Joint Modeling of Pedestrian and Bicycle Crashes: A Copula Based Approach

Tammam Nashad
Department of Civil, Environmental & Construction Engineering
University of Central Florida
Email: tammam24@knights.ucf.edu

Shamsunnahar Yasmin
Department of Civil Engineering & Applied Mechanics
McGill University
Tel: 514-398-6823, Fax: 514-398-7361
Email: shamsunnahar.yasmin@mail.mcgill.ca

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

Jaeyoung Lee
Department of Civil, Environmental & Construction Engineering
University of Central Florida
Email: jaeyoung.lee@ucf.edu

Mohamed A. Abdel-Aty
Professor
Department of Civil, Environmental & Construction Engineering
University of Central Florida
Tel: (407) 823-4535, Fax: 407-823-3315
Email: m.aty@ucf.edu

November, 2015

Submitted for **Presentation and Publication** to Committee on Safety Data, Analysis and Evaluation (ANB20)

95th Transportation Research Board Annual Meeting

Word count: 6452 words + 3 tables + 2 figures =7702 equivalent words

*Corresponding author

ABSTRACT

The study contributes to safety literature on active mode transportation safety by employing a copula based model for count frequency analysis at a macro-level. Most studies in the transportation safety area identify a single count variable (such as vehicular, pedestrian or bicycle crash counts) for a spatial unit and study the impact of exogenous variables. While the traditional count models perform adequately in the presence of a single count variable, it is necessary to modify these approaches to examine multiple dependent variables for each study unit. To that extent, the current research effort contributes to literature by developing a multivariate model by adopting a copula based bivariate negative binomial model for pedestrian and bicyclist crash frequency analysis. The proposed approach also accommodates for potential heterogeneity (across zones) in the dependency structure. The formulated models are estimated using pedestrian and bicycle crash count data at the Statewide Traffic Analysis Zone (STAZ) level for the state of Florida for the years 2010 through 2012. The STAZ level variables considered in our analysis include exposure measures, socio-economic characteristics, road network characteristics and land use attributes. A policy analysis is also conducted along with a representation of hotspot identification to illustrate the applicability of the proposed model for planning purposes. The development of such spatial profiles will allow planners to identify high risk zones for screening and treatment purposes.

INTRODUCTION

Urban regions in North America are encouraging the adoption of active modes of transportation by proactively developing infrastructure for these modes. According to data from the 2009 National Household Travel Survey (NHTS), about 37.6% of the trips by private vehicles in the United States (US) are less than 2 miles long. Even if a small proportion of the shorter private vehicle trips (around dense urban cores) are substituted with public transportation and active transportation trips, it offers substantial benefits to individuals, cities and the environment. However, a strong impediment to the increasing adoption of active modes of transportation is the risk associated with these modes. In fact, in the US between 2004 and 2013, bicycle and pedestrian fatalities as a percentage of total traffic crash related fatalities have increased from 1.7% to 2.3% and 11% to 14%, respectively (*1*). For increasing the adoption of active transportation, there is a need to reduce the risk to pedestrians and bicyclists on roadways. The safety risk posed to active transportation users in Florida is exacerbated compared to active transportation users in the US. While the national average for pedestrian (bicyclist) fatalities per 100,000 population is 1.50 (2.35), the corresponding number for the state of Florida is 2.56 (6.80), which clearly present a clear picture of the challenge faced in Florida. An important tool to determine the critical factors affecting occurrence of pedestrian and bicycle crashes; and identifying vulnerable locations is the application of planning level crash prediction models.

Traffic crashes aggregated at a certain spatial scale are non-negative integer valued random events. Naturally, these integer counts are examined employing count regression approaches that quantify the influence of exogenous factors on crash counts. Most studies in the transportation safety area identify a single count variable (such as vehicular, pedestrian or bicycle crash counts) for a spatial unit and study the impact of exogenous variables. In this context, the crash prediction model structures considered include Poisson (*2*-*3*), Poisson-Lognormal, Poisson-Gamma regression (also known as negative binomial (NB)), Poisson-Weibull, and Generalized Waring models (*4*-*10*). Among these model structures, the NB model, which offers a closed form expression while relaxing the equal mean variance equality constraint of Poisson regression, serves as the workhorse for crash count modeling.

MULTIPLE DEPENDENT VARIABLES

While the above models perform adequately in the presence of a single count variable, it is necessary to modify these approaches to examine multiple dependent variables for each study unit. To elaborate, for a study unit, if multiple dependent variables are available it is plausible to imagine that common observed and unobserved factors that affect one dependent variable might also affect the other dependent variables. Accommodating for the impact of observed factors is relatively straightforward within count regression models by estimating distinct count models for every dependent variable. 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. In traditional discrete choice models, there are three ways that such joint processes are examined can be accommodated. The first approach considers the dependent variables being investigated as marginal distributions within a bivariate (or multivariate) distribution by developing a joint error distribution. The distribution parameters estimated will allow us to evaluate the correlation between the dependent variables. If permissible, the approach usually results in closed form parametric formulations. These formulations thus allow for analytical computation of log-likelihood and offer more stable inference conclusions. Examples of such approaches include bivariate normal or logistic distributions, bivariate NB distributions or the flexible bivariate copula based approaches (for example see (*11-13*)). Of course, the flexibility of the approach is restricted by the potential parametric alternatives available. In the transportation safety area, to our best knowledge, no count models have been developed employing this approach.

The second approach to addressing multiple dependent variables involves the development of multivariate function as described in the first approach. However, as the estimation of the multivariate approach is computationally intractable, an approximation approach to evaluating the multivariate function is considered. The approach – referred to as the composite marginal likelihood approach (CML) – has received considerable attention in transportation literature in recent years (*14-16*). In terms of crash count modeling, the approach has been employed by Narayanamoorthy et al. (*17*) for bicycle and pedestrian crash counts by severity type.

The third approach to accommodate for the dependency between the dependent variables allows for stitching by considering unobserved error components that jointly affect the dependent variables. The approach, usually, partitions the error components of the dependent variables to accommodate for a common term and an independent term across dependent variables. The common error term across the dependent variables allows for the possible unobserved effects. Of course, the common term is considered with a distribution that has a zero mean. Thus, any computation of probability requires an integral across the error term distribution. The probability computation is dependent on the distributional assumption and no longer has a closed form expression. Thus, the estimation procedure requires the adoption of maximum simulated likelihood (MSL) approach or Markov Chain Monte Carlo (MCMC) approach in the Bayesian realm. MSL and MCMC methods provide substantial flexibility in accommodating for unobserved heterogeneity. However, in MSL and MCMC methods, the probability computation is sensitive to number of draws as well as random number generation procedures. Further, these approaches are more prone to efficiency loss due to inaccuracy in retrieving the variance covariance parameters that is critical for inference (see (*18*) for more detailed discussion on issue with MSL approaches). A majority of the count modeling approaches employed in the safety area have adopted the third approach. Specifically, the model structures employed in literature include multivariate-Poisson model (*19*), Poisson-lognormal models (*7*, *20-22*) and simultaneous equation models (*23-24*).

