2016

Joint modeling of traffic related crashes: a Copula based approch

Tammam Nashad
University of Central Florida

JOINT MODELING OF TRAFFIC RELATED CRASHES: A COPULA BASED APPROACH

by

TAMMAM NASHAD
B.S. University of Benghazi, 2012

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science
in the Department of Civil, Environmental and Construction Engineering
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Spring Term
2016

Major Professor: Naveen Eluru

ABSTRACT

The study contributes to safety literature on transportation safety by employing copula based models 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 two multivariate models based on copula methodology. First, a copula based bivariate negative binomial model for pedestrian and bicyclist crash frequency analysis is developed. Second, a multivariate negative binomial model for crashes involving non-motorized road users, passenger cars, vans, light trucks and heavy trucks is proposed. The proposed approaches also accommodate for potential heterogeneity (across zones) in the dependency structure. The formulated models are estimated using traffic 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.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE: INTRODUCTION

Improving traffic safety was, is and will continue to be a high priority on the national transportation agenda due to the significant social and financial implications of motor vehicle crashes including injuries, deaths and economic losses among others. In the years 2010 and 2011 the fatalities rate per 100,000 Population in the state of Florida due to traffic related crashes was (12.97) and (12.58) respectively, which is clearly higher than the national Fatalities rate of (10.67) and (10.39) respectively (1).Moreover, 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 Florida is 2.56 (6.80). The statistics present a clear picture of the challenge faced in the state of Florida.

An important tool to identify the factors affecting occurrence of traffic related crashes; and identifying vulnerable locations is the application of planning level crash prediction models.

1.1 Thesis Structure

The remainder of the thesis is organized as follows. Chapter 2 discusses the earlier studies on modeling crash count. Chapter 3 focuses on joint modeling of traffic related crashes at the macro-level crash frequency by employing a copula based bivariate NB and multivariate NB modeling framework. Chapter 4 discusses the data source used and sample formation techniques in detail. Chapter 5 summarizes the results of the empirical application of the bivariate NB, introduces policy analysis and Spatial Distribution of Hotspot. Chapter 6.discusses model estimation results for multivariate NB. Chapter 7 presents the conclusions and recommendations based on the empirical results of the study.

CHAPTER TWO: LITERATURE RIVEW

Crash counts has been extensively researched in the safety analysis literature. This chapter reviews earlier crash counts studies that considered bivariate and multivariate analysis.

2.1 Earlier Research

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 offers a closed form expression while relaxing the equal mean variance equality constraint and serves as the workhorse for crash count modeling.

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 second dependent variable. 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 dependency 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 negative binomial 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 frequency 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 - has received considerable attention in transportation literature in recent years (14-16) ;(15). In terms of safety count modeling, the approach has been employed by (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) approaches or Markov Chain Monte Carlo (MCMC) 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 (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 (for example see (19), Poisson-lognormal models (for example see (20), (7), (21), (22) and simultaneous equation models (23), (24).

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 frequency variables. The current research effort contributes to literature on the first approach – developing a multivariate model by First, a copula based bivariate negative binomial model for pedestrian and bicyclist crash frequency analysis is developed. Second, a multivariate negative binomial model for crashes involving non-motorized road users, passenger cars, vans, light trucks and heavy trucks is proposed. The approach proposed here has been employed in econometrics (25). To the best of the authors' knowledge, this is the first attempt to employ such copula based bivariate and multivariate count models for safety literature. To be sure, copula models for ordered and unordered discrete outcome variables have been adopted in safety literature (see (26-29)). However, these approaches are not directly transferable to the count modeling. In this paper, we apply the copula based models for count frequency analysis. Empirically, the study examines the influence of several exogenous variables (exposure measures, socio-economic 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.

2.2 Summary

This chapter presented a summary of the existing bivariate and multivariate literature on traffic crash counts. This is the first attempt to employ such copula based bivariate or multivariate count models for safety literature. The next chapter presents the methodology adopted in this study.

CHAPTER THREE: METHODOLOGY

The focus of our study is to jointly model the macro-level pedestrian crash frequency and bicycle crash frequency by employing a copula based bivariate NB modeling framework. Also jointly model the macro-level traffic related crashes frequency by employing a copula based multivariate NB modeling framework. The econometric framework for the joint model is presented in this section.

