A Joint Econometric Framework for Modeling Crash Counts by Severity

Shamsunnahar Yasmin
Postdoctoral Associate
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

This paper proposes an innovative joint econometric framework for examining total crash count and crash proportion by different crash severity. Specifically, we propose to consider a crash frequency model for total crashes and a fractional split model that considers proportion by crash severity. The model ties total crash counts and crash proportion by accommodating for the potential common unobserved heterogeneity (across study unit) in the joint framework. In our proposed approach, irrespective of the number of crash frequency variables the dimensions to be investigated is *'two'*, offering substantial benefits in terms of parameter stability and computational time as opposed to the traditional multivariate approaches. The proposed model is demonstrated in the study by employing a joint Negative Binomial-Ordered Logit Fractional Split (NB-OLFS) model framework. In the joint econometric framework, we also allow for the potential unobserved heterogeneity to vary across study units. The empirical analysis is conducted using zonal level crash count data for different crash severity levels from the state of Florida for the year 2015. The models are estimated employing a comprehensive set of exogenous variables − sociodemographic characteristics, socioeconomic characteristics, built environment, transport infrastructure and traffic characteristics. We also generate a comprehensive set of measures to evaluate model performance and data fit of the proposed framework. The results clearly highlight the superiority of the joint model in terms of data fit compared to independent model. The applicability of the proposed framework is demonstrated by generating spatial distribution of predicted motor vehicle crash frequency and predicted crash counts by severity levels.

Keywords: count model, crash count by severity, negative binomial model, ordered fractional split model, joint model, crash prediction model

1. INTRODUCTION

Road traffic crashes and their consequences (property damage, injuries and fatalities) remain a global health concern given the extent of societal, emotional and economic impacts of these unfortunate events. However, many developed countries have been able to achieve a reduction in road crash related fatalities through coordinated multi-sectoral responses, devised and implemented based on evidence-based data-driven strategies. Crash frequency analysis is a major component for devising and evaluating these road safety policies. The analysis is focused on identifying attributes that result in traffic crashes and propose effective countermeasures to improve roadway infrastructure design and operational efficiency. The outcome of these models is also useful to devise safety-conscious decision support tools to facilitate a proactive approach in assessing medium and long term policy based countermeasures.

Traffic crashes aggregated at a certain spatial scale are non-negative integer valued random events. Researchers have employed a wide array of econometric approaches (linear regression, count regression and discrete outcome models) for quantifying the impact of exogenous factors at different scales - micro-level such as intersection or segment and macro-level such as traffic analysis zone or census tract. The application of traditional Poisson and negative binomial (NB) regression models remains predominant in examining univariate crash count events. However, as argued in different studies, crash counts across different attributes (crash severity, vehicle types, road user groups, crash types) are likely to be correlated and hence are multivariate in nature. With the emergence of increased computational power, examining such "jointness" is becoming more tractable and useful. In fact, a number of studies have employed multivariate econometric frameworks for examining multiple crash count variables – often referred to as multivariate count models (Mannering et al., 2016).

The current research effort contributes to the safety literature in examining crash count events methodologically and empirically by proposing an alternative crash count model formulation for multivariate count variables. Specifically, we propose an innovative joint econometric framework for examining total crash count and crash proportion by crash severity. In our approach, crash count is analyzed using a NB model and the crash proportions by severity is represented as an ordered fractional split model. The joint NB-ordered fractional split model ties total crash counts and crash severity proportions by accommodating for the influence of common unobserved heterogeneity. Such an integrated approach is appealing compared to traditional multivariate count framework for many reasons. From the methodological perspective, the proposed approach (1) is computationally less burdensome irrespective of the number of count dimensions – the proposed approach requires only <u>two equations</u>, (2) allows for unobserved heterogeneity across and within count and fractional split components, (3) recognizes the inherent ordering of the dependent variable in the fractional split component and (4) allows for a parsimonious specification while retaining the benefits of both the count and fractional split approaches. From the empirical perspective, the proposed approach (1) provides a complete picture of count events with respect to total counts and contribution of each count dimension under consideration and (2) provides a platform to perform policy scenario analysis considering possible change in total count events as well as changes within and across different count dimensions.

To the best of the authors' knowledge, this is the first attempt to employ such a joint framework within an ordered framework for examining count events. The proposed joint NB-Fractional split econometric framework is generic and applicable for examining count and ordered events simultaneously for any domain. In current study context, the application of the proposed model is demonstrated by using zonal level motor vehicle crash count data for different crash severity levels from the state of Florida employing a comprehensive set of exogenous variables from a host of variable groups including − sociodemographic characteristics, socioeconomic characteristics, built environment, transport infrastructure and traffic characteristics.

2. EARLIER STUDIES AND CURRENT STUDY IN CONTEXT

2.1 Summary of Earlier Studies

In our paper, we reviewed studies employing econometric framework for examining crash counts. A number of research efforts have examined traffic crashes aggregated at a certain spatial scale to gain a comprehensive understanding of the factors that affect crash occurrences (see Lord and Mannering, 2010; Mannering and Bhat, 2014; Mannering et al., 2016 and Yasmin and Eluru, 2016 for detailed reviews). In general, crash count studies can be grouped into three broad categories based on the dimensions of dependent variables examined: (1) univariate crash count models, (2) multivariate crash count models and (3) crash proportion models.

The *first group of studies* in the transportation safety area identify a single count variable for different crash attribute level (road user group, crash severity, crash types, or vehicles types), for a spatial unit and study the impact of exogenous variables. Among different road user groups, considerable research has been carried out for examining total crash risk and motor vehicle crash risk (Shin and Washington, 2012; Huang et al., 2016; Roshandeh et al., 2016; Lee et al., 2015). Given the increased patronage for active mode of transportation, studies examining pedestrian and/or bicycle crash risk have also increased over the last decade (Cai et al., 2016; Wei and Lovegrove 2013). It is beyond the scope of the paper to review all the research on transportation crash frequency employing univariate crash count models (see Lord and Mannering, 2010 and Yasmin and Eluru, 2016 for a detailed review of this group of studies). With respect to crash frequency by crash attribute levels, a significant number of studies have developed crash count models by severity levels: fatal crash count, fatal/serious injury crash count, injury crash count and no injury crash count (Dong et al., 2017; Abdel-Aty et al., 2011; Lee et al., 2014; Naderan and Shahi, 2010). Another crash classification that has been considered in examining a single count variable is crash type (Lee and Mannering, 2002; Hosseinpour et al., 2014). In examining crash counts in a univariate modeling system, statistical modeling approaches considered include negative binomial regression model, generalized linear modeling techniques, ordinary least square regression, Poisson-lognormal, generalized Poisson regression, negative multinomial regression, random effect negative binomial, geographically weighted Poisson regression, geographically weighted negative binomial regression, bayesian Poisson lognormal, quasi induced exposure method and bayesian spatial regression.

While these approaches perform adequately in the presence of a single count variable, these models ignore the correlations across different levels of crash attributes. For instance, crash frequencies by different severity levels are likely to be dependent for the same observation unit resulting in a multivariate crash event set. For a study unit, if multiple dependent variables are available it is plausible to imagine that common unobserved factors that affect one dependent variable might also affect the other dependent variables. The process of incorporating the impact of unobserved factors poses methodological challenges. Essentially, accommodating the impact of unobserved factors recognizes that the multiple dimensions of interest have common error terms that affect the dependent variables. At the same time, ignoring the presence of such potential correlation may result in biased parameter estimates and thus lead to inaccurate policy implications (Chamberlain, 1980; Eluru and Bhat, 2007; Washington et al., 2010).

The *second group of studies* – multivariate crash count models examine multiple dependent variables for each study unit. A summary of earlier studies employing multivariate crash count models is presented in Table 1. The information provided in the table includes the study unit considered, the methodological approach employed, the dependent variables analysed and the number of dimensions examined in the multivariate frameworks. The following observations can be made from the table. The most prevalent study unit considered is roadway segment for micro-level analysis. The model structures employed in developing multivariate count model include multivariate-Poisson model, multivariate Poisson-lognormal model, multivariate random-parameters zero-inflated negative binomial model, multinomial-generalized Poisson model, multivariate random parameter model with spatial heterogeneity, copula based bivariate model, multivariate conditional autoregressive model, multivariate tobit model, multivariate Poisson gamma mixture count model, multivariate mixture latent class multivariate model and simultaneous equation models. Within the multivariate scheme, studies have predominantly explored crash frequency by severity level, frequency by crash type, frequency by crash location, crash counts by active mode of transportation, road user group and vehicle type. The dimensions considered in the multivariate econometric framework varies from 2 to 6 based on the number of dependent variables considered in the modeling exercise.

The multivariate count modeling approaches presented in Table 1, 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). The common error term across the dependent variables allows for the possible unobserved effects. Any 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[1]. Thus, the estimation procedure requires the adoption of maximum simulated likelihood (MSL) approach in the classical approach

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

or Markov Chain Monte Carlo (MCMC) approach in the Bayesian realm. MSL and MCMC methods provide substantial flexibility in accommodating for unobserved heterogeneity. However, the probability evaluation with high dimensional integrals is affected by the challenges in generating high dimensionality of random numbers and longer computational run times. Furthermore, the stability of the variance-covariance matrix is often sensitive to model specification and number of simulation draws.