CURRENT STUDY IN CONTEXT

From the above literature review it is evident that transportation safety literature of count modeling realm has predominantly focused on the third approach to examining multivariate count variables. The current research effort contributes to literature on the first approach – developing a multivariate model by adopting a copula based bivariate NB model for pedestrian and bicyclist crash frequency analysis. The proposed approach has three major advantages relative to existing methods. <u>First</u>, in the earlier research attempts (from second and third groups described above), a particular distributional assumption on the nature of the correlation across the multiple dependent variables is considered. However, it is possible that the distributional assumption might influence the results. In a copula based approach, we can empirically compare the different dependency structures thus enhancing the flexibility of the multivariate approach. Thus the copula approach subsumes any bivariate modeling approach. <u>Second</u>, the copula based approach for count modeling results in an analytical formulation as opposed to an approximation (as in CML methods) or simulation (in MSL or MCMC approaches). Thus, the parameter estimates are likely to be more accurate. Finally, it is possible that several exogenous factors might affect the dependency profile between the multiple variables. We accommodate for these impacts by parameterizing the dependency profile to allow for such potential heterogeneity (across zones).

A simpler version of the approach proposed here has been employed in econometrics (*25*). In their study, the copula dependency is considered to be the same across the entire dataset. To the best of the authors' knowledge, this is the first attempt to employ such copula based bivariate count models for examining crash count events. To be sure, copula models for ordered and unordered discrete outcome variables have been adopted in safety literature (see (*26*-*29*)). In this paper, we employ the copula based models for count events analysis. Empirically, the study examines the influence of several exogenous variables (exposure measures, socioeconomic characteristics, road network characteristics and land use attributes) on pedestrian and bicycle crash count events at the Statewide Traffic Analysis Zone (STAZ) level for the state of Florida.

MODEL FRAMEWORK

The focus of our study is to jointly model pedestrian crash frequency and bicycle crash frequency at zonal level by employing a copula based bivariate NB modeling framework. The econometric framework for the joint model is presented in this section.

Let i be the index for STAZ ($i = 1,2,3, \ldots, N$) and y_{qi} be the index for crashes occurring over a period of time in a STAZ i; where q takes the value of 1 for pedestrian crashes and 2 for bicycle crashes. The NB probability expression for random variable y_{qi} can be written as (*25*):

$$P_{qi}(y_{qi}|\mu_{qi}, \alpha_q) = \frac{\Gamma(y_{qi}+\alpha_q^{-1})}{\Gamma(y_{qi}+1)\Gamma(\alpha_q^{-1})}\left(\frac{1}{1+\alpha_q\mu_{qi}}\right)^{\frac{1}{\alpha_q}}\left(1-\frac{1}{1+\alpha_q\mu_{qi}}\right)^{y_{qi}} \quad (1)$$

where, $\Gamma(\cdot)$ is the Gamma function, α_q is the NB dispersion parameter specific to road user group q and μ_{qi} is the expected number of crashes occurring in STAZ i over a given period of time for vulnerable road user group q. We can express μ_{qi} as a function of explanatory variable ($\boldsymbol{x}_{qi}$) by using a log-link function as: $\mu_{qis} = E(y_{qi}|\boldsymbol{x}_{qi}) = exp(\boldsymbol{\beta}_q\boldsymbol{x}_{qi})$, where $\boldsymbol{\beta}_q$ is a vector of parameters to be estimated specific to road user group q.

The correlation or joint behaviour of random variables y_{1i} and y_{2i} are explored in the current study by using a copula based approach. A copula is a mathematical device that identifies dependency among random variables with pre-specified marginal distribution ((*30*-*31*) provide a detailed description of the copula approach). In constructing the copula dependency, let us assume that $\Lambda_1(y_{1i})$ and $\Lambda_2(y_{2i})$ are the marginal distribution functions of the random variables y_{1i} and y_{2i}, respectively; and $\Lambda_{12}(y_{1i}, y_{2i})$ is the joint distribution for the bivariate case with corresponding marginal distribution. Subsequently, the bivariate distribution $\Lambda_{12}(y_{1i}, y_{2i})$ can be generated as a joint cumulative probability distribution of uniform [0, 1] marginal variables U_1 and U_2 as below (*11*):

$$\Lambda_{12}(y_{1i}, y_{2i}) = Pr(U_1 \le y_{1i}, U_2 \le y_{2i}) \quad (2)$$

$$= Pr[\Lambda_1^{-1}(U_1) \leq y_{1i},\ \Lambda_2^{-1}(U_2) \leq y_{2i}\ ]$$

$$= Pr[U_1 < \Lambda_1(y_{1i}),\ U_2 < \Lambda_2(y_{2i})\ ]$$

The joint distribution (of uniform marginal variable) in equation 2 can be generated by a function $C_{\theta i}(.,.)$ (*32*), such that:

$$\Lambda_{12}(y_{1i}, y_{2i}) = C_{\theta i}(U_1 = \Lambda_1(y_{1i}), U_2 = \Lambda_2(y_{2i})) \tag{3}$$

where, $C_{\theta i}(.,.)$ is a copula function and θ_i is the dependence parameter defining the link between y_{1i} and y_{2i}. In the case of two continuous random variables, the bivariate density (or joint density) can be derived from partial derivatives for the continuous case. However, in our study, y_{1i} and y_{2i} are nonnegative integer valued events. For such count data, following Cameron et al. (*25*), the probability mass function ($\zeta_{\theta i}$) is presented (instead of continuous derivatives) by using finite differences of the copula representation as follows:

$$\begin{aligned} \zeta_{\theta i}\big(\Lambda_1(y_{1i}), \Lambda_2(y_{2i})\big) = C_{\theta i}&(\Lambda_1(y_{1i}), \Lambda_2(y_{2i});\ \theta_i) \\ -C_{\theta i}&(\Lambda_1(y_{1i}-1), \Lambda_2(y_{2i});\ \theta_i) \\ -C_{\theta i}&(\Lambda_1(y_{1i}), \Lambda_2(y_{2i}-1);\ \theta_i) \\ +C_{\theta i}&(\Lambda_1(y_{1i}-1), \Lambda_2(y_{2i}-1);\ \theta_i) \end{aligned} \tag{4}$$

Given the above setup, we specify $\Lambda_1(y_{1i})$ and $\Lambda_2(y_{2i})$ as the cumulative distribution function (cdf) of the NB distribution. The cdf of NB probability expression (as presented in equation 1) for y_{qi} can be written as:

$$\Lambda_q\big(y_{qi}|\mu_{qi}, \alpha_q\big) = \sum_{k=0}^{y_{qi}} P_{qi}\big(y_{qi}|\mu_{qi}, \alpha_q\big) \tag{5}$$

