2.525 (0.100)
Land use mix	-1.100 (0.115)	-1.755 (0.155)	-1.149 (0.110)	-1.354 (0.089)	-0.264 (0.104)	-1.118 (0.112)	-1.087 (0.152)	-1.203 (0.120)

Traffic Characteristic								
AADT	0.157 (0.016)	0.135 (0.029)	0.164 (0.017)	0.141 (0.015)	0.107 (0.018)	0.099 (0.016)	0.220 (0.031)	0.187 (0.018)
Truck AADT	0.027 (0.005)	0.073 (0.007)	0.026 (0.005)	--	0.044 (0.005)	0.017 (0.005)	--	0.043 (0.005)
Dispersion parameter	0.739 (0.023)	0.479 (0.076)	0.511 (0.026)	0.116 (0.025)	0.394 (0.036)	0.585 (0.026)	0.460 (0.056)	0.495 (0.036)
Correlation 1	--	--	--	0.897 (0.018)	0.897 (0.018)	--	0.897 (0.018)	--
Correlation 2	1.064 (0.012)	1.064 (0.012)	1.064 (0.012)	--	--	1.064 (0.012)	--	1.064 (0.012)

*S.E. = Standard Error

TABLE 5 In-Sample Predictive Performance Measure

	Log-Likelihood		**MPB (Disaggregate level)**		**MAD (Disaggregate level)**		**MSPE (Disaggregate level)**		**MPE (Disaggregate level)**		**MAPE (Disaggregate level)**	
	RPMNB	Joint NB-MNLFS	RPMNB	Joint NB-MNLFS	RPMNB	Joint NB-MNLFS	RPMNB	Joint NB-MNLFS	RPMNB	Joint NB-MNLFS	RPMNB	Joint NB-MNLFS
Across Crash types												
Rear-end	**-30693.973**	-30735.310	34.101	**28.955**	47.142	**37.189**	50265.214	**19320.034**	3.373	**2.745**	3.839	**3.035**
Head-on	-10752.476	**-8575.469**	**0.633**	1.945	**2.067**	2.846	**50.734**	60.991	**0.237**	0.735	**0.629**	0.923
Angular	-26073.451	**-25517.422**	**14.840**	15.877	22.788	**20.607**	13484.108	**7891.729**	**2.069**	2.425	**2.554**	2.698
Off-road	-20435.015	**-19233.875**	**3.299**	7.864	**6.452**	8.950	**419.719**	492.010	**0.847**	2.045	**1.369**	2.231
Other-Single Vehicle	-16448.825	**-14967.931**	**1.887**	4.292	**3.839**	5.254	**161.141**	206.474	**0.696**	1.581	**1.157**	1.793
Other-Multiple Vehicle	-26346.472	**-25669.352**	**14.723**	15.770	23.742	**21.730**	16052.810	**10176.374**	**2.301**	2.792	**2.788**	3.069
Rollover	-8079.455	**-5769.824**	**0.269**	0.909	**0.954**	1.294	**7.518**	10.973	**0.013**	0.235	**0.351**	0.409
Sideswipe	-22111.728	**-21444.219**	10.154	**8.898**	14.995	**12.098**	8463.525	**3782.182**	2.093	**2.053**	2.593	**2.354**
Across Observation (8518 records)	-18.894	**-17.834**	**79.906**	84.511	121.980	**109.968**	88904.768	**41940.766**	**11.629**	14.613	**15.280**	16.512
(Multivariate NB vs Joint NB-MNLFS) – (3 vs 3 at STAZ level)												

***Better model is in bold format**