The *third group of studies* is based on examination of crash proportions. One of the recently employed analytical framework for examining proportions is known as the fractional split approach (Eluru et al., 2013; Papke and Wooldridge, 1996). In a traditional count modeling approach, crash counts by different crash attributes are estimated using separate crash propensity equations for each attribute under consideration. In multivariate count modeling approaches that study frequency across different attributes in a joint framework, 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 (studies discussed in second group above) *i.e.* there is no interaction of observed effects across the multiple count models. For forecasting and policy evaluation, it might be beneficial to evaluate the impact of exogenous variables in a framework that directly relates a single exogenous variable to all count variables simultaneously. Such specification is not feasible within the traditional univariate or multivariate count modeling approaches. The fractional split approach provides an alternative approach toward achieving such an objective.

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 through a single equation allowing us to obtain a parsimonious specification of exogenous variable impacts. 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 and crash severity simultaneously by building on multinomial-Poisson transformation (Chiou and Fu, 2013, 2015; Chiou et al., 2014). However, the approach employed in these studies intrinsically ignores the inherent ordering between ordinal attribute levels (for crash severity from no injury to fatal) (see Lee et al., 2016 for a similar approach within fractional split framework). Also, the number of equations in the multinomial-Poisson approach increases with the increase in crash severity dimensions. The reader would also note that the multinomial-Poisson transformation is significantly different in mathematical formulation from the multinomial fractional split approach. Recently, Yasmin et al. (2016) developed an ordered outcome fractional split model that allows the analysis of proportion for variables with multiple alternatives while also recognizing the inherent ordering in the severity[2].

2.2 Current Study in Context

The current study proposes a new joint modeling approach that builds on earlier research from second and third group of research efforts. Specifically, we propose to consider a crash frequency model for total crashes in conjunction with a fractional split model that considers proportion by each crash attributes. Similar to the multivariate studies, a simulation based framework is employed to accommodate for the influence of common unobserved effects in frequency and proportion components. The proposed approach offers many advantages. ***First***, in the proposed joint approach, the dimensions that define the joint distribution are no longer tied to the number of crash frequency variables. In our approach, irrespective of the number of crash frequency variables the dimensions to be investigated is *'two'*, offering substantial benefits in terms of parameter stability and computational time as opposed to the traditional multivariate approaches. ***Second***, the proposed approach retains the benefit of the fractional split model that allows observed variables to affect the proportion across crash frequency variables. ***Third***, the proposed approach also recognizes the inherent ordering of the dependent variable in the fractional split component. ***Finally***, the proposed approach allows for a parsimonious specification for the components under consideration. To be sure, the proposed approach is not suggested as a replacement to existing multivariate approaches but as an alternative approach that could potentially augment the available approaches for crash frequency analysis.

In the current study context, we demonstrate the application of the proposed approach by employing a Negative Binomial-Ordered Logit Fractional Split (NB-OLFS) model framework[3]. Our study also allows for the potential unobserved heterogeneity to vary across the study unit in the joint framework. We also generate a comprehensive set of measures to evaluate model performance and data fit of the proposed framework. In the current study context, the proposed model is estimated using zonal level crash count data for different crash severity levels from the state of Florida employing a comprehensive set of exogenous variables − sociodemographic characteristics, socioeconomic characteristics, built environment, transport infrastructure and traffic characteristics. The outcomes of this macro-level crash count model can be used to devise safety-conscious decision support tools to facilitate proactive approach in assessing medium and long term policy based countermeasures considering possible change in total count events as well as changes within and across different severity dimensions.

[2] The reader would note that there might be other approaches to combining counts and severity. For example, see Pei et al. (2011) for an approach that employs MCMC based joint model estimation of crash counts and crash counts by severity. Also, Wang et al. (2011) and Xu et al. (2014) developed two-stage model by incorporating a sequential estimation of Poisson-mixed multinomial and bivariate logistic-Tobit model, respectively.

[3] It is worthwhile to recognize that, the proposed approach can also be implemented with unordered or generalized ordered fractional split approaches. Moreover, the approach can be employed in developing both macro and micro-level count models.

The rest of the paper is organized as follows. Section 3 provides details of the econometric model framework used in the analysis. In Section 4, the data and dependent variable formation procedures are described. Model comparison results and estimation results are presented in Section 5. Section 6 concludes the paper.

3. ECONOMETRIC FRAMEWORK

3.1 Model Structure

The focus of our study is to jointly model "total number of crashes" and "proportion of crashes by severity". Let us assume that i $(i = 1,2,3, \dots, N)$ be the index for Statewide Traffic Analysis Zone (STAZ) and k $(k = 1,2,3, \dots, K)$ be the index to represent injury severity categories. In this empirical study, k take the values of 'no injury' $(k = 1)$, 'minor injury' $(k = 2)$, 'incapacitating injury' $(k = 3)$ and 'fatal injury' $(k = 4)$. For the joint approach, the equation system for modeling total crash count in the usual NB formulation can be written as:

$$P(c_i) = \frac{\Gamma\left(c_i + \frac{1}{\alpha}\right)}{\Gamma(c_i + 1)\Gamma\left(\frac{1}{\alpha}\right)} \left(\frac{1}{1 + \alpha\mu_i}\right)^{\frac{1}{\alpha}} \left(1 - \frac{1}{1 + \alpha\mu_i}\right)^{c_i} \tag{1}$$

where, c_i be the index for crashes occurring over a period of time in STAZ i. $P(c_i)$ is the probability that STAZ i has c_i number of crashes. $\Gamma(\cdot)$ is the gamma function, α is NB overdispersion parameter and μ_i is the expected number of crashes occurring in STAZ i over a given time period. In equation 1, we can express μ_i as a function of explanatory variables by using a log-link function:

$$\mu_i = E(c_i|\boldsymbol{z}_i) = exp((\boldsymbol{\delta} + \boldsymbol{\zeta}_i)\boldsymbol{z}_i + \ln(Area) + \varepsilon_i + \eta_i) \tag{2}$$

where, $\boldsymbol{z}_i$ is a vector of explanatory variables associated with STAZ i. $Area$ is the STAZ area used as an offset variable in the NB model specification[4]. $\boldsymbol{\delta}$ is a vector of coefficients to be estimated. $\boldsymbol{\zeta}_i$ 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: $\boldsymbol{\zeta}_i \sim N(0, \boldsymbol{\pi}^2)$. ε_i is a gamma distributed error term with mean 1 and variance α. η_i captures unobserved factors that simultaneously impact total number of crashes and proportion of crashes by severity for STAZ i.

In the joint model framework, the modeling of crash proportions by severity levels is undertaken using the Ordered Logit Fractional Split (OLFS) model. In the ordered outcome

[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 in order 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.

framework, the actual injury severity proportions (y_{ik}) are assumed to be associated with an underlying continuous latent variable (y_i^*). The latent propensity equation is typically specified as the following linear function:

$$y_i^* = ((\boldsymbol{\beta} + \boldsymbol{\rho_i})\boldsymbol{x}_i + \xi_i \pm \eta_i), y_{ik} = k \; if \; \tau_{k-1} < y_i^* < \tau_k \tag{3}$$

The latent propensity y_i^* is mapped to the actual severity proportion categories y_{ik} by τ thresholds ($\tau_0 = -\infty \; and \; \tau_K = +\infty$) as presented in equation 3. $\boldsymbol{x}_i$ is a vector of attributes (not including a constant) that influences the propensity associated with severity proportion categories. $\boldsymbol{\beta}$ is the corresponding vector of mean effects. $\boldsymbol{\rho_i}$ is a vector of unobserved factors on severity proportion propensity for STAZ i and its associated zonal characteristics assumed to be a realization from standard normal distribution: $\boldsymbol{\rho} \sim N(0, \boldsymbol{\sigma}^2)$. ξ_i is an idiosyncratic error term assumed to be identically and independently standard logistic distributed across STAZ i. η_i term generates the correlation between equations for total number of crashes and crash proportions by severity levels. To estimate the model presented in equation 3, we assume that:

$$E(y_{ik}|\boldsymbol{x}_i) = H_{ik}(\lambda, \psi), 0 \le H_{ik} \le 1, \sum_{k=1}^{K} H_{ik} = 1 \tag{4}$$