Thus, the log-likelihood function (LL) with the joint probability expression in equation 4 can be written as:

$$LL = \sum_{i=1}^{N} \zeta_{\theta i}\big(\Lambda_1(y_{1i}), \Lambda_2(y_{2i})\big) \tag{6}$$

In our empirical analysis we select six different copula structures: 1) Gaussian, 2) Farlie-Gumbel-Morgenstern (FGM), 3) Clayton, 4) Gumbel, 5) Frank and 6) Joe (a detailed discussion of these copulas is available in (*30*)). Among these copulas; Gaussian, FGM and Frank copulas represent symmetric dependency structures that ensure higher dependency for unobserved variables around the mean of the distribution. Clayton copula allows for stronger dependency between the unobserved variables for the lower tails of the distribution. Gumbel and Joe distributions offer the mirror image to Clayton copula by allowing for stronger dependency toward the positive tails of the distribution. Between Joe and Gumbel copula, Joe copula allows for a stronger positive tail dependency (for more details see Figure 1 in (*30*)).

It is important to note here that, the level of dependence between the random variables can vary across STAZs. Therefore, in the current study, the dependence parameter θ_i is parameterized as a function of observed attributes as follows:

$$\theta_i = fn(\boldsymbol{\gamma}_q \boldsymbol{s}_{qi}) \tag{7}$$

where, $\boldsymbol{s}_{qi}$ is a column vector of exogenous variable, $\boldsymbol{\gamma}_q$ is a row vector of unknown parameters (including a constant) specific to road user group q and fn represents the functional form of parameterization. Based on the dependency parameter permissible ranges, alternate parameterization forms for the six copulas are considered in our analysis. For Normal, FGM and Frank Copulas we use $\theta_i = \boldsymbol{\gamma}_q \boldsymbol{s}_{qi}$, for the Clayton copula we employ $\theta_i = exp(\boldsymbol{\gamma}_q \boldsymbol{s}_{qi})$, and for Joe and Gumbel copulas we employ $\theta_i = 1 + exp(\boldsymbol{\gamma}_q \boldsymbol{s}_{qi})$. The parameters to be estimated in the model of Equation 6 are: $\boldsymbol{\beta}_q$, α_q and $\boldsymbol{\gamma}_q$. The parameters are estimated using maximum likelihood approaches. The model estimation is achieved through the log-likelihood functions programmed in Gauss (*33*).

DATA DESCRIPTION

This study is focused on pedestrian and bicycle crashes at the STAZ level. There are 8,518 STAZs in the State of Florida (Figure 1). Data for the empirical study is obtained from Florida for the year 2010 through 2012. The pedestrian and bicycle crash records are collected and compiled from Florida Department of Transportation CAR (Crash Analysis Reporting) 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). Crash data records from short and long form databases are compiled in order to have more complete information on road crashes and hence is used for the purpose of analysis in the current study context.

In addition to the crash database, the explanatory attributes considered in the empirical study are also aggregated at the STAZ level accordingly. For the empirical analysis, the selected explanatory variables can be grouped into four broad categories: exposure measures, socioeconomic characteristics, road network characteristics and land use attributes. The exposure measures, socioeconomic characteristics, and land use attributes are obtained from the US Census Bureau and FDOT District Offices/MPOs (or FDOT Central Office). Moreover, the road network characteristics and traffic related attributes are collected from FDOT Transportation Statistics Office (TRANSTAT). STAZ data are collected from Florida Department of Transportation District Offices/MPOs (or Florida Department of Transportation Central Office), the U.S. Census Bureau, and Florida Geographic Data Library (FGDL). Table 1 offers a summary of the sample characteristics of the count and exogenous variables and the definition of variables considered for final model estimation along with the zonal minimum, maximum, average and standard deviation values for Florida. From Table 1, we can see that for the three years, Florida has a record of 16,240 pedestrian crashes with an average of 1.90 crashes (ranging from 0 to 39 crashes) per STAZ. On the other hand, the state has an average of 1.79 crashes (ranging from 0 to 88) per TAZ with a total record of 15,307 bicycle crashes for the three years period.