Let us assume that 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:

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

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: $\mu_{qis} = E\left(y_{qi}|\boldsymbol{x}_{qi}\right) = exp\left(\boldsymbol{\beta}_q\boldsymbol{x}_{qi}\right)$, 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) and

(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:

$$\begin{aligned}\Lambda_{12}(y_{1i}, y_{2i}) &= Pr(\, U_1 \leq y_{1i},\, U_2 \leq y_{2i}) \\ &= Pr[\Lambda_1^{-1}(U_1) \leq y_{1i},\, \Lambda_2^{-1}(U_2) \leq y_{2i}\,] \qquad (2) \\ &= Pr[U_1 < \Lambda_1(y_{1i}),\, U_2 < \Lambda_2(y_{2i})\,]\end{aligned}$$

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

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

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 (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:

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

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:

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

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, Farlie-Gumbel-Morgenstern (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.

3.1 Summary

The present chapter described in detail the econometric framework employed in modeling traffic related crash counts in the present study. Next Chapter will discuss how the data was collected in detail

CHAPTER FOUR: DATA DESCRIPTION

In the previous chapter the econometric framework employed in the current study for modeling traffic related crashes is described in detail. In the present chapter, the data source employed for the empirical analysis of traffic related crashes is discussed. The section also explains how the data is aggregated to the STAZ level

4.1 Data Source

This study is focused on traffic related crashes at the STAZ level. There are 8,518 STAZs in the State of Florida (Figure 4.1). Data for the empirical study is sourced 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, socio-economic characteristics, road network characteristics and land use attributes. The exposure measures, socio-economic 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), U.S. Census Bureau, and Florida Geographic Data Library (FGDL). Table 4.1 offers a summary of the sample characteristics of the count and exogenous variables. Table 4.1 also represents the definition of variables[1] considered for final model estimation along with the zonal minimum, maximum and average values for Florida. From Table 4.1, we can see that for the three years the state of 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.

4.2 Summary

In the present chapter, the data source employed for the empirical analysis of traffic related crashes is discussed. The section also explains how the data is aggregated to the STAZ level. The results of the empirical application of the modeling framework are presented in the subsequent chapter.

[1] In estimating the models, several functional forms and variable specifications are explored. The functional form that gave the best result is used for final model specifications and, in Table 2, the variable definitions are presented based on these final functional form of variables.

Table 4.1 Sample Statistics for the State of Florida

Variable Names	Variables Descriptions	Minimum	Maximum	Average
Dependent variable				
Pedestrian crashes per STAZ	Total number of pedestrian crashes per STAZ	0	39	1.907
Bicycle crashes per STAZ	Total number of bicycle crashes per STAZ	0	88	1.797
Non- Motor crashes per STAZ	Total number of Non- Motor crashes per STAZ	0	89	2.931
Car crashes per STAZ	Total number of Car crashes per STAZ	0	1063	49.279
Van crashes per STAZ	Total number of Van crashes per STAZ	0	139	7.016
Light Trucks crashes per STAZ	Total number of Light Trucks crashes per STAZ	0	505	22.861
Med Trucks crashes per STAZ	Total number of Med Trucks crashes per STAZ	0	57	2.352
Exposure measures				
VMT	Natural Log of vehicle miles travel (VMT) in STAZ	0	13.437	9.039
Proportion of heavy vehicles	Total heavy vehicle VMT in STAZ /Total vehicles VMT in STAZ	0	0.519	0.067
Total population	Natural log of total population in STAZ	0	10.571	6.437
Proportion of families with no vehicle	Total number of families with no vehicle in STAZ/Total number of families in STAZ	0	1	0.095
Socio-economic characteristics				
Bicycle commuters	Natural log of total bicycle commuters in STAZ	0	6.654	0.847
Public transit commuters	Natural log of total commuters using public transportation in STAZ	0	6.841	1.416
Walk commuters	Natural log of total walk commuters in STAZ	0	7.162	1.629
Total employment	Natural log of total employment in STAZ	0	10.371	5.857
Proportion of industrial employment	Total number of industrial employment in STAZ/Total number of employment in STAZ	0	1	0.176
School enrollment density	Natural Log of total school enrollment per square miles in STAZ	0	12.45	2.715
Road network characteristics				
Proportion of urban area	Total urban area in STAZ/Total area in STAZ	0	1	0.722
Proportion of local roads	Total length of local roads in STAZ/Total length of all roads in STAZ	0	1	0.572
Proportion of arterial roads	Total length of arterial roads in STAZ/Total length of all roads in STAZ	0	1	0.221
Traffic signal density	Natural log of total number of traffic signals per miles of road in STAZ	0	8.756	0.227
Sidewalk length	Natural log of total length of sidewalk miles in STAZ	0	3.284	0.477
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	10.392	1.549
Distance to nearest urban area	Distance of the STAZ to the nearest urban area in miles	0	44.101	2.14

CHAPTER FIVE: EMPIRICAL ANALYSIS AND BIVARIATE RESUALTS

The results of the empirical analysis carried out on the data described on the previous chapter are presented in this chapter for the bivariate model. In addition this chapter provide policy analysis and Spatial Distribution of Hotspot.Two models are estimated in the present study. The first model is a bivariate NB model (pedestrian and bicyclist) is discussed in this chapter and the second model is a multivariate NB is discussed in the following chapter.