H_{ik} in our model takes the ordered logistic probability (Λ) form for the severity category k. Given these relationships across different parameters, the resulting probability (Λ) for the OLFS model takes the following form:

$$\Lambda(y_{ik} = k) = \varphi\{\tau_k - ((\boldsymbol{\beta} + \boldsymbol{\rho_i})\boldsymbol{x}_i \pm \eta_i)\} - \varphi\{\tau_{k-1} - ((\boldsymbol{\beta} + \boldsymbol{\rho_i})\boldsymbol{x}_i \pm \eta_i)\} \tag{5}$$

where, $\varphi(\cdot)$ is the standard logistic cumulative distribution function. In employing fractional split approach within an ordered framework, previous studies have typically assumed $\varphi(\cdot)$ as standard normal distributed (Papke and Wooldridge, 1996; Eluru et al., 2013). However, as indicated by Papke and Wooldridge, (2008), logistic function form of $\varphi(\cdot)$ is also feasible. The $\pm$ sign in front of η_i in equation 5 indicates that the correlation in unobserved individual factors between total crashes and crash proportions by severity levels 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 severe crashes. On the other hand, negative sign implies that STAZs with higher number of crashes 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 total number of crashes and crash proportions by severity levels can vary across STAZs. Therefore, in the current study, the correlation parameter η_i is parameterized as a function of observed attributes as follows:

$$\eta_i = \boldsymbol{\gamma}_i \boldsymbol{s}_i \tag{6}$$

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

3.2 Model Estimation

In examining the model structure of total crash count (equation 1) and proportions of crashes by severity levels (equation 5), it is necessary to specify the structure for the unobserved vectors $\boldsymbol{\zeta}, \boldsymbol{\rho}$ and $\boldsymbol{\gamma}$ represented by Ω. In this paper, it is assumed that these elements are drawn from independent realization from normal population: $\Omega \sim N(0, (\boldsymbol{\pi^2}, \boldsymbol{\sigma^2}\ \boldsymbol{\nu}^2))$. Thus, conditional on Ω, the likelihood function for the joint probability can be expressed as:

$$L_i = \int_{\Omega} P(c_i) \times \prod_{k=1}^{K} \big(\Lambda(y_{ik} = k)\big)^{\varpi_i d_{ik}} \, d\Omega \tag{7}$$

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. d_{ik} is the proportion of crashes in severity category k. Finally, the log-likelihood function is:

$$LL = \sum_i Ln(L_i) \tag{8}$$

All the parameters in the model are estimated by maximizing the logarithmic function LL presented in equation 8. The parameters to be estimated in the model are: $\boldsymbol{\delta}$, α, $\boldsymbol{\beta}$, τ, $\boldsymbol{\pi}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\nu}$. 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 (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

4.1 Study Area

The study area, the state of Florida is associated with 8,518 STAZ with a population of about 18 million and approximately 9 million households. More than 90% of these population lives in urban area which covers only about 14% of total land area of Florida (U.S. Census Bureau, 2012). Like many other state of the United States, Florida is also an auto oriented state where nearly 80% of Floridians commute to work by automobiles and 93% of households (HH) have access to at least one passenger vehicle (FDOT and CUTR, 2015). Data for the empirical study is compiled from Florida Department of Transportation Crash Analysis Reporting (CAR) and Signal Four Analytics (S4A) databases. Florida Department of Transportation CAR and S4A are long and short forms of crash reports in the State of Florida, respectively. The long form crash report includes higher injury severity level or crash related to criminal activities (such as hit-and-run or Driving Under Influence). The Short Form Report is used to report all other types of traffic crashes. Crash data records from short and long form databases are compiled in order to generate complete information on road crashes and hence are used for the purpose of analysis in the current study context.

4.2 Dependent Variables and Data Descriptions

This study is focused on crashes involving motor vehicles at the zonal level - pedestrian or bicycle involved crashes were excluded. The geocoded crash data involving motorized vehicles are aggregated at the level of STAZ for the year 2015 – dependent variables for count model component (represented as NB model) of the joint system. For the year 2015, Florida has a record of 494,831 motor vehicle crashes with an average of 58.09 crashes per STAZ (ranging from 0 to 864 crashes). These crashes are further classified by crash severity outcomes (no injury, minor injury, incapacitating injury and fatal injury) at the zonal level. In this case of four severity levels, the dependent variable for fractional split component (represented as OLFS model) can be represented as proportions (number of specific severity level/total number of all crashes) as follows: (1) proportion of no injury crashes, (2) proportion of minor injury crashes, (3) proportion of incapacitating injury crashes and (4) proportion of fatal crashes. The dependent variables and sample size for both components are presented in upper row panel of Table 2. From Table 2, we can observe that, as expected, number of no injury crashes has the highest proportion followed by proportion of minor injury crashes.

In addition to the crash database, the explanatory attributes considered in the empirical study are also aggregated at the STAZ level. The selected explanatory variables can be grouped into five broad categories: sociodemographic characteristics, socioeconomic characteristics, built environment attributes, transport infrastructure and traffic characteristics. These variables are collected from different data sources including: 2010 US census data, 2009-2013 American Community Survey (ACS), Florida Geographic Data Library (FDGL), Florida Department of Transportation (FDOT) and Signal Four Analytics (S4A) databases. <u>Sociodemographic characteristics</u> included are household (HH) density, dependence[5], proportion of female population, proportion of Caucasian population, proportion of Asian population, proportion of Hispanic population and proportion of African-American population. <u>Socioeconomic</u>

[5] Dependence is defined as the ratio of youth (15 years or younger) and elderly (65 years or more) to working age persons.

characteristics included are automobile commuters, transit commuters, walk commuters, bike commuters, number of workers that work from home, employment density and population below poverty level. Built environment attributes included are proportion of urban area, law enforcement offices, restaurants, park and recreational centers, transportation hubs, shopping centers, land use mix, proportion of retail and office land use and proportion of industrial land use. Transport infrastructures included are local roads, major roads, traffic signal density and intersection density. Traffic characteristics included are vehicle miles travelled (VMT), proportion of heavy vehicle miles travelled, average speed and traffic intensity.

Table 2 offers a summary of the sample characteristics of the exogenous factors in the estimation dataset. The table represents the definition of variables considered for final model estimation along with the zonal minimum, maximum and average values. The final specification of the model development was based on removing the statistically insignificant variables in a systematic process based on statistical significance (90% significance level). The specification process was also guided by prior research and parsimony considerations. In estimating the models, several functional forms and variable specifications were explored. The functional form that provided the best result was used for the final model specifications and, in Table 2, the variable definitions are presented based on these final functional forms.

5. EMPIRICAL ANALYSIS

5.1 Model Specification and Overall Measures of Fit

The empirical analysis involves estimation of three different models: 1) an independent Negative Binomial (NB) and Ordered Logit Fractional Split (OLFS) model system, 2) joint NB-OLFS model without correlation parameterization and 3) joint NB-OLFS model with correlation parameterization. The independent model (separate NB and OLFS models) were estimated to establish a benchmark for comparison. Prior to discussing the estimation results, we compare the performance of these models in this section. We employ the Bayesian Information Criterion (BIC) to determine the best model between independent and joint models. The BIC for a given empirical model is equal to:

$$BIC = -2LL + K\,ln(Q) \tag{9}$$

where LL is the log likelihood value at convergence, K is the number of parameters, and Q is the number of observations. The model with the lower BIC is the preferred model.

The log-likelihood values at convergence for the models estimated are as follows: (1) Independent NB-OLFS (with 45 parameters) is -44872.33 (2) joint NB-OLFS model without correlation parameterization (with 46 parameters) is -44863.11 and (3) joint NB-OLFS model with correlation parameterization (with 47 parameters) is -44856.98. The BIC values for the final specifications of the independent, joint NB-OLFS models without and with correlation parameterization are 90150.76, 90142.52 and 90139.31, respectively. The comparison exercise clearly highlights the superiority of the joint model with the correlation parameterization in terms of data fit compared to independent model.

5.2 Estimation Results

In presenting the effects of exogenous variables, we will restrict ourselves to the discussion of the joint model with the correlation parameterization[6]. For the ease of presentation, the crash count component (NB model) and crash proportion component (OLFS model) are presented and discussed separately. Table 3 presents the estimation results of the joint NB-OLFS model with NB component results in second and third column panels of the table and OLFS component results in fourth and fifth column panels. The correlation parameters within joint model specification are presented in the last row panel of Table 3.

5.2.1 Crash Count Component (NB Model)

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.

Sociodemographic Characteristics: Several sociodemographic characteristics considered are found to be significant determinants of motor vehicle crash risk at the zonal level. As expected, we find that as HH density increases, the risk of motorized vehicle crashes also increases. The result is in accordance with the finding from Hadayeghi et al. (2003). Dependence (a surrogate for non-working group of population) has significant negative impact on motor vehicle crashes risk, perhaps indicating a lower exposure to motor vehicles for this group of population. Increased proportion of female population at the zonal level is negatively associated with increased number of motor vehicle crashes. The result can be explained by overall low risk taking attitude of this group of population (Stamatiadis and Puccini, 2000). The estimation results also indicate that STAZs with greater proportions of Caucasian and Asian population are likely to experience less motor vehicle crashes. On the other hand, motor vehicle crash risk increases with increasing proportion of Hispanic population, a result also observed in Lee et al. (2014).

Socioeconomic Characteristics: In terms of commuters by mode choice, total number of bike and walk commuters are found to have significant impact in NB component of the joint model system. As expected, increases in both of these variables reduce the likelihood of motorized vehicle crash risk at the zonal level. The findings are contradictory to Rifaat et al. (2010), perhaps indicative of the difference in contexts - Rifaat et al. (2010) examined total number of crashes whereas in our study we examine total crashes involving motor vehicles only. The result associated with

---

[6] Estimation results of independent NB-OLFS and joint NB-OLFS without correlation parameterization are presented in Table A and B, respectively, in the APPENDIX section.

employment density reflects that an increase in employment density increases the likelihood of motor vehicle crash risk (see Khondakar et al., 2010 for similar result).