FIGURE 1 State Traffic Analysis Zones (STAZs) for the state of Florida

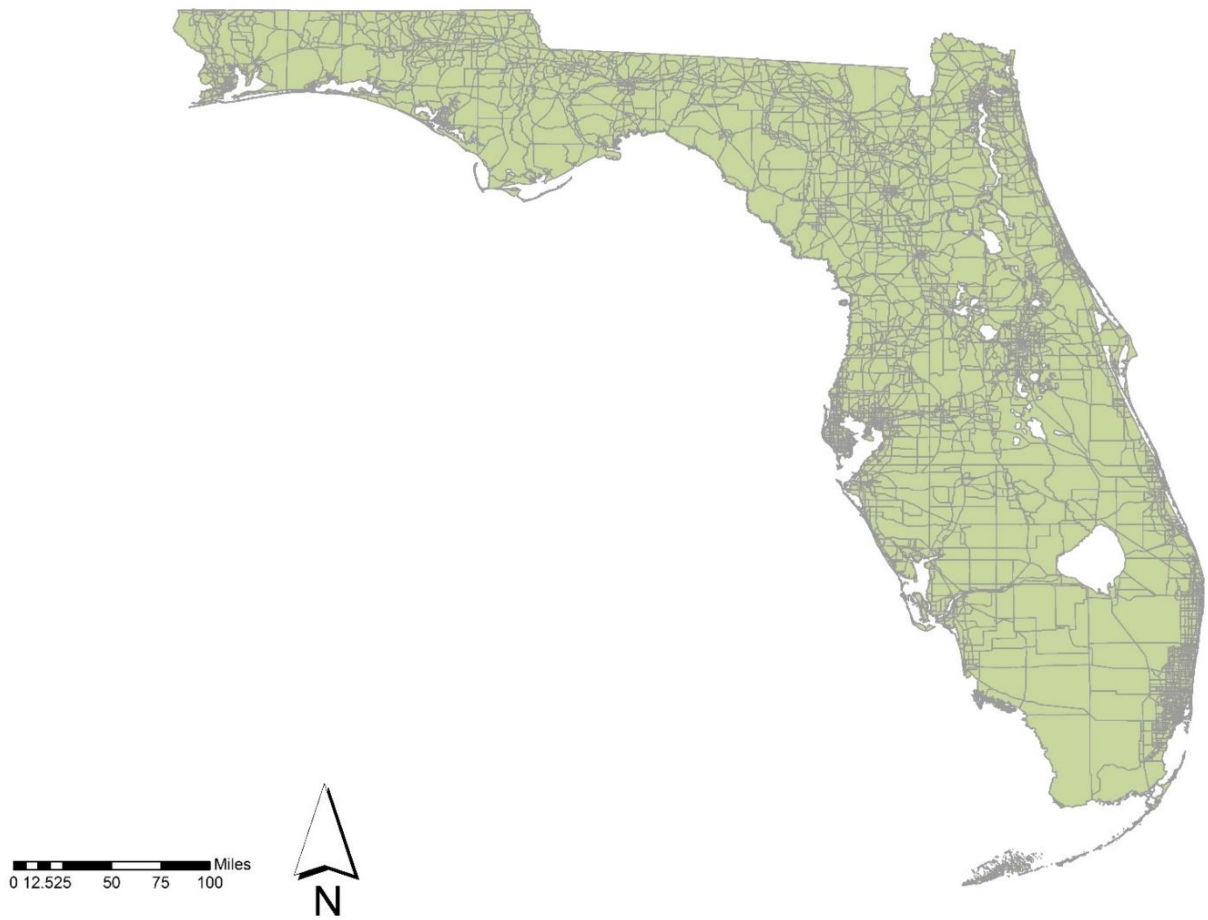


TABLE 1 Sample Statistics for the State of Florida

Variable Names	Variables Descriptions	Zonal				Percentiles	
		Minimum	Maximum	Average	Std. Deviation	20th	80th
Dependent variable							
Pedestrian crashes per STAZ	Total number of pedestrian crashes per STAZ	0.000	39.000	1.907	3.315	0.000	3.000
Bicycle crashes per STAZ	Total number of bicycle crashes per STAZ	0.000	88.000	1.797	3.309	0.000	3.000
Exposure measures							
VMT	Natural Log of vehicle miles travel (VMT) in STAZ	0.000	13.437	9.039	2.659	7.978	10.870
Proportion of heavy vehicles	Total heavy vehicle VMT in STAZ /Total vehicles VMT in STAZ	0.000	0.519	0.067	0.052	0.031	0.095
Total population	Natural log of total population in STAZ	0.000	10.571	6.437	2.144	4.990	8.233
Population density	Natural log of population density (per sqmi)	0.000	11.052	6.481	2.257	4.542	8.267
Proportion of families with no vehicle	Total number of families with no vehicle in STAZ/Total number of families in STAZ	0.000	1.000	0.095	0.123	0.020	0.133
Socio-economic characteristics							
Public transit commuters	No of commuters using public transportation	0.000	934.000	18.812	54.273	0.000	18.000
Bicycle commuters	No of commuters using bicycle	0.000	775.000	5.844	19.263	0.000	6.000
Walk commuters	No of commuters by walking	0.000	1288.000	14.352	34.681	0.000	20.000
Total employment	Natural log of total employment in STAZ	0.000	10.371	5.857	2.017	4.382	7.523
Proportion of service employment	Proportion of service employment	0.000	1.000	0.525	0.257	0.294	0.760
Proportion of industrial employment	Proportion of industrial employment	0.000	1.000	0.176	0.232	0.000	0.333
Proportion of commercial employment	Proportion of commercial employment	0.000	1.000	0.299	0.235	0.072	0.498
School enrollment density	Natural Log of total school enrollment per square miles in STAZ	0.000	12.450	2.715	3.143	0.000	6.278
Road network characteristics							
Proportion of urban area	Total urban area in STAZ/Total area in STAZ	0.000	1.000	0.722	0.430	0.007	1.000

Proportion of local roads	Total length of local roads in STAZ/Total length of all roads in STAZ	0.000	1.000	0.572	0.329	0.177	0.858
Proportion of collector roads	Proportion of collectors	0.000	1.000	0.191	0.246	0.000	0.323
Proportion of arterial roads	Total length of arterial roads in STAZ/Total length of all roads in STAZ	0.000	1.000	0.221	0.275	0.006	0.369
Traffic signal density	Natural log of total number of traffic signals per miles of road in STAZ	0.000	8.756	0.227	0.578	0.000	0.269
Bike lane length	Bike lane length	0.000	28.637	0.303	1.096	0.000	0.030
Sidewalk length	Sidewalk length	0.000	25.683	0.993	1.750	0.000	1.735
Land use attributes							
Density of hotel/ motel/timeshare room	Natural log of total number of hotel, motel, timeshare room per square mile in STAZ	0.000	10.392	1.549	2.365	0.000	3.924
Distance to nearest urban area	Distance of the STAZ to the nearest urban area in miles	0.000	44.101	2.140	5.441	0.000	1.606

EMPIRICAL ANALYSIS

Model Specification and Overall Measures of Fit

The empirical analysis involves the estimation of models by using six different copula structures: 1) Gaussian, 2) FGM, 3) Clayton, 4) Gumbel, 5) Frank and 6) Joe. The empirical analysis involved a series of model estimations. *First*, an independent copula model (separate NB models for pedestrian and bicycle crash counts) were estimated to establish a benchmark for comparison. *Second*, six different models were estimated by considering the dependency parameter in the copula model to be the same across all STAZs. *Third*, different copula models were also estimated by considering the parameterization for copula dependency profile. Finally, to determine the most suitable copula model (including the independent copula model), a comparison exercise was undertaken. The alternative copula models estimated are non-nested and hence, cannot be tested using traditional log-likelihood ratio test. We employ the Bayesian Information Criterion (BIC) to determine the best model among all copula models (*26*, *31*-*32*, *34*). The BIC for a given empirical model is equal to:

$$BIC = -2LL + K\, ln(Q) \tag{8}$$

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 copula model. The BIC value for independent copula model was 48747.45. The following copula models (BIC) without parameterization offered improved data fit: Clayton (48343.15), FGM (48388.16) and Frank (48340.05). Gaussian, Gumbel and Joe copulas collapsed to independent copula model. For copula dependency profile parameterization, the variables effects were significant only for Clayton copula. Overall, Clayton copula with dependency profile parameterization (48271.85) outperformed all other copula models as well the independent model. The copula model BIC comparisons confirm the importance of accommodating dependence between pedestrian and bicycle crash count events in the macro-level analysis.

Estimation Results

In presenting the effects of exogenous variables in the joint model specification, we will restrict ourselves to the discussion of the Clayton Copula specification. Table 2 presents the estimation results of the joint model. For the ease of presentation, the pedestrian crash count component (3rd and 4th columns of Table 2) and bicycle crash count component (5rd and 6th columns of Table 2) results are discussed together in the following section by variable groups. The copula parameters are presented in the last row panel of Table 2.

Exposure measures

In terms of exposure measures, the estimates indicate that both pedestrian and bicycle crashes are positively associated with higher vehicle-miles traveled (VMT) at the zonal level. The result related to VMT represents the higher crash risk faced by non-motorized (pedestrian and bicyclist) road user groups with increasing VMT (*35*). Further, the results in Table 2 indicate reduced crash propensity for both pedestrian and bicyclists with higher proportion of heavy vehicle VMT at the zonal level. With respect to total population, the joint model estimation results reveal that both pedestrian and bicycle crashes are positively associated with higher zonal population (*36*-*38*).