5.1 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 (a detailed discussion of these copulas is available in (30). 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 (see (31), (33), and (26)). 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.

5.2 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 5.2 presents the estimation results of the joint model. For the ease of presentation, the pedestrian crash count component (3^{rd} and 4^{th} columns of Table 5.2) and bicycle crash count component (5^{rd} and 6^{th} columns of Table 5.2) results are discussed together in the following section by variable groups. The copula parameters are presented in the last row panel of Table 5.2.

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 is represents the higher crash risk faced by non-motorized (pedestrian and bicyclist) road user groups with increasing VMT (34). Further, the results in Table 5.2 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 (see (35-37)).

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 38, 39), 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 (40).

Socio-economic 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 (41). 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 (42). 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 event relative to pedestrian crash count events. As found in previous studies (38, 40), 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 (43, 44), 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.

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 (45, 46). Moreover, tourists/visitors might be unfamiliar/less familiar with local driver behavior and road regulations (47), 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 5.2 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.

5.3 Policy Analysis

5.4 Elasticity Effects and Implications

The parameter effects of exogenous variables in Table 5.2 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 investigate the effect as 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 5.3 (see (48) for a discussion on the methodology for computing elasticities).

The following observations can be made based on the elasticity effects presented in Table 3. First, the results in Table 5.3 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.

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.

5.5 Spatial Distribution of HotSpot

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 hotspot in Figure 5.2. 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 based on for enhancing non-motorized road user's safety features of these high crash risk zones.

Table 5.1 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		

Table 5.2 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
Socio-economic 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