Built Environment: In NB component of the joint model system, a higher proportion of urban area results in higher motor vehicle crash risk, plausibly indicating higher interactions of vehicles and in turn, higher vehicular conflicts within an urbanized road environment. Among different point of interests considered, the results reveal that motor vehicle crashes are negatively associated with higher number of law enforcement offices, parks and recreational centers and transportation hubs. Further, the results reveal that the presence of more restaurants in a STAZ is positively associated with motor vehicle crash risk. Among land use characteristics, land use mix and proportion of retail and office area have significant impact on the crash count component. Interestingly, crashes are negatively associated with higher land use mix in a zone. The result shows that STAZs with higher proportion of retail and office area have higher likelihood of motorized vehicular crash risk (see Ng et al, 2002; Rifaat et al, 2009; Rifaat et al, 2010 for similar results).

Transport Infrastructure: In crash count component, proportion of major roads has significant impact on motor vehicle crash risk. We find that in presence of more major roads in a STAZ, the possibility of crash risk decreases. As explained in prior literature (Huang et al., 2010), the result can be explained by better road design of major roads. An increase in traffic signal density in a STAZ increases the likelihood of motorized vehicular crash risk. With respect to intersection density, the model estimation result indicates an expected positive correlation of higher intersection density with motorized vehicular crashes, a result also observed in several previous studies (Jiang et al., 2016; Abdel-Aty et al., 2013).

Traffic Characteristics: With respect to the traffic characteristics, none of the variables are found to affect motor vehicle crash risk at the zonal level. The reader should note that we considered VMT in the count propensity component of the joint model. However, the model estimation offered a statistically insignificant parameter. In the macro-level model for our study area, it is possible that the influence of VMT is represented by other attributes that serve as surrogates for VMT (such as household density).

5.2.2 Crash Proportion Component (OLFS Model)

In OLFS model, the positive (negative) coefficient corresponds to increased (decreased) proportion for severe injury categories.

Sociodemographic Characteristics: From Table 3, we can see that HH density is highly significant in the crash severity proportion component. As expected, the variable has an opposing effect in fractional split component than crash count component of the joint model. HH density has negative impact on proportion of crash severity outcomes implying a reduced likelihood of more severe crashes. We find that crash proportion for severe outcome levels is lower in the STAZs with higher proportion of Caucasian, Hispanic and African-American population.

Socioeconomic Characteristics: The results for the number of commuters based on different commute modes reveal that STAZs with higher number of automobile commuters increase the likelihood of more severe crashes. The result associated with transit commuter reflects lower probability of severe crash proportions. Higher number of workers working from home is negatively associated with more severe crash proportions. The result can be explained by overall lower exposure of this group of people to traffic (Abdel-Aty et al., 2013). In our joint model specification, employment density has significant impact in OLFS model component as well. From the model estimates we find that the likelihood of higher proportion of severe crash outcomes decreases with increasing employment density. Proportion of population below poverty level, an indicator for area deprivation, reveals positive impact on proportion of crashes by severity levels. Huang et al. (2010) and Aguero-Valverde and Jovanis (2006) also found a similar impact in examining the impact of variables on severe crashes.

Built Environment: As found in previous studies (Noland and Quddus, 2004), we also find that the possibility of more severe crashes decreases with increasing share of urbanized area of a STAZ, presumably due to the congested and/or slower traffic on roadways of urbanized environment. With respect to point of interests, number of law enforcement offices, restaurants, parks and recreational centers and shopping centers are negatively associated with crash severity proportions. In the OLFS component, the result for proportion of industrial land use category reveals that STAZs with higher share of industrial land use increase the likelihood of higher severe crash proportions. A similar positive relationship between industrial land use and severe crash occurrence is documented by Hadayeghi et al. (2007).

Transport Infrastructure: The only transport infrastructure variables influencing motor vehicle crash risk proportions is the zonal level length of local roads. Crash severity proportions are negatively associated with higher length of local roads.

Traffic Characteristics: Several traffic characteristics considered are found to be significant determinants of crash proportions by severity levels. Among traffic characteristics, crash proportion of severe crashes is found to be higher for STAZs with higher vehicle miles travelled (VMT). The result is in line with several previous studies and can be attributable to higher exposure and /or adaptation of drivers to different levels of traffic volume (see Milton et al., 2008; Dong et al., 2014; Lee et al., 2014; and Hadayeghi et al., 2003 for similar results). The OLFS model results reveal higher proportion of severe crash outcomes for higher proportion of heavy vehicular miles travelled at the STAZ level, consistent with earlier research findings (Li et al., 2013). Average zonal speed limit is found to be a significant determinant of crash proportion by severity outcomes. The estimate for average speed has a positive coefficient suggesting that proportion of severe crashes increases with increasing zonal level average speed. In line with findings from previous studies (Hadayeghi et al., 2003; Li et al., 2013), we find that higher traffic intensity (defined as the ratio of VMT and total length of roadways in miles) decreases the possibility of higher proportions of severe crash outcomes, attributable to lower travel speed of motor vehicles in more congested roadway environment.

5.2.3 Unobserved Effects

Significance of the unobserved heterogeneity parameters presented in the last row panel of Table 3 highlights the presence of common unobserved factors affecting crash count and crash severity proportion components. As indicated earlier, we parameterize the correlation profile as a function of observed exogenous variables. In terms of exogenous variables, we find that the correlation between the two dimensions of the joint model system is moderated by the proportion of heavy vehicle miles travelled. This provides support to our hypothesis that the correlation is not constant across the entire database. Both the constant and proportion of heavy vehicle miles travelled are introduced with a $"-"$ sign before η_i in the crash proportion component (as described in econometric framework section) since it was the expected effect and also provided a substantially better fit compared to introducing them with a $"+"$ sign.

5.3 Predictive Performance Evaluation

In order to demonstrate the predictive performance of the estimated models, we also perform computation of several in-sample goodness-of-fit measures. In doing so, performance of joint NB-OLFS model with correlation parameterization is compared with the predictive performance of independent NB-OLFS for verifying the improvement of incorporating correlation in estimating crash count and crash severity proportions simultaneously. To evaluate the in-sample predictive performance, we employ three different fit measures: mean absolute error (MAE), mean percentage error (MPE) and mean absolute percentage error (MAPE)[7]. These fit measures quantify the error associated with model predictions and the model with lower fit measures provides better predictions of the observed data. We compute these measures at the disaggregate level by generating measures at the study unit level (STAZ) and compute the average measures across all units.

Table 4 presents the values for these measures for independent NB-OLFS and joint NB-OLFS model with correlation parameterization. Other than total crash counts and crash proportions across different severity levels, from the estimated joint models, we can also generate crash counts by severity levels by using equations 2 and 5 as follows:

$$E(\boldsymbol{k}_i) = \mu_i * \Lambda(y_{ik} = k) \quad (10)$$

[7] These measures can be computed as $MAE = \frac{\sum_{n=1}^{N}(\hat{y}_n - y_n)}{N}, MPE = \sum_{n=1}^{N}(\frac{\hat{y}_n - y_n}{y_n})\ and\ MAPE = \sum_{n=1}^{N}\left|\frac{\hat{y}_n - y_n}{y_n}\right|$, where, $\hat{y}_n$ and y_n are the predicted and observed values across different study units n (n = 1,1,2,…8518).

where, $E(\boldsymbol{k}_i)$ is the expected number of crashes by injury severity k for STAZ i. Thus the framework allows us to predict total crash counts, proportion of crash counts by crash severity levels and crash counts across different severity levels within a single econometric framework. In evaluating predictive performance, we compute errors in crash predictions for total crash counts (from NB component), crash severity proportions (from OLFS component) and counts for different severity levels (from joint distribution). It is worthwhile to recognize here that in independent NB-OLFS model $\eta_i = 0$ in equation 2 and 5, while in joint NB-OLFS model with correlation parameterization η_i is estimated by using structure as presented in equation 6. The resulting fit measures for comparing the predictive performance clearly indicate that overall the joint NB-OLFS with correlation parameterization model offers superior fit compared to the independent NB-OLFS model. The independent model performs marginally better than the joint model in OLFS component with respect to MAE and MAPE, while the joint model provides superior predictions across all other fit measures. These prediction results further confirm the benefit of accommodating correlation and heterogeneity in modeling crash counts and crash severity proportions at a zonal level.

5.4 Model Implications

The model results and performance evaluation from the previous sections clearly highlight the value of the proposed joint NB-OLFS model. The model findings have important implications in terms of countermeasures for zones with higher number of crashes. Moreover, the findings can be used to identify zones with greater risk of severe crashes. To illustrate the model applicability, we employ the model results to plot the spatial distribution of predicted motor vehicle crash frequency and predicted crash counts by severity levels (calculated by using equation 10). These plots are presented in Figure 1. The reader would note that the Figure also identifies major urban regions in Florida - Tallahassee, Jacksonville, Orlando, Tampa and Miami. From Figure 1, we can see that STAZs with higher number of total crashes are also in general associated with higher crash counts across different severity levels. Further, the figures indicate that high crash risk zones are dispersed throughout the state with visible clustering. From spatial representation, we can also observe that zones with higher number of crashes are, in general, close to the major cities. This spatial illustration can easily be used to prioritize STAZs based on total crash risk and crash risk across different severity levels in enhancing motor vehicle safety of these high crash risk zones.

6. CONCLUSIONS

The paper proposed, formulated and estimated an innovative joint econometric framework for examining total crash count and crash proportion by different crash attribute levels (such as crash severity, different crash types or different road user groups involved in crashes). Specifically, we proposed to consider a crash frequency model for total crashes in conjunction with a fractional split model that considers proportion by crash attribute levels. The model ties total crash counts and crash proportions by accommodating for the potential common unobserved heterogeneity (across study unit) in the joint framework. To the best of the authors' knowledge, this is the first attempt to employ such a joint framework for examining count events.