As expected, both pedestrian and bicycle crash risk are found to be higher for the STAZs with higher proportion of households without access to private vehicles (see *39*-*40*), but the magnitude of the impact is more pronounced for pedestrian crashes relative to bicycle crashes. The results can be explained by the fact that members of the households with access to no private vehicles would use alternate mode of transportation for daily activities resulting in higher pedestrian and bicycling exposure in these STAZs. The variable is also surrogate indicator for low-income level of zone, where people are less likely to receive safety education and hence are exposed to higher potential crash risk (*41*).

Socioeconomic characteristics

The results for the number of commuters based on different commute modes are also found to significantly influence pedestrian and bicycle crash risk in the current study context. An increase in the number of transit commuters increases the likelihood of pedestrian and bicycle crashes at the STAZ level. The result in pedestrian crash model intuitively suggests higher demand and supply of public transit in zones with higher number of transit commuters which are determinants of pedestrian activities (*42*). The variable indicating transit commuters in bicycle crash model is possibly representing greater bicycle exposure from higher cycle-transit integrated mode share (popularly known as "bike-and-ride") for access and egress at transit stations (*43*). In terms of walk and bicycle commuters, the results reveal that STAZs with higher number of walk and bike commuters increase the likelihoods of both pedestrian and bicycle crashes. These variables can be considered as proxy measures for pedestrian and bicycle exposure in the zones. It is interesting to note that both non-motorized commute variables have larger impact in bicycle crash count events relative to pedestrian crash count events. As found in previous studies (*39, 41*), our study also found that more employment within a TAZ leads to higher probability of bicycle crashes. However, increasing proportion of industrial employment has negative association with pedestrian and bicycle crashes at the STAZ level. Also, an increase in school enrollment density in a STAZ increases the likelihoods of crash risk in count model components for both non-motorized road user group.

Road network characteristics

Proportion of urban area, a proxy for non-motorized activity, reflects that an increase in the proportion of urban area in a zone increases the likelihood of both pedestrian and bicycle crash risk. The results associated with functional class of roadways show that pedestrian and bicycle crash risk are positively correlated with higher proportion of arterial and local roads. Consistent with several previous studies (*44*-*45*), our study results also show that higher density of signalized intersections are positively associated with more pedestrian- and bicycle-motor vehicle crashes. With respect to sidewalk length, the model estimation results indicate higher likelihood of pedestrian and bicycle crashes with increasing length of sidewalk in a zone.

TABLE 2 Pedestrian-Bicycle Joint Model Estimation Results – Clayton Copula

Variable Names	Pedestrian		Bicycle	
	Estimate	t-stat	Estimate	t-stat
Constants	-4.238	-38.738	-4.272	-41.469
Exposure measures				
VMT	0.118	20.646	0.128	20.775
Proportion of heavy vehicles	-0.902	-2.444	-3.145	-8.786
Total population	0.137	17.447	0.138	15.339
Proportion of families with no vehicle	1.323	12.040	0.244	1.976
Socio-economic characteristics				
Bicycle commuters	0.036	3.841	0.144	16.754
Public transit commuters	0.171	21.750	0.097	11.480
Walk commuters	0.070	7.286	0.081	8.129
Total employment	0.172	16.812	0.136	14.087
Proportion of industrial employment	-0.242	-3.632	-0.191	-2.794
School enrollment density	0.012	3.022	0.011	2.638
Road network characteristics				
Proportion of urban area	0.272	5.146	0.658	11.170
Proportion of local roads	0.564	8.752	0.565	8.157
Proportion of arterial roads	0.306	3.949	0.422	5.040
Traffic signal density	0.289	12.716	0.184	7.281
Sidewalk length	0.272	12.963	0.309	14.754
Land use attributes				
Density of hotel/motel/timeshare room	0.029	5.943	0.018	3.429
Distance to nearest urban area	-0.039	-7.031	-0.084	-9.363
Copula Parameters				
Variable Names	**Estimate**		**t-stat**	
Constant	-0.973		--	
Public transit commuters	0.141		4.373	
School enrollment density	0.049		2.728	

Land use attributes

The result associated with hotel/motel/timeshare room density in STAZ reflects that an increase in hotel/motel/timeshare room density increases the likelihood of both pedestrian and bicycle crash risk, presumably indicating higher level of non-motorized road user activity in the proximity of these facilities in a zone (*46-47*). Moreover, tourists/visitors might be unfamiliar/less familiar with local driver behavior and road regulations (*48*), which might further exacerbate crash risk for these non-motorized road user groups. The possibilities of pedestrian and bicycle crash risk increase with increasing distance to the nearest urban area from the STAZ. STAZs close to urban area are associated with shorter, more walkable and/or cyclable travel distances which in turn increase the exposure of non-motorized road user groups resulting in increased likelihood of crash risks.

Dependence Effects

As indicated earlier, the estimated Clayton copula based bivariate NB model provides the best fit in incorporating the correlation between the pedestrian and bicycle crash count events. An examination of the copula parameters presented in the last row panel of Table 2 highlights the presence of common unobserved factors affecting pedestrian and bicycle crash frequency. The various exogenous variables that contribute to the dependency include school enrollment density and public transit commuters. This provides support to our hypothesis that the dependency structures are not constant across all STAZs. For the Clayton copula, the dependency is entirely positive and the coefficient sign and magnitude reflects whether a variable increases or reduces the dependency and by how much. The proposed framework by allowing for such parameterizations allows us to improve data fit.

POLICY ANALYSIS

Elasticity Effects and Implications

The parameter effects of exogenous variables in Table 2 do not provide the magnitude of the effects on zonal level crash counts. For this purpose, we compute aggregate level “elasticity effects” of exogenous variables for both pedestrian and bicycle crash events. We compute the percentage change in the expected total zonal crash counts due to the change in exogenous variable for pedestrian and bicycle separately to identify the policy measures based on most critical contributory factors. The computed elasticities are presented in Table 3 (see (*49*) for a discussion on the methodology for computing elasticities). The results in Table 3 represent the percentage change in the number of crashes for 100% change in the independent variable, other characteristics being equal. For example, the elasticity estimate for VMT variable indicates that a 100% increase in VMT will result in a 25.1% increase in pedestrian crashes and a 26.3% increase in bicyclist crashes.

The following observations can be made based on the elasticity effects presented in Table 3. First, the results in Table 3 indicate that there are differences in the elasticity effects across the expected number of pedestrian and bicycle crash counts. Second, the most significant variable in terms of increase in the expected number of both pedestrian and bicycle crash counts include: VMT, total population and total employment. Third, pedestrian crashes have higher elasticities relative to bicycle crashes for total population, total employment, public transit commuters, proportion of families with no vehicle, traffic signal density and density of hotel/motel/timeshare room. Finally, based on the elasticity estimates it is evident that the influence of exposure and socio-economic characteristics is substantially larger than the influence of roadway and land-use characteristics.