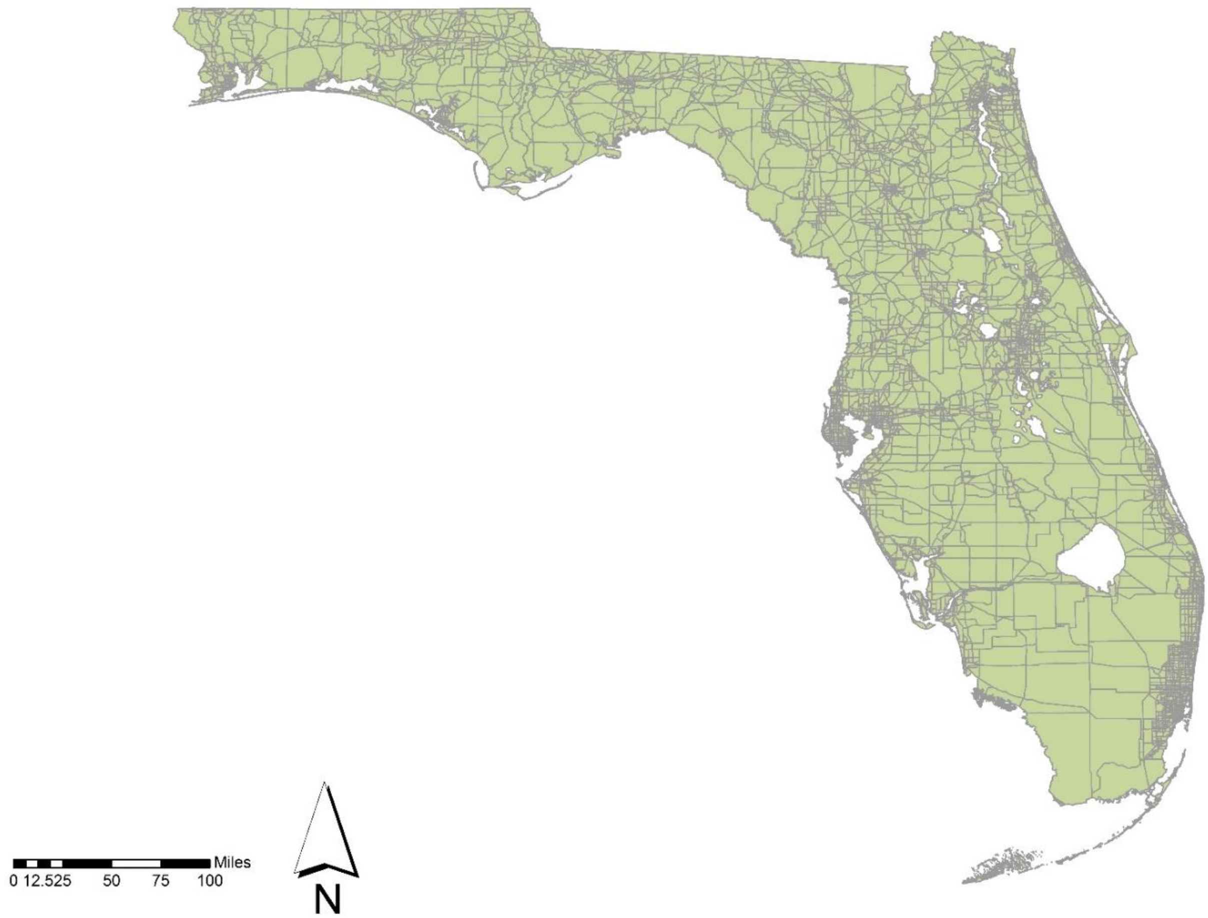

Figure 5.1 State Traffic Analysis Zones (STAZs) for the state of Florida

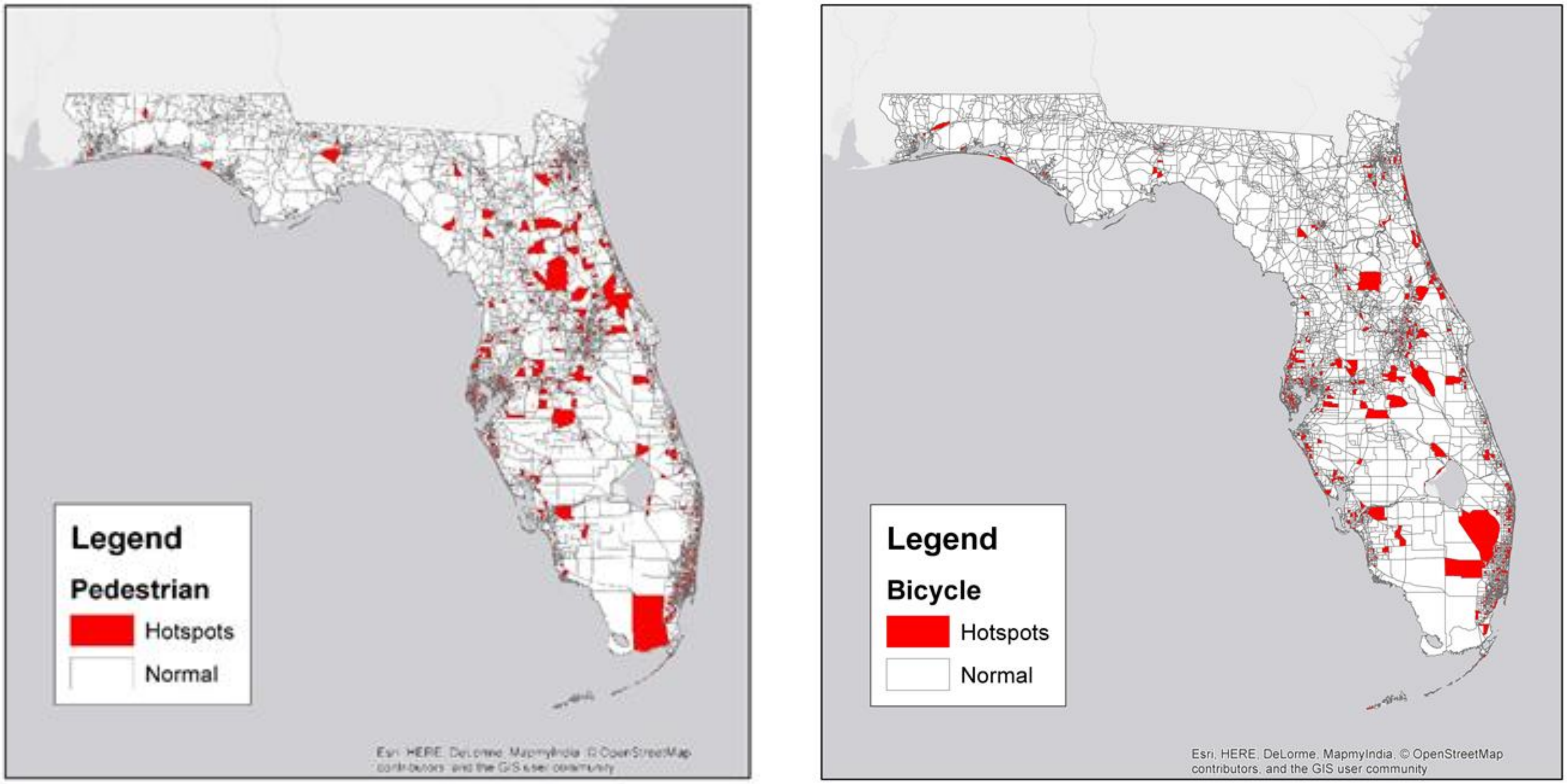

Figure 5.2 Spatial distribution of Hotspots for Pedestrian and Bicycle Crash Risk of Florida