In this study, we demonstrated the application of the proposed approach by employing a Negative Binomial-Ordered Logit Fractional Split (NB-OLFS) model framework. We also allowed the potential unobserved heterogeneity to vary across study units in the joint framework. The empirical analysis was conducted by using zonal level crash count data for different crash severity levels from the state of Florida for the year 2015. The models were estimated employing a comprehensive set of exogenous variables − sociodemographic characteristics, socioeconomic characteristics, built environment, transport infrastructure and traffic characteristics. The empirical analysis involved estimation of three different model systems: 1) an independent Negative Binomial (NB) and Ordered Logit Fractional Split (OLFS) model system, 2) joint NB-OLFS model without correlation parameterization and 3) joint NB-OLFS model with correlation parameterization. The comparison exercise, based on information criterion metrics, highlighted the superiority of the joint model with the correlation parameterization in terms of data fit. According to our results, the impacts of exogenous variables (in sign) between two components of the joint model were different for some variables. An in-sample validation exercise is conducted to compare the performance of the joint NB-OLFS model with correlation parameterization to the performance of the independent NB-OLFS model. The prediction results clearly highlight the superior performance of the joint model. To further illustrate the model applicability, we employed the model results to plot the spatial distribution of predicted motor vehicle crash frequency and predicted crash counts by severity levels.

The paper is not without limitations. In our research effort, we employed aggregate level crash count data at a zonal level. However, we have not explored spatial correlation across different zones. It will be an interesting exercise to model the impact of spatial correlation across zones. Moreover, it might be interesting to explore the transferability of models developed for crash count and crash severity simultaneously by estimating similar models for multiple spatial units and several years.

Acknowledgements:

The authors would also like to gratefully acknowledge Signal Four Analytics (S4A) for providing access to Florida crash data.

References

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: Policy and Practice 49, 62-75.

Abdel-Aty, M., Siddiqui, C., Huang, H., Wang, X., 2011. Integrating trip and roadway characteristics to manage safety in traffic analysis zones. Transportation Research Record: Journal of the Transportation Research Board 2213 (1), 20-28.

Aguero-Valverde, J., Jovanis, P., 2009. Bayesian multivariate poisson lognormal models for crash severity modeling and site ranking. Transportation Research Record: Journal of the Transportation Research Board (2136), 82-91.

Aguero-Valverde, J., Jovanis, P.P., 2006. Spatial analysis of fatal and injury crashes in pennsylvania. Accident Analysis & Prevention 38 (3), 618-625.

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 & Prevention 87, 8-16.

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.

Anastasopoulos, P.C., Shankar, V.N., Haddock, J.E., Mannering, F.L., 2012a. A multivariate tobit analysis of highway accident-injury-severity rates. Accident Analysis & Prevention 45, 110-119.

Aptech 2015, Aptech Systems Inc, accessed from http://www.aptech.com/ on September 19th 2015.

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.

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.

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

Brijs, T., Karlis, D., Van Den Bossche, F., Wets, G., 2007. A bayesian model for ranking hazardous road sites. Journal of the Royal Statistical Society: Series A (Statistics in Society) 170 (4), 1001-1017.

Cai, Q., Lee, J., Eluru, N., Abdel-Aty, M., 2016. Macro-level pedestrian and bicycle crash analysis: Incorporating spatial spillover effects in dual state count models. Accident Analysis & Prevention 93, 14-22.

Chamberlain, G., 1980. Analysis of covariance with qualitative data. The Review of Economic Studies 47 (1), 225-238.

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 & Prevention 99, Part A, 330-341.

Chiou, Y.-C., Fu, C., 2013. Modeling crash frequency and severity using multinomial-generalized poisson model with error components. Accident Analysis & Prevention 50, 73-82.

Chiou, Y.-C., Fu, C., 2015. Modeling crash frequency and severity with spatiotemporal dependence. Analytic Methods in Accident Research 5, 43-58.

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.

Dong, C., Clarke, D.B., Nambisan, S.S., Huang, B., 2016. Analyzing injury crashes using random-parameter bivariate regression models. Transportmetrica A: Transport Science 12 (9), 794-810.

Dong, C., Clarke, D.B., Yan, X., Khattak, A., Huang, B., 2014. Multivariate random-parameters zero-inflated negative binomial regression model: An application to estimate crash frequencies at intersections. Accident Analysis & Prevention 70, 320-329.

Dong, C., Nambisan, S.S., Xie, K., Clarke, D.B., Yan, X., 2017. Analyzing the effectiveness of implemented highway safety laws for traffic safety across us states. Transportmetrica A: Transport science 13 (2), 91-107.

El-Basyouny, K., Sayed, T., 2009. Collision prediction models using multivariate poisson-lognormal regression. Accident Analysis & Prevention 41 (4), 820-828.

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

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 & Prevention 40 (3), 1033-1054.

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 & Prevention 59, 125-134.

FDOT and CUTR, 2015. Florida Transportation Trends & Conditions, Florida Department of Transportation Office of Policy Planning (FDOT) with support from Center for Urban Transportation Research (CUTR).

Hadayeghi, A., Shalaby, A., Persaud, B., 2003. Macrolevel accident prediction models for evaluating safety of urban transportation systems. Transportation Research Record: Journal of the Transportation Research Board (1840), 87-95.

Hadayeghi, A., Shalaby, A., Persaud, B., 2007. Safety prediction models: Proactive tool for safety evaluation in urban transportation planning applications. Transportation Research Record: Journal of the Transportation Research Board (2019), 225-236.

Heydari, S., Fu, L., Miranda-Moreno, L.F., Jopseph, 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.

Hosseinpour, M., Yahaya, A.S., Sadullah, A.F., 2014. Exploring the effects of roadway characteristics on the frequency and severity of head-on crashes: Case studies from malaysian federal roads. Accident Analysis & Prevention 62, 209-222.

Huang, H., Abdel-Aty, M., Darwiche, A., 2010. County-level crash risk analysis in florida: Bayesian spatial modeling. Transportation Research Record: Journal of the Transportation Research Board (2148), 27-37.

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, H., Zhou, H., Wang, J., Chang, F., Ma, M., 2017. A multivariate spatial model of crash frequency by transportation modes for urban intersections. Analytic Methods in Accident Research 14, 10-21.

Jiang, X., Abdel-Aty, M., Hu, J., Lee, J., 2016. Investigating macro-level hotzone identification and variable importance using big data: A random forest models approach. Neurocomputing 181, 53-63.

Khondakar, B., Sayed, T., Lovegrove, G., 2010. Transferability of community-based collision prediction models for use in road safety planning applications. Journal of Transportation Engineering 136 (10), 871-880.

Ladron De Guevara, F., Washington, S., Oh, J., 2004. Forecasting crashes at the planning level: Simultaneous negative binomial crash model applied in tucson, arizona. Transportation Research Record: Journal of the Transportation Research Board (1897), 191-199.

Lee, J., Abdel-Aty, M., Jiang, X., 2014. Development of zone system for macro-level traffic safety analysis. Journal of Transport Geography 38 (0), 13-21.

Lee, J., Abdel-Aty, M., Jiang, X., 2015. Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level. Accident Analysis & Prevention 78, 146-154.

Lee, J., Mannering, F., 2002. Impact of roadside features on the frequency and severity of run-off-roadway accidents: An empirical analysis. Accident Analysis & Prevention 34 (2), 149-161.

Lee, J., Yasmin, S., Eluru, N., Abdel-Aty, M., Cai, Q., 2016. Macroscopic analysis of crash proportion by mode: fractional split multinomial logit modeling approach. Presented in the 95th Annual Meeting of the Transportation Research Board (TRB) of the National Academies, January 10-14, 2016, Washington, D.C., USA.

Li, Z., Wang, W., Liu, P., Bigham, J.M., Ragland, D.R., 2013. Using geographically weighted poisson regression for county-level crash modeling in california. Safety science 58, 89-97.

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

Ma, J., Kockelman, K., 2006. Bayesian multivariate poisson regression for models of injury count, by severity. Transportation Research Record: Journal of the Transportation Research Board (1950), 24-34.

Ma, J., Kockelman, K.M., Damien, P., 2008. A multivariate poisson-lognormal regression model for prediction of crash counts by severity, using bayesian methods. Accident Analysis & Prevention 40 (3), 964-975.

Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: Methodological frontier and future directions. Analytic Methods in Accident Research 1, 1-22.

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.C., Shankar, V.N., Mannering, F.L., 2008. Highway accident severities and the mixed logit model: An exploratory empirical analysis. Accident Analysis & Prevention 40 (1), 260-266.

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.

Naderan, A., Shahi, J., 2010. Aggregate crash prediction models: Introducing crash generation concept. Accident Analysis & Prevention 42 (1), 339-346.

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: Journal of the Transportation Research Board (2601), 119-127.

Ng, K.-S., Hung, W.-T., Wong, W.-G., 2002. An algorithm for assessing the risk of traffic accident. Journal of Safety Research 33 (3), 387-410.

Noland, R.B., Quddus, M.A., 2004. A spatially disaggregate analysis of road casualties in england. Accident Analysis & Prevention 36 (6), 973-984.