These results have important implications in improving the safety situation for non-motorized road users and promoting active mode of transportation. For instance, results indicating auto-oriented (VMT) and public transit-oriented (public transit commuters) neighborhoods have important implications in terms of engineering measures. Traffic calming measures should be provided in these zones to reduce road crashes involving pedestrians and bicyclists. Engineering infrastructure (such as overpasses, shaded walkways for pedestrian traffic and bike box at intersections, bike paths for bicycle traffic) that separate non-motorized traffic flow from motorized traffic flow in the road network system should be installed and regulated in the zones with more population and more employment. Public awareness efforts and traffic education for safe walking and cycling are needed for both non-motorists and motorists of zones with more transit, bike and walk commuters. Moreover, education campaigns in the communities with less access to private vehicles are needed to improve non-motorists' safety situation. Further, targeted enforcement strategies should be regulated in the zones with more local roads and sidewalks to make the neighborhoods more walkable and bikeable. Overall, the elasticity analysis conducted provides an illustration on how the proposed model can be applied to determine the critical factors contributing to increase in pedestrian and bicycle crash counts.

TABLE 3 Elasticity Effects

Variable Names	Pedestrian	Bicycle
Exposure measures		
VMT	25.076	26.318
Proportion of heavy vehicles	-0.938	-2.887
Total population	22.014	21.407
Proportion of families with no vehicle	2.973	0.442
Socioeconomic characteristics		
Bicycle commuters	1.147	5.097
Public transit commuters	9.831	5.018
Walk commuters	3.760	4.257
Total employment	25.730	19.239
Proportion of industrial employment	-0.582	-0.421
School enrollment density	1.034	0.916
Road network characteristics		
Proportion of urban area	0.208	0.505
Proportion of local roads	7.198	7.016
Proportion of arterial roads	0.944	1.214
Traffic signal density	1.809	0.922
Sidewalk length	4.840	5.538
Land use attributes		
Density of hotel/motel/timeshare room	1.207	0.691
Distance to nearest urban area	-0.224	-0.210

Spatial Distribution of Hotspots

The model findings have also important implications in terms of identifying hotspot at the zonal level for non-motorized road user safety planning. To identify the hotspots, the Highway Safety Manual approach that computes the Excess Predicted Average Crash Frequency defined as observed frequency minus predicted crash frequency. Based on the measure the 10% of the zones are labelled as hot zones and others are labelled Normal.

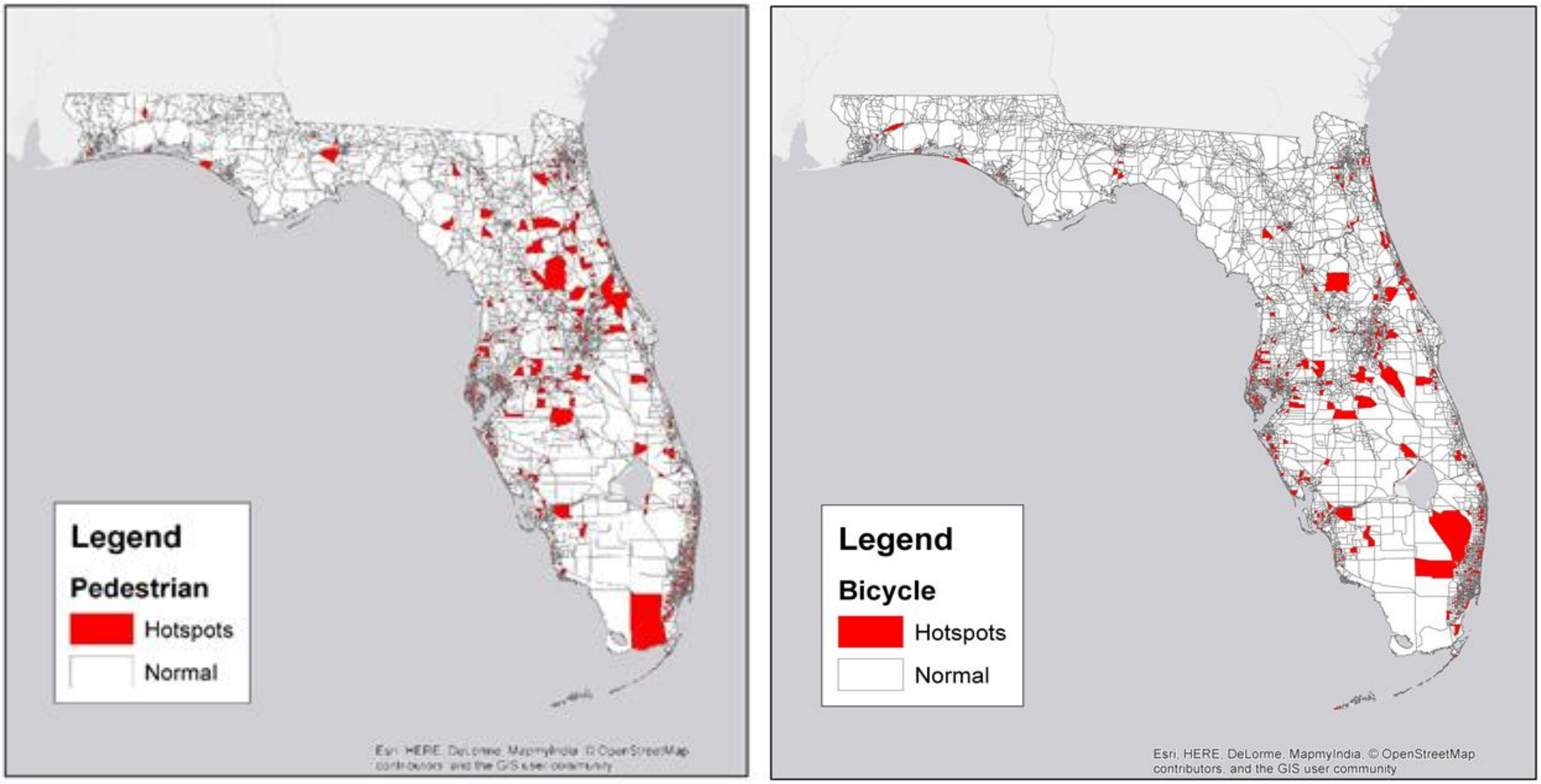
We present the identified hotspots in Figure 2. From the spatial hotspot distribution we can see that hotspots for both pedestrian and bicycle crashes are dispersed throughout Florida. Also we can see that risk of getting involved in pedestrian-motor vehicle or bicycle-motor vehicle crashes is higher in most urban zones. This spatial illustration can be used to prioritize STAZs for enhancing non-motorized road users' safety in high crash risk zones.