CHPTER SIX: MULTIVARIATE RESULTS

Model estimation results of the bivariate NB model was introduced in the previous chapter along with some policy analysis and spatial distribution of Hotspots. In this chapter the model estimation of the multivariate model is presented

6.1Estimation Results

In presenting the effects of exogenous variables in the joint model specification, we will restrict ourselves to the discussion of the Frank Copula specification. Table 6.1 presents the estimation results of the joint model. For the ease of presentation, the non-motorized crash count component (3^{rd} and 4^{th} columns of Table 6.1), car crash count component (5^{rd} and 6^{th} columns of Table 6.1), van crash count component (7^{rd} and 8^{th} columns of Table 6.1), light truck crash count component (9^{rd} and 10^{th} columns of Table 6.1), and heavy truck crash count component (11^{rd} and 12^{th} columns of Table 6.1). Results are discussed together in the following section by variable groups.

<u>Exposure measures:</u> In terms of exposure measures, the estimates indicate that traffic related 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 (34). Further, the results in Table 7 indicate reduced crash propensity for all road users (motorized and non-motorized) with higher proportion of heavy vehicle VMT at the zonal level. With respect to total population, the joint model estimation results reveal that traffic related crashes are positively associated with higher zonal population (see 35-37). Also the model estimations show a positive correlation

between population density and crashes involving both vans and cars. In terms of square miles the results reveal a higher probability for cars, vans heavy trucks, and light trucks crashes in the STAZs with higher square mileage.

Socio-economic characteristics: The results for the number of commuters based on different commute modes are also found to be significantly influencing the traffic related crash risk in current study context. An increase in number of transit and walk commuters increases the likelihood of traffic related crashes at the STAZ level. The result in non- motor 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 (41). Moreover, bicycle exposure from higher cycle-transit integrated mode share (popularly known as "bike-and-ride") for access and egress at transit stations (42). For non-motorized, the results can be related to the frequent stops made by public transit especially when there is no designated public transit lane. It is interesting to note that non-motorized commute variable has a larger impact on both the number of transit and walk commuter count events relative to non-motorized crash count events. The result associated with bicycle commuters is positively correlated with crashes for non- motor users. However, it is negatively correlated with vans and heavy trucks. Our study also found that more employment within a TAZ leads to a higher probability of non-motor, van and heavy trucks. However, increasing proportion of industrial employment has negative association with non-motor crashes at the STAZ level but not on heavy trucks. Also, an increase in school enrollment density in a STAZ increases the likelihoods of crash risk in count model components for both non-motorized and motorized road user group.

Road network characteristics: Proportion of urban area, reflects that an increase in the proportion of urban area in a zone increases the likelihood of traffic related crash risk. The results associated with functional class of roadways show that pedestrian, bicycle, and vans' crash risk are positively correlated with higher proportion of arterial and local roads while both light and heavy trucks are negatively correlated. Consistent with several previous studies (43, 44), our study results also show that higher density of signalized intersections is positively associated with more traffic related crashes. With respect to bike lane length, the model estimation results indicate higher likelihood of non-motorized and car crashes with increasing length of bike lane in a zone. In terms of sidewalk length the results shows that an increase of the sidewalk length will increase the probability of car and heavy trucks.

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 non-motorized, cars and vans' crash risk, presumably indicating higher level of non-motorized road user activity in the proximity of these facilities in a zone (45, 46). Moreover, tourists/visitors might be less familiar with local driver behavior and road regulations (47), which might further exacerbate crash risk for these road user groups. The possibilities of traffic related crash risk decreases 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 increases the exposure of non-motorized road user groups resulting in increased likelihood of crash risks.