Papke, L.E., 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.

Park, E., Lord, D., 2007. Multivariate poisson-lognormal models for jointly modeling crash frequency by severity. Transportation Research Record: Journal of the Transportation Research Board (2019), 1-6.

Park, E.S., Park, J., Lomax, T.J., 2010. A fully bayesian multivariate approach to before–after safety evaluation. Accident Analysis & Prevention 42 (4), 1118-1127.

Pei, X., Wong, S., Sze, N.-N., 2011. A joint-probability approach to crash prediction models. Accident Analysis & Prevention 43 (3), 1160-1166.

Rifaat, S., Tay, R., De Barros, A., 2010. Effect of street pattern on road safety: Are policy recommendations sensitive to aggregations of crashes by severity? Transportation Research Record: Journal of the Transportation Research Board (2147), 58-65.

Rifaat, S., Tay, R., Perez, A., Barros, A.D., 2009. Effects of neighborhood street patterns on traffic collision frequency. Journal of Transportation Safety & Security 1 (4), 241-253.

Roshandeh, A.M., Agbelie, B.R., Lee, Y., 2016. Statistical modeling of total crash frequency at highway intersections. Journal of traffic and transportation engineering (English edition) 3 (2), 166-171.

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.

Shin, K., Washington, S.P., 2012. Empirical bayes method in the study of traffic safety via heterogeneous negative multinomial model. Transportmetrica 8 (2), 131-147.

Song, J.J., Ghosh, M., Miaou, S., Mallick, B., 2006. Bayesian multivariate spatial models for roadway traffic crash mapping. Journal of Multivariate Analysis 97 (1), 246-273.

Stamatiadis, N., Puccini, G., 2000. Socioeconomic descriptors of fatal crash rates in the southeast USA. Injury control and safety promotion 7 (3), 165-173.

Tunaru, R., 2002. Hierarchical bayesian models for multiple count data. Austrian Journal of statistics 31 (3), 221-229.

Wang, C., Quddus, M.A., Ison, S.G., 2011. Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model. Accident Analysis & Prevention 43 (6), 1979-1990.

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 & Prevention 99, Part A, 6-19.

Washington, S.P., Karlaftis, M.G., Mannering, F., 2010. Statistical and econometric methods for transportation data analysis CRC press.

Wei, F., Lovegrove, G., 2013. An empirical tool to evaluate the safety of cyclists: Community based, macro-level collision prediction models using negative binomial regression. Accident Analysis and Prevention 61, 129-137.

Xu, X., Wong, S., Choi, K., 2014. A two-stage bivariate logistic-tobit model for the safety analysis of signalized intersections. Analytic Methods in Accident Research 3, 1-10.

Yasmin, S., Eluru, N., 2016. Latent segmentation based count models: Analysis of bicycle safety in montreal and toronto. Accident Analysis & Prevention, Volume 95, Part A, October 2016, Pages 157-171.

Yasmin, S., Eluru, N., Lee, J., Abdel-Aty, M.A., 2016. An ordered fractional split approach for aggregate injury severity modeling. Transportation Research Record Vol. 2583, pp. 119-126.

Yasmin. S., and N. Eluru (2013), "Evaluating Alternate Discrete Outcome Frameworks for Modeling Crash Injury Severity," Accident Analysis & Prevention, 59 (1), pp. 506-521

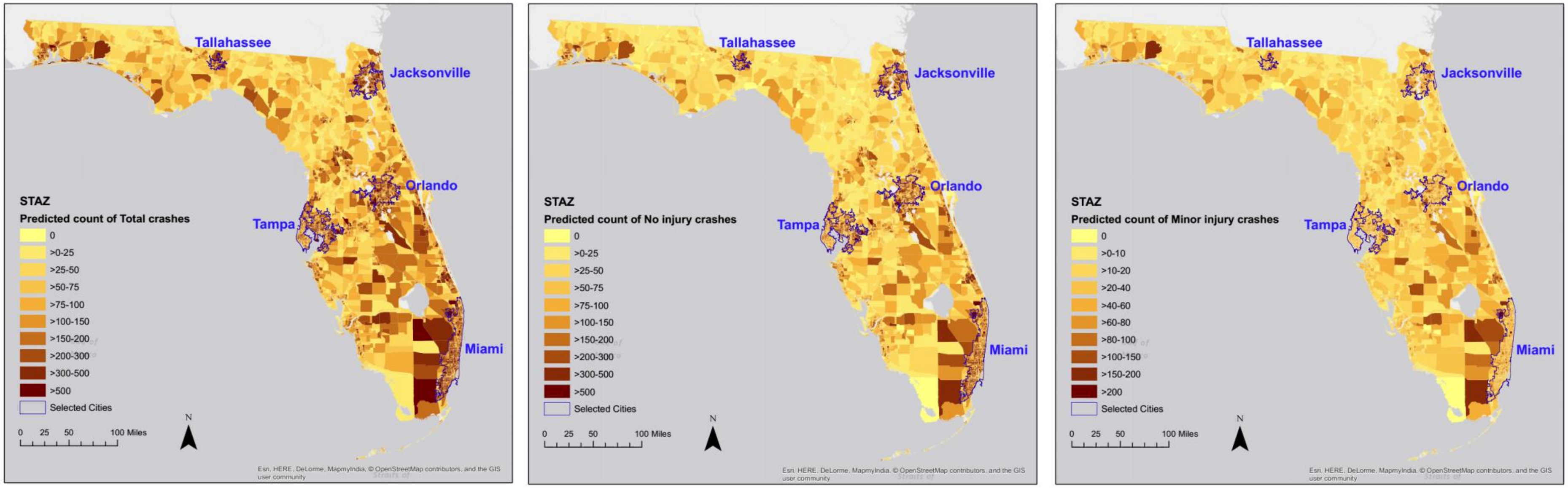
Ye, X., Pendyala, R.M., Shankar, V., Konduri, K.C., 2013. A simultaneous equations model of crash frequency by severity level for freeway sections. Accident Analysis & Prevention 57, 140-149.

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., 2017. A multivariate random-parameters tobit model for analyzing highway crash rates by injury severity. Accident Analysis & Prevention 99, 184-191.

Zhan, X., Aziz, H.M.A., Ukkusuri, S.V., 2015. An efficient parallel sampling technique for multivariate poisson-lognormal model: Analysis with two crash count datasets. Analytic Methods in Accident Research 8, 45-60.

FIGURE 1 Spatial Distribution of Expected Motor Vehicle Crash Frequency for Total Crash Counts and Counts by Severity Levels