CONCLUSIONS

This paper formulated and estimated a multivariate count model by adopting a copula based bivariate negative binomial model for pedestrian and bicyclist crash frequency analysis. To the best of the authors' knowledge, this is the first attempt to employ such copula based bivariate count models for safety literature. Moreover, the study contributes to safety literature by examining the influence of several exogenous variables (exposure measures, socioeconomic characteristics, road network characteristics and land use attributes) on pedestrian and bicycle crash count events at the Statewide Traffic Analysis Zone (STAZ) level for the state of Florida. The empirical analysis involves estimation of models by using six different copula structures: 1) Gaussian, 2) FGM, 3) Clayton, 4) Gumbel, 5) Frank and 6) Joe. The comparison between copula and the independent models, based on information criterion metrics, confirmed the importance of accommodating dependence between pedestrian and bicycle crash count events in the macro-level analysis. The most suitable copula model is obtained for Clayton copula with parametrization for dependence profile. The model estimates were also augmented by conducting policy analysis including elasticity analysis and a spatial representation of hotspots for pedestrian and bicycle separately. Elasticity effects indicated that exogenous variables exhibit differences for the expected number of pedestrian and bicycle crash counts. Moreover, the most significant variable in terms of increase in the expected number of both pedestrian and bicycle crash counts included: VMT, total population and total employment. The spatial distribution of hotspots indicated that higher pedestrian and bicycle crash prone zones are dispersed throughout Florida with evidence of clustering along the urban zones. Overall, the policy analysis conducted provided an illustration on how the proposed model can be applied to determine the critical factors contributing to increase in pedestrian and bicycle crash counts.

The paper is not without limitations. In modeling pedestrian and bicyclist crashes we did not have access to pedestrian and non-motorized exposure. To accommodate for this, in our effort, we employed surrogate measures such as population density, VMT and proportion of heavy vehicles. It would be useful to compile pedestrian and bicyclist exposure data to enhance the model frameworks developed in our work.

FIGURE 2 Spatial distribution of Hotspots for Pedestrian and Bicycle Crash Risk of Florid



REFERENCES

1. NHTSA. *Traffic Safety Facts 2013: A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System*. Publication DOT HS 812 139. National Highway Traffic Safety Administration (NHTSA), U.S. Department of Transportation, 2015.
2. Jovanis, P. P., and H. Chang. Modeling The Relationship of Accidents to Miles Traveled. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1068*, Transportation Research Board of the National Academies, Washington, D.C., 1986, pp. 42-51.
3. Miaou, S.-P., and H. Lum. A Statistical Evaluation of the Effects of Highway Geometric Design on Truck Accident Involvements. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1407,* Transportation Research Board of the National Academies, Washington, D.C., 1993, pp. 11–23.
4. Abdel-Aty, M., and E. Radwan .Modeling Traffic Accident Occurrence and Involvement. *Accident Analysis & Prevention*, Vol. 32, 2000, pp. 633–642.
5. Miaou, S.-P., J. J. Song, and B. K. Mallick. Roadway Traffic Crash Mapping: A Space-Time Modeling Approach. *Journal of Transportation and Statistics,* Vol. 6, 2003, pp. 33-58.
6. Aguero-Valverde, J., Jovanis, P.P. Analysis of Road Crash Frequency with Spatial Models. . *In Transportation Research Record: Journal of the Transportation Research Board, No. Transportation Research Record, 2061*, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp.55-63.
7. Lord, D., and L.F. Miranda-Moreno. Effects of Low Sample Mean Values and Small Sample Size on the Estimation of the Fixed Dispersion Parameter of Poisson-gamma Models for Modeling Motor Vehicle Crashes: A Bayesian Perspective. *Safety Science*, Vol. 46, No. 5, 2008, pp. 751-770.
8. Maher, M., and L. Mountain. The Sensitivity of Estimates of Regression to the Mean. *Accident Analysis & Prevention*, Vol. 41, No. 4, 2009, pp. 861-868.
9. Cheng, L., S. R. Geedipally, and D. Lord. The Poisson–Weibull Generalized Linear Model for Analyzing Motor Vehicle Crash Data. *Safety Science,* Vol. 54, 2013, pp. 38-42.
10. Peng, Y., D. Lord, and Y. Zou. Applying the Generalized Waring Model for Investigating Sources of Variance in Motor Vehicle Crash Analysis. *Accident Analysis & Prevention,* Vol. 73, 2014, pp. 20-26.
11. Bhat C. R., and N. Eluru. A Copula-Based Approach to Accommodate Residential Self Selection Effects in Travel Behavior Modeling. *Transportation Research Part B: Methodological*. Vol. 43. No. 7, 2009, pp. 749-765.
12. Yasmin. S., N. Eluru, A. R. Pinjari and R. Tay .Examining Driver Injury Severity in Two Vehicle Crashes - A Copula Based Approach,. *Accident Analysis & Prevention*, Vol. 66, 2014, pp. 120-135.
13. Wang K., S. Yasmin, K. C. Konduri, N. Eluru, and J. N. Ivan, A Copula Based Joint Model of Injury Severity and Vehicle Damage in Two-Vehicle Crashes Presented at the *Transportation Research Board (TRB) Annual Meeting, Washington D.C.,* 2015.
14. Sener, I.N., N. Eluru, and C.R. Bhat. On Jointly Analyzing the Physical Activity Participation Levels of Individuals in a Family Unit. *Journal of Choice Modeling,* Vol 3, No. 3, 2010, pp. 1-38.