Table 6.1 Traffic related crashes Joint Model Estimation Results – Frank Copula

Variable Names	Non-Motor		Car		Van		Light Trucks		Heavy Trucks	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Constants	-6.3627	-31.56	0.7475	22.916	-2.3967	-38.249	0.7475	22.916	-0.3777	-3.856
Alpha	0.4571	35.396	0.6326	63.923	0.6066	46.362	0.6657	55.556	0.6708	39.558
Exposure measures										
VMT	2.57	54.149	2.174	77.258	2.2562	61.04	2.174	77.258	2.267	45.497
Square Miles	-	-	0.3608	8.608	0.2539	6.006	0.3608	8.608	0.402	8.832
Proportion of heavy vehicles	-2.1472	-9.22	-2.1008	-13.662	-1.1739	-6.258	-2.1008	-13.662	-	-
Population density	0.1534	25.596	-	-	0.0435	8.055	-	-	-	-
Total population	0.4258	10.281	-	-	0.3896	7.712	-	-	-	-
Proportion of families with no vehicle	-	-	-	-	-	-	-	-	0.6549	8
Socio-economic characteristics										
Bicycle commuters	3.6352	12.079	-	-	-0.0358	-4.444	-	-	-0.0453	-4.939
Public transit commuters	3.5891	17.295	0.1351	24.351	0.0909	13.626	0.1351	24.351	0.0599	8.283
Walk commuters	1.4997	7.293	0.0977	16.843	0.0461	6.288	0.0977	16.843	-	-
Total employment	0.1434	20.319	-	-	0.1165	20.613	-	-	0.1061	16.105
Proportion of service employment	-	-	-	-	-	-	-	-	-	-
Proportion of commercial employment	-0.0726	-2.053	0.2537	7.545	-	-	0.2537	7.545	-0.2152	-4.666
Proportion of industrial employment	-0.0924	-2.1	-	-	-	-	-	-	0.6553	15.661
School enrollment density	0.0136	4.669	0.035	12.341	0.0162	5.235	0.035	12.341	-	-
Road network characteristics										
Proportion of urban area	0.3913	10.875	0.4885	18.57	0.4178	13.752	0.4885	18.57	0.0905	3.024
Proportion of local roads	-	-	-	-	-	-	-	-	-	-
Proportion of collector roads	1.9697	10.463	-0.3913	-14.552	-	-	-0.3913	-14.552	-2.0915	-23.262
Proportion of arterial roads	2.3191	12.29	-	-	0.6474	15.38	-	-	-1.7359	-20.164
Traffic signal density	0.1789	10.113	0.1713	8.736	0.2212	12.041	0.1713	8.736	0.0986	4.432
Bike Lane length	0.0927	4.093	0.0328	1.274	-	-	0.0328	1.274	-	-
Sidewalk length	-	-	0.0951	5.713	-	-	0.0951	5.713	0.0206	1.135
Land use attributes										
Density of hotel/motel/timeshare room	0.023	6.251	0.0185	5.394	-	-	0.0185	5.394	-	-
Distance to nearest urban area	-0.0535	-1.764	-0.3403	-18.812	-0.2745	-11.236	-0.3403	-18.812	-0.173	-6.988
over desperstion	7.3159									

CHAPTER SEVEN: CONCLUSIONS

This thesis formulated and estimated a multivariate count model by developing two multivariate models based on copula methodology. First, a copula based bivariate negative binomial model for pedestrian and bicyclist crash frequency analysis is developed. Second, a multivariate negative binomial model for crashes involving non-motorized road users, passenger cars, vans, light trucks and heavy trucks is proposed. 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, socio-economic characteristics, road network characteristics and land use attributes) on traffic related 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.

APPENDIX: TABLES

APPENDIX 5.3 Pedestrian-Bicycle Poisson Model Estimation Results

Variable Names	Pedestrian		Bicycle	
	Estimate	**t-stat**	**Estimate**	**t-stat**
Constants	-4.000	-41.921	-4.017	-40.978
Exposure measures				
VMT	0.193	26.525	0.219	28.173
Proportion of heavy vehicles	-0.934	-3.410	-3.474	-10.914
Proportion of families with no vehicle	1.385	19.281	0.477	5.577
Socio-economic characteristics				
Bicycle commuters	0.023	4.008	0.140	23.455
Public transit commuters	0.208	37.945	0.110	19.813
Walk commuters	0.082	12.133	0.104	14.914
Total employment	0.166	19.817	0.141	16.437
Proportion of industrial employment	-0.278	-6.039	-0.234	-4.884
School enrollment density	0.016	5.527	0.010	3.507
Road network characteristics				
Proportion of urban area	0.185	4.288	0.562	11.040
Proportion of local roads	0.723	14.333	0.562	14.197
Proportion of arterial roads	0.155	2.460	-	-
Traffic signal density	0.226	13.251	0.178	9.749
Sidewalk length	0.264	19.260	0.254	18.431
Land use attributes				
Density of hotel/motel/timeshare room	0.020	5.783	-	-
Distance to nearest urban area	-0.049	-8.718	-0.096	-10.258