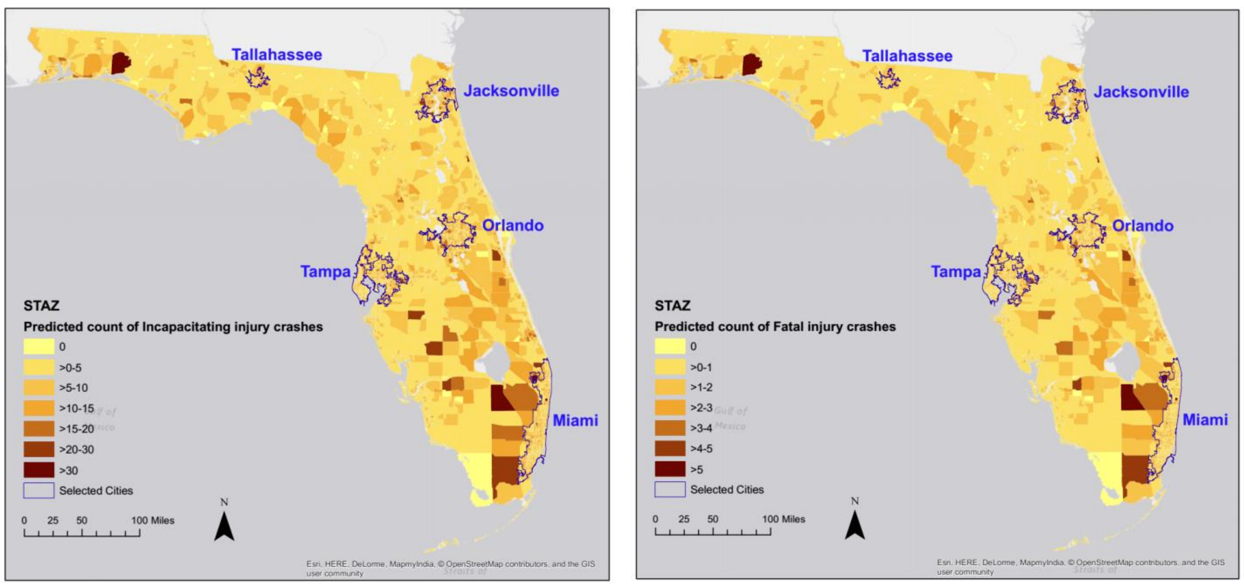


TABLE 1 Summary of Existing Multivariate Crash Frequency Studies

Studies	Study Unit (Scale)	Methodology	Dependent Variables Analyzed	Number of Dimensions Examined
(Ye *et al.* 2013)	Multilane freeway segment (Micro)	Joint Poisson regression model	Crash frequencies by severity level - property damage only, possible injury, and injury/fatality	3
(Aguero-Valverde and Jovanis 2009)	State-maintained rural two-lane roads (Micro)	Full Bayes multivariate Poisson lognormal models	Crash frequencies by severity level - Fatalities, major injuries, moderate injuries, minor injuries, and PDO	5
(Tunaru 2002)	Single carriageway link sites (Micro)	Bayesian Multivariate Poisson-log Normal Model	Crash frequencies by severity level – slight injury, serious/fatal injury by number of vehicles involved (1 and 2+)	4
(Ladron de Guevara *et al.* 2004)	Traffic analysis zone (Macro)	Simultaneous negative binomial model	Crash frequencies by severity level – no injury, injury and fatal	3
(Song *et al.* 2006)	District (Macro)	Bayesian multivariate conditional autoregressive model	Crash frequencies by location – intersection crash, intersection-related crash, driveway-access crash, and non-intersection crash	4
(Ma and Kockelman 2006)	Highway segments (Micro)	Multivariate Poisson regression model	Count of victims by severity type - fatal, disabling injury, non-disabling injury, possible injury, and non-injury	5
(Park and Lord 2007)	Intersection (Micro)	Multivariate Poisson–lognormal Model	Crash frequencies by severity level – fatal, incapacitating-injury, non-incapacitating injury, minor injury, property damage only	5
(Ye *et al.* 2009)	Intersection (Micro)	Multivariate Poisson regression model	Crash frequencies by crash types – head-on, rear-end, sideswipe (same direction), sideswipe (opposite direction) and pedestrian-involved crashes	6
(Ma *et al.* 2008)	Road segment (Micro)	Bayesian multivariate Poisson log-normal model	Count of victims by severity type - fatal, disabling injury, non-disabling injury, possible injury, and non-injury	5
(Aguero-Valverde *et al.* 2016)	Roadway segment (Micro)	Multivariate Poisson log-normal spatial model	Crash frequencies by crash types – same direction, opposite direction, angle and hit-fixed object crashes	4
(Dong *et al.* 2014)	Intersection (Micro)	Multivariate random-parameters zero-inflated negative binomial model	Crash frequencies by vehicles involved – car only crash, car-truck crash and truck only crash	3
(El-Basyouny and Sayed 2009)	Intersection (Micro)	Multivariate Poisson log-normal regression model	Crash frequencies by crash severity levels – PDO and injury/fatal crashes	2

(Song *et al.* 2006)	District (Macro)	Multivariate conditional autoregressive (CAR) models	Crash frequencies by roadway locations – intersection, intersection-related, driveway access and non-intersection locations	4
(Anastasopoulos *et al.* 2012a)	Roadway segments (Micro)	Multivariate tobit regression	Rates of crashes (per distance travelled) by crash severity levels - no-injury, possible injury and injury crashes	3
(Park *et al.* 2010)	Roadway segment (Micro)	Fully Bayesian multivariate Poisson regression model	Crash frequencies by different crash characteristics – total crash, speed related crash and crashes for different severity and situational characteristics	4
(Brijs *et al.* 2007)	Intersection (Micro)	Bayesian multivariate Poisson regression model	Crash frequencies by crash outcome levels – total crashes, fatal crashes and slight/serious injury crashes	3
(Barua *et al.* 2016)	Road segment (Micro)	Bayesian multivariate random parameters spatial model	Crash frequencies by severity levels – no injury and injury/fatal crashes	2
(Anastasopoulos 2016)	Roadway segment (Micro)	Random parameter multivariate tobit model, Multivariate zero-inflated negative binomial model	Crash frequency and crash rate (per distance travelled) by severity type – PDO, injury and fatality	3
(Mothafer *et al.* 2016)	Multilane freeway segment (Micro)	Multivariate Poisson gamma mixture count model	Crash frequency by crash type – rear-end, sideswipe, fixed object and other collision types (same direction, overturn, head-on, and miscellaneous type)	4
(Serhiyenko *et al.* 2016)	Limited access highway segment (Micro)	multivariate Poisson Lognormal model	Crash frequency by crash type – single vehicle, same direction and opposite direction crashes	3
(Huang *et al.* 2017)	Urban intersection (Micro)	Multivariate spatial conditional autoregressive (CAR) models	Crash frequency by travel mode – pedestrian, bicycle and motor vehicle	3
(Barua *et al.* 2014)	Road segment (Micro)	Multivariate Poisson lognormal model	Crash frequency by crash severity – no injury and injury/fatal crashes	2
(Zhan *et al.* 2015)	Census tract (Macro) Roadway segment (Micro)	Multivariate Poisson-lognormal model	Crash frequency of pedestrian-vehicle crashes by severity levels – fatal and severe injury crashes Crash frequency by crash severity – no injury, possible injury and evident injury	2, 3
(Zeng *et al.* 2017)	Road segment (Micro)	Multivariate random parameter tobit model	Crash frequency by severity levels – slight injury crash and killed/seriously injured crashes	2
(Heydari *et al.* 2017)	Intersection (Micro)	Bayesian latent class flexible mixture multivariate model	Crash frequency by crash type – pedestrian and bicycle crashes	2
(Wang *et al.* 2017)	Roadway segment and intersections (Micro)	Integrated Nested Laplace Approximation Multivariate Poisson Lognormal model	Crash frequency by crash types –same-direction, intersection-direction, opposite direction and single vehicle crashes	4, 3

			Crash frequency by severity outcomes – no injury, possible/non-incapacitating injury and fatal/incapacitating injury crashes	
(Nashad *et al.* 2016)	Statewide Traffic Analysis Zone (Macro)	Copula based bivariate negative binomial model	Crash frequency by crash type – pedestrian and bicycle crashes	2
(Cheng *et al.* 2017)	Intersection (Micro)	Multivariate Poisson lognormal temporal and spatial models	Crash frequency by crash type - Rear-end, Head-on, Side-swipe, Broad-side, Hit object, and Others crashes	6
(Dong *et al.* 2016)	Intersection (Micro)	Random parameter bivariate zero-inflated negative binomial model	Crash frequency by severity – disabling injury and non-disabling injury	2

TABLE 2 Sample Statistics for the State of Florida

Variable Names	Definition	Zonal		
		Minimum	**Maximum**	**Average**
Dependent Variables				
Count of total crashes	Total number of crashes in STAZ	0.000	864.000	58.092
Proportion of no injury crashes	Total number of no injury crashes in STAZ/ Total number of crashes in STAZ	0.000	1.000	0.693
Proportion of minor injury crashes	Total number of minor injury crashes in STAZ/ Total number of crashes in STAZ	0.000	1.000	0.218
Proportion of incapacitating injury crashes	Total number of incapacitating injury crashes in STAZ/ Total number of crashes in STAZ	0.000	1.000	0.034
Proportion of fatal crashes	Total number of fatal injury crashes in STAZ/ Total number of crashes in STAZ	0.000	1.000	0.006
Sociodemographic Characteristics				
HH density	Ln(Number of HH in STAZ/Total area of STAZ in square miles)	-9.816	10.182	5.619
Dependence	Ratio of youth (15 years or younger) and elderly (65 years or more) to working age persons	0.000	12.622	0.607
Proportion of female population	Number of female residents in STAZ/Total number of population in STAZ	0.000	0.834	0.500
Proportion of Caucasian population	Number of Caucasian residents in STAZ/Total number of population in STAZ	0.000	1.000	0.759
Proportion of Asian population	Number of Asian residents in STAZ/Total number of population in STAZ	0.000	0.508	0.020
Proportion of Hispanic population	Number of Hispanic residents in STAZ/Total number of population in STAZ	0.000	1.000	0.171
Proportion of African - American population	Number of African - American population residents in STAZ/Total number of population in STAZ	0.000	1.000	0.172
Socioeconomic Characteristics				
Automobile commuters	Ln(Total passenger vehicle commuters in STAZ)	-11.428	9.847	5.325
Transit commuters	Ln(Total public transit commuters in STAZ)	-18.900	6.928	0.095
Walk commuters	Ln(Total walk commuters in STAZ)	-21.362	7.024	0.100
Bike commuters	Ln(Total bike commuters in STAZ)	-20.711	6.407	-0.360
Number of workers that work from home	Ln(Total number of workers that worked from home in STAZ)	-17.688	8.414	1.640

	Employment density	Total number of jobs in STAZ/Total number of population in STAZ	-10.514	11.386	-0.673
	Population below poverty level	Total number of population below poverty level in STAZ/Total number of population in STAZ	0.000	0.790	0.182
Built Environment					
	STAZ area	Ln(Total area of STAZ in square miles)	-18.517	6.786	-0.171
	Proportion of Urban area	Urban area in STAZ/Total area of STAZ	0.000	1.000	0.731
	Law enforcement offices	Number of law enforcement offices in STAZ	0.000	4.000	0.116
	Restaurants	Count of restaurants in STAZ/10	0.000	11.000	0.410
	Park and recreational centers	Count of park and recreational centers/10	0.000	5.400	0.079
	Transportation hubs	Count of transportation hubs/10	0.000	5.200	0.016
	Shopping centers	Count of shopping centers/10	0.000	18.900	0.583
	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 a STAZ	0.000	0.859	0.046
	Proportion of retail and office land use	Retail and office land use in STAZ/Total area of STAZ	0.000	0.786	0.008
	Proportion of industrial land use	Industrial land use in STAZ/Total area of STAZ	0.000	0.871	0.002
Transport Infrastructure					
	Local roads	Ln(Length of local roads in STAZ in meter)	-4.240	11.713	1.935
	Major roads	Ln(Length of major roads in STAZ in meter)	-4.036	11.155	6.288
	Traffic signal density	Total number of traffic signal in STAZ/Total roads length in STAZ in miles	-4.144	6.840	-0.078
	Intersection density	Ln(Total number of intersections/Total roads length in STAZ in miles)	-2.638	8.464	1.815
Traffic Characteristics					
	Vehicle miles travelled (VMT)	Ln(Total vehicle miles travelled in STAZ)	0.000	13.524	9.442
	Proportion of heavy vehicle miles travelled	Heavy vehicle miles travelled in STAZ/Total vehicle miles travelled in STAZ	0.000	0.848	0.037
	Average speed	Ln(Average posted speed limit in square miles per hour in STAZ)	0.000	4.248	3.390
	Traffic intensity	Vehicle miles travelled/Total length of roadways in miles in STAZ	0.000	13.094	8.798