15. Ferdous, N., N. Eluru, C.R. Bhat, and I. Meloni. A Multivariate Ordered Response Model System for Adults' Weekday Activity Episode Generation by Activity Purpose and Social Context, *Transportation Research Part B*, Vol. 44, No. 8-9, 2010,pp. 922-943.
16. Bhat, C.R. The Composite Marginal Likelihood (CML) Inference Approach with Applications to Discrete and Mixed Dependent Variable Models. *Foundations and Trends in Econometrics*, Vol. 7, No. 1, 2014, pp. 1-117.
17. Narayanamoorthy, S., R. Paleti, and C.R. Bhat .On Accommodating Spatial Dependence in Bicycle and Pedestrian Injury Counts by Severity Level, *Transportation Research Part B*, Vol. 55, 2014, pp. 245-264.
18. Bhat, C.R. The Maximum Approximate Composite Marginal Likelihood (MACML) Estimation of Multinomial Probit-Based Unordered Response Choice Models. *Transportation Research Part B*, Vol. 45, No. 7, 2011, pp. 923-939.
19. Ma, J., and K. Kockelman. Bayesian Multivariate Poisson Regression for Models of Injury Count by Severity. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1950*, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp.24-34.
20. Park, E.S., and D. Lord, Multivariate Poisson-Lognormal Models for Jointly Modeling Crash Frequency by Severity. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2019*, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 1–6.
21. El-Basyouny, K., and T. Sayed. Collision Prediction Models Using Multivariate Poisson-Lognormal Regression. *Accident Analysis & Prevention*, Vol. 41, No. 4, 2009, pp. 820-828.
22. Lee, J., M. Abdel-Aty, and X. Jiang. Multivariate Crash Modeling for Motor Vehicle and Non-Motorized Modes at the Macroscopic Level. *Accident Analysis & Prevention*, Vol. 78, 2015, pp. 146-154.
23. Kim, D.G., and Washington, S. The Significance of Endogeneity Problems in Crash Models: An Examination of Left-Turn Lanes in Intersection Crash Models. *Accident Analysis & Prevention*, Vol. 38, No. 6, 2006, pp. 1094-1100.
24. Ye, X., R. M. Pendyala, S.P. Washington, K. Konduri, and J. Oh. A Simultaneous Equations Model of Crash Frequency by Collision Type for Rural Intersections, *Safety Sci*ence, Vol. 47, No. 3, 2009, pp. 443–452.
25. Cameron, A. C., Li, T., Trivedi, P. K., and Zimmer, D. M. Modelling the Differences in Counted Outcomes Using Bivariate Copula Models with Application to Mismeasured Counts. *The Econometrics Journal*, Vol. 7, No. 2, 2004, pp. 566-584.
26. Eluru, N., R. Paleti, R.M. Pendyala, and C.R. Bhat. Modeling Multiple Vehicle Occupant Injury Severity: A Copula-Based Multivariate Approach. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2165*, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 1-11.
27. Rana T., S. Sikder, and A. Pinjari. A Copula-Based Method to Address Endogeneity in Traffic Crash Injury Severity Models: Application to Two-Vehicle Crashes. *In Transportation Research Record: Journal of the Transportation Research Board, No. 2147,* Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 75–87.
28. Yasmin. S., N. Eluru, A. R. Pinjari, and R. Tay ,Examining Driver Injury Severity in Two Vehicle Crashes - A Copula Based Approach. *Accident Analysis & Prevention*, Vol. 66, 2010, pp. 120-135.

29. Wang K., S. Yasmin, K. C. Konduri, N. Eluru and J. N. Ivan, A Copula Based Joint Model of Injury Severity and Vehicle Damage in Two-Vehicle Crashes. Presented at the *Transportation Research Board (TRB) Annual Meeting, Washington D.C., 20*15.
30. Bhat, C.R., and N. Eluru , A Copula-Based Approach to Accommodate Residential Self-Selection Effects in Travel Behavior Modeling, *Transportation Research Part B*, Vol. 43, No. 7, 2009, pp. 749-765.
31. Trivedi, P. K., and D. M. Zimmer. *Copula Modeling: An introduction for Practitioners. Foundations and Trends in Econometrics*, Now Publishers Inc. 2007.
32. Sklar, A. Random Variables, Joint Distribution Functions, and Copulas. *Kybernetika*, Vol. 9, No. 6, 1973, pp. 449- 460.
33. GAUSS. Aptech Systems, Inc, Chandler, Arizona, 2012.
34. Quinn, C. The Health-Economic Applications of Copulas: Methods in Applied Econometric Research. Health. *Econometrics and Data Group (HEDG)*, Working Paper 07/22,2007, Department of Economics, University of York, 2007.
35. Elvik, R. The Non-Linearity of Risk and the Promotion of Environmentally Sustainable Transport. *Accident Analysis & Prevention*, Vol. 41, No. 4, 2009, pp. 849-855.
36. Ukkusuri, S., L. F. Miranda-Moreno, G. Ramadurai, and J. Isa-Tavarez. The Role of Built Environment on Pedestrian Crash Frequency. *Safety Science*, Vol. 50, No. 4, 2012, pp. 1141-1151.
37. LaScala, E. A., D. Gerber, and Gruenewald, P. J. Demographic and Environmental Correlates of Pedestrian Injury Collisions: A Spatial Analysis. *Accident Analysis & Prevention*, Vol. 32, No. 5, 2000, pp. 651-658.
38. Loukaitou-Sideris, A., R. Liggett, and H.-G. Sung. Death on the Crosswalk A Study of Pedestrian-Automobile Collisions in Loss Angeles. *Journal of Planning Education and Research*, Vol. 26, No. 3, 2007, pp. 338-351.
39. Cottrill, C. D., and P. V. Thakuriah. Evaluating Pedestrian Crashes in Areas with High Low-Income or Minority Populations. *Accident Analysis & Prevention*, Vol. 42, Vol. 6, 2010, pp. 1718-1728.
40. Siddiqui, C., M. Abdel-Aty, and K. Choi. Macroscopic Spatial Analysis of Pedestrian and Bicycle Crashes. *Accident Analysis & Prevention*, Vol. 45, 2012, pp. 382-391.
41. Martinez, R., and A. V. Richard. A Challenge in Injury Prevention-The Hispanic Population. *Academic Emergency Medicine*, Vol. 3, No. 3, 1996, pp. 194-197.
42. Wier, M., J. Weintraub, E. H. Humphreys, E. Seto, and R. Bhatia. An Area-Level Model of Vehicle-Pedestrian Injury Collisions with Implications for Land Use and Transportation Planning. *Accident Analysis & Prevention*, Vol. 41, No. 1, 2009, pp. 137-145.
43. Givoni, M., and P. Rietveld. The Access Journey to the Railway Station and its Role in Passengers’ Satisfaction with Rail Travel. *Transport Policy*, Vol. 14, Vol. 5, 2007, pp. 357-365.
44. Wei, F., and G. Lovegrove. An Empirical tool to evaluate the Safety of Cyclists: Community Based, Macro-Level Collision Prediction Models Using Negative Binomial Regression. *Accident Analysis & Prevention*, Vol. 61, 2013, pp. 129-137.
45. Pulugurtha, S. S., and V. Thakur. Evaluating the Effectiveness of On-Street Bicycle Lane and Assessing Risk to Bicyclists in Charlotte, North Carolina. *Accident Analysis & Prevention*, Vol. 76, 2015, pp. 34-41.
46. Qin, X., and J. Ivan. Estimating Pedestrian Exposure Prediction Model in Rural Areas. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1773*,

Transportation Research Board of the National Academies, Washington, D.C., 2001, pp. 89-96.

47. Kim, K., and E. Yamashita. Motor Vehicle Crashes and Land Use: Empirical Analysis from Hawaii. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1784*, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 73-79.
48. Lee, J., M. Abdel-Aty, and X. Jiang. Development of Zone system for Macro-Level Traffic Safety Analysis. *Journal of Transport Geography*, Vol. 38, 2014, pp. 13-21.
49. Eluru, N., and C.R. Bhat. A Joint Econometric Analysis of Seat Belt Use and Crash-Related Injury Severity. *Accident Analysis & Prevention*, Vol. 39, No. 5, 2014, pp. 1037-1049.