APPENDIX 5.4 Pedestrian-Bicycle Negative Binomial Model Estimation Results

Variable Names	Pedestrian		Bicycle	
	Estimate	**t-stat**	**Estimate**	**t-stat**
Constants	-4.350	-35.553	-4.274	-32.588
Exposure measures				
VMT	0.124	15.002	0.136	15.292
Proportion of heavy vehicles	-0.892	-2.389	-3.275	-7.556
Total population	0.141	11.986	0.146	11.587
Proportion of families with no vehicle	1.337	11.713	0.360	2.806
Socio-economic characteristics				
Bicycle commuters	0.029	3.015	0.138	13.595
Public transit commuters	0.171	20.149	0.098	10.634
Walk commuters	0.062	5.984	0.081	7.371
Total employment	0.178	15.208	0.124	10.171
Proportion of industrial employment	-0.238	-3.499	-0.169	-2.304
School enrollment density	0.012	2.889	0.012	2.644
Road network characteristics				
Proportion of urban area	0.265	4.748	0.637	9.763
Proportion of local roads	0.572	8.317	0.545	7.332
Proportion of arterial roads	0.314	3.720	0.337	3.708
Traffic signal density	0.289	10.828	0.189	6.388
Sidewalk length	0.269	12.332	0.297	12.807
Land use attributes				
Density of hotel/motel/timeshare room	0.029	5.288	0.016	2.735
Distance to nearest urban area	-0.038	-6.309	-0.085	-8.556

APPENDIX 6.2 Traffic related crashes Poisson Model Estimation Results

Variable Names	Non-Motor		Car		Van		Light Trucks		Heavy Trucks	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Constants	-7.152	-31.995	1.990	236.299	-4.402	0.056	-1.479	-51.953	-3.749	-32.430
Exposure measures										
VMT	2.818	39.179	-	-	4.339	0.042	4.313	177.075	3.677	44.709
Square Miles	0.210	4.540	-	-	0.405	0.022	-	-	0.106	4.592
Proportion of heavy vehicles	-2.079	-8.520	-	-	-	-	-2.001	-27.189	4.388	27.429
Population density	0.165	19.580	0.117	85.647	0.229	0.005	-	-	-	-
Total population	0.146	8.022	-	-	0.465	0.010	-	-	0.079	3.495
Proportion of families with no vehicle	0.916	13.698	-1.335	-69.472	-	-	-	-	0.900	11.838
Socio-economic characteristics										
Bicycle commuters	-	-	0.054	33.275	0.031	0.003	-	-	-0.020	-3.378
Public transit commuters	0.119	25.099	0.159	109.805	-	-	-	-	0.064	12.731
Walk commuters	0.084	15.522	0.140	83.972	-	-	0.117	64.856	0.027	4.554
Total employment	0.153	21.567	-	-	-	-	-	-	0.184	27.000
Proportion of service employment	-	-	-	-	-	-	-0.053	-5.058	-	-
Proportion of commercial employment	-0.169	-5.003	0.218	24.531	-	-	-	-	-0.310	-7.955
Proportion of industrial employment	-0.156	-3.843	-	-	-	-	-	-	0.507	14.482
School enrollment density	-	-	0.044	65.414	-	-	0.026	29.892	-	-
Road network characteristics										
Proportion of urban area	0.180	4.339	-	-	0.268	0.022	0.155	15.351	-	-
Proportion of local roads	-	-	-	-	-	-	-	-	-	-
Proportion of collector roads	1.973	10.028	-	-	-0.535	0.027	-0.764	-50.416	-1.260	-16.754
Proportion of arterial roads	2.186	11.226	-	-	-	-	-0.453	-36.514	-0.957	-13.484
Traffic signal density	0.256	19.249	0.227	79.678	0.225	0.006	0.275	64.679	0.214	14.347
Bike Lane length	-0.069	-4.860	-	-	-	-	0.021	3.926	-	-
Sidewalk length	0.319	26.943	-	-	-	-	0.175	40.485	0.145	11.603
Land use attributes										
Density of hotel/motel/timeshare room	0.021	7.562	-0.007	-8.601	-	-	-	-	-	-
Distance to nearest urban area	-0.406	-7.593	-	-	-0.367	0.027	-0.295	-26.723	-0.270	-10.576

APPENDIX 6.3 Traffic related crashes Negative Binomial Model Estimation Results

Variable Names	Non-Motor		Car		Van		Light Trucks		Heavy Trucks	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Constants	-7.0446	-25.201	0.5666	11.225	-3.1542	-34.44	-0.0998	-1.726	-1.047	-5.563
Exposure measures										
VMT	2.4728	29.227	2.0442	53.117	2.1779	38.609	2.5286	59.338	2.2102	24.559
Square Miles	-	-	0.4159	7.532	0.2322	4.431	0.4668	8.178	0.337	6.043
Proportion of heavy vehicles	-2.1951	-5.881	-2.1703	-9.205	-1.2117	-4.277	-2.1916	-8.787	-	-
Population density	0.1564	15.181	-	-	0.0584	7.45	-	-	-	-
Total population	0.7792	16.608	-	-	0.6686	14.186	1.3936	28.097	0.296	5.676
Proportion of families with no vehicle	-	-	-	-	-	-	-	-	0.5992	4.694
Socio-economic characteristics										
Bicycle commuters	3.6506	7.703	-	-	-0.0295	-3.135	-	-	-0.0317	-2.915
Public transit commuters	3.6235	12.408	0.1862	26.8	0.1275	16.028	-	-	0.0884	9.69
Walk commuters	1.4721	4.46	0.1185	15.404	0.0538	5.837	-	-	-	-
Total employment	0.2564	23.616	-	-	0.2102	25.496	-	-	0.2236	21.87
Proportion of service employment	-	-	-	-	-	-	-	-	-	-
Proportion of commercial employment	-0.2523	-4.038	0.2316	5.073	-	-	0.1397	2.84	-0.3588	-5.351
Proportion of industrial employment	-0.2527	-3.552	-	-	-	-	-	-	0.6623	10.391
School enrollment density	0.0162	3.857	0.0395	11.542	0.0162	4.186	0.0338	9.126	-	-
Road network characteristics										
Proportion of urban area	0.2329	4.004	0.3719	10.277	0.1836	4.115	0.2788	7.302	-0.2767	-5.984
Proportion of local roads	-	-	-	-	-	-	-	-	-	-
Proportion of collector roads	1.6779	6.623	-0.5081	-11.544	-	-	-0.6926	-13.804	-2.4044	-14.599
Proportion of arterial roads	2.1038	8.39	-	-	0.5214	8.104	-0.1561	-3.151	-1.8959	-11.993
Traffic signal density	0.3357	12.959	0.1774	9.305	0.3391	13.945	0.3734	14.002	0.203	7.142
Bike Lane length	0.1262	4.596	0.0764	2.794	-	-	-	-	-	-
Sidewalk length	-	-	0.2538	12.576	-	-	-	-	0.2059	8.554
Land use attributes										
Density of hotel/motel/timeshare room	0.0266	4.997	0.0206	4.718	-	-	0.0301	6.429	-	-
Distance to nearest urban area	-0.2574	-4.497	-0.468	-18.276	-0.3743	-9.601	-0.3282	-12.103	-0.2765	-6.892
over desperstion	0.5257	30.568	0.7197	59.472	0.6487	43.629	0.7783	56.812	0.7085	32.511

REFRENCES

1. NHTSA. Traffic Safety Facts 2013. Report No. DOT HS, Vol. 812 139, 2015, pp.
2. Paul P. Jovanis , Hsin-li Chang.,1986 .Modeling the relationship of accidents to miles traveled In *Transportation Research Record*, No. 1068, TRB, *National Research Council*
3. S.-P. Miaou, H. Lum “A statistical evaluation of the effects of highway geometric design on truck accident involvements.*Transportation Research Record*, 1407 (1993), pp. 11–23
4. M. Abdel-Aty, E. Radwan .Modeling traffic accident occurrence and involvement. *Accident Analysis & Prevention*, 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. (2008). Analysis of Road Crash Frequency with Spatial Models. *Transportation Research Record* 2061, 55-63.
7. Lord, D., and L.F. Miranda-Moreno (2008) 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, 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 Board (TRB) Annual Meeting, Washington D.C., 2009*

12. Yasmin. S., N. Eluru, A. R. Pinjari and R. Tay (2014).Examining Driver Injury Severity in Two Vehicle Crashes - A Copula Based Approach,. *Accident Analysis and Prevention* Vol. 66, 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 (2010), "On Jointly Analyzing the Physical Activity Participation Levels of Individuals in a Family Unit," *Journal of Choice Modeling,* Vol 3, No. 3, pp. 1-38

15. Ferdous, N., N. Eluru, C.R. Bhat, and I. Meloni (2010), "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, pp. 922-943

16. Bhat, C.R. (2014), "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, pp. 1-117

17. Narayanamoorthy, S., R. Paleti, and C.R. Bhat (2013), "On Accommodating Spatial Dependence in Bicycle and Pedestrian Injury Counts by Severity Level," *Transportation Research Part B*, Vol. 55, pp. 245-264
18. Bhat, C.R. (2011), "The Maximum Approximate Composite Marginal Likelihood (MACML) Estimation of Multinomial Probit-