TABLE 3 Joint NB-OLFS with Correlation Parameterization Model Results for Florida

Variable Names	NB Model Component		OLFS Model Component	
	Estimate	t-stat	Estimate	t-stat
Constant	1.393	7.234	---	---
Threshold Parameters	---	---	---	---
Threshold between no injury and minor injury	---	---	-0.121	-0.571
Threshold between minor and incapacitating injury	---	---	2.058	9.660
Threshold between incapacitating and fatal injury	---	---	3.992	18.003
Sociodemographic Characteristics				
HH density	0.476	17.313	-0.049	-4.117
Dependence	-0.312	-7.423	---	---
Proportion of female population	-1.263	-3.419	---	---
Proportion of Caucasian population	-0.948	-12.200	-1.021	-5.531
Proportion of Asian population	-2.243	-3.901	---	---
Proportion of Hispanic population	1.038	13.718	-0.541	-10.692
Proportion of African-American population	---	---	-1.067	-5.632
Socioeconomic Characteristics				
Automobile commuters	---	---	0.039	3.149
Transit commuters	---	---	-0.007	-2.200
Walk commuters	-0.041	-6.045	---	---
Bike commuters	-0.023	-3.440	---	---
Number of workers that work from home	---	---	-0.011	-2.242
Employment density	0.317	17.713	-0.033	-3.645
Population below poverty level	---	---	0.310	3.292
Built Environment				
Proportion of Urban area	1.360	13.084	-0.239	-5.201

Law enforcement offices	-0.126	-3.996	-0.084	-4.592
Restaurants	0.157	7.647	-0.093	-5.798
Park and recreational centers	-0.444	-5.287	-0.202	-3.645
Transportation hubs	-0.398	-4.326	---	---
Shopping centers	---	---	-0.021	-2.349
Land use mix	-0.907	-10.541	---	---
Proportion of retail and office land use	1.623	4.849	---	---
Proportion of industrial land use	---	---	0.602	2.476
Transport Infrastructure				
Local roads	---	---	-0.018	-5.601
Major roads	-0.041	-5.653	---	---
Traffic signal density	0.338	12.448	---	---
Intersection density	0.343	10.179	---	---
Traffic Characteristics				
Vehicle miles travelled	---	---	0.053	3.929
Proportion of heavy vehicle miles travelled	---	---	0.432	1.968
Average speed	---	---	0.064	3.792
Traffic intensity	---	---	-0.060	-4.256
Dispersion parameter	1.728	53.801	---	---

Correlation Parameters

Variables	**Estimate**	**t-stat**
Constant	0.144	2.559
Proportion of heavy vehicle miles travelled	1.755	2.922

TABLE 4 In-Sample Predictive Performance Evaluation

Models	Fit measures	Error in					
		Total counts	Fractions	Counts for different severity levels			
		NB component	OLFS component	No injury	Minor injury	Incapacitating injury	Fatal injury
Independent NB-OLFS	MAE	570.783	0.255	426.927	122.356	18.620	3.367
	MPE	13.196	0.248	12.249	12.648	7.299	0.605
	MAPE	13.385	0.638	12.420	12.797	7.405	0.648
Joint NB-OLFS with correlation parameterization	MAE	532.401	0.262	403.121	110.151	16.638	3.021
	MPE	12.199	0.243	11.608	11.216	6.509	0.531
	MAPE	12.400	0.652	11.795	11.373	6.621	0.577

APPENDIX

TABLE A Independent NB-OLFS Model Results for Florida

Variable Names	NB Model Component		OLFS Model Component	
	Estimate	t-stat	Estimate	t-stat
Constant	1.471	7.208	---	---
Threshold Parameters	---	---	---	---
Threshold between no injury and minor injury	---	---	-0.121	-0.625
Threshold between minor and incapacitating injury	---	---	2.056	10.560
Threshold between incapacitating and fatal injury	---	---	3.990	19.544
Sociodemographic Characteristics				
HH density	0.461	15.076	-0.053	-4.524
Dependence	-0.316	-7.322	---	---
Proportion of female population	-1.240	-3.080	---	---
Proportion of Caucasian population	-0.937	-11.614	-1.020	-6.061
Proportion of Asian population	-2.243	-3.793	---	---
Proportion of Hispanic population	1.033	12.921	-0.536	-10.845
Proportion of African-American population	---	---	-1.063	-6.118
Socioeconomic Characteristics				
Automobile commuters	---	---	0.049	4.375
Transit commuters	---	---	-0.007	-2.208
Walk commuters	-0.041	-5.726	---	---
Bike commuters	-0.025	-3.380	---	---
Number of workers that work from home	---	---	-0.011	-2.348
Employment density	0.312	15.817	-0.027	-3.093
Population below poverty level	---	---	0.316	3.385
Built Environment				

Proportion of Urban area	1.394	12.577	-0.231	-5.117
Law enforcement offices	-0.131	-4.056	-0.087	-4.757
Restaurants	0.156	7.383	-0.098	-6.215
Park and recreational centers	-0.447	-5.364	-0.205	-3.744
Transportation hubs	-0.402	-4.304	---	---
Shopping centers	---	---	-0.023	-2.565
Land use mix	-0.935	-10.597	---	---
Proportion of retail and office land use	1.616	4.791	---	---
Proportion of industrial land use	---	---	0.602	2.567
Transport Infrastructure				
Local roads	---	---	-0.018	-5.554
Major roads	-0.043	-5.415	---	---
Traffic signal density	0.348	11.533	---	---
Intersection density	0.341	9.651	---	---
Traffic Characteristics				
Vehicle miles travelled	---	---	0.050	3.760
Proportion of heavy vehicle miles travelled	---	---	0.435	2.347
Average speed	---	---	0.065	3.852
Traffic intensity	---	---	-0.060	-4.323
Dispersion parameter	1.760	50.424	---	---

TABLE B Joint NB-OLFS without Correlation Parameterization Model Results for Florida

Variable Names	NB Model Component		OLFS Model Component	
	Estimate	**t-stat**	**Estimate**	**t-stat**
Constant	1.420	7.285	---	---
Threshold Parameters	---	---	---	---
Threshold between no injury and minor injury	---	---	-0.119	-0.601
Threshold between minor and incapacitating injury	---	---	2.061	10.312
Threshold between incapacitating and fatal injury	---	---	3.996	19.107
Sociodemographic Characteristics				
HH density	0.472	17.427	-0.049	-4.087
Dependence	-0.309	-7.403	---	---
Proportion of female population	-1.296	-3.540	---	---
Proportion of Caucasian population	-0.942	-12.202	-1.031	-5.955
Proportion of Asian population	-2.250	-3.947	---	---
Proportion of Hispanic population	1.037	13.547	-0.534	-10.750
Proportion of African-American population	---	---	-1.076	-6.027
Socioeconomic Characteristics				
Automobile commuters	---	---	0.039	3.227
Transit commuters	---	---	-0.007	-2.141
Walk commuters	-0.041	-6.088	---	---
Bike commuters	-0.024	-3.506	---	---
Number of workers that work from home	---	---	-0.011	-2.256
Employment density	0.316	18.259	-0.032	-3.477
Population below poverty level	---	---	0.302	3.214
Built Environment				
Proportion of Urban area	1.372	13.465	-0.243	-5.297
Law enforcement offices	-0.127	-4.049	-0.085	-4.596

Restaurants	0.159	7.763	-0.096	-6.029
Park and recreational centers	-0.448	-5.246	-0.195	-3.550
Transportation hubs	-0.400	-4.382	---	---
Shopping centers	---	---	-0.022	-2.441
Land use mix	-0.919	-10.605	---	---
Proportion of retail and office land use	1.636	4.897	---	---
Proportion of industrial land use	---	---	0.602	2.514
Transport Infrastructure				
Local roads	---	---	-0.018	-5.591
Major roads	-0.040	-5.556	---	---
Traffic signal density	0.337	12.670	---	---
Intersection density	0.340	10.302	---	---
Traffic Characteristics				
Vehicle miles travelled	---	---	0.054	4.023
Proportion of heavy vehicle miles travelled	---	---	0.435	2.313
Average speed	---	---	0.065	3.818
Traffic intensity	---	---	-0.060	-4.294
Dispersion parameter	1.732	53.880	---	---

Correlation Parameters		
Variables	**Estimate**	**t-stat**
Constant	0.177	4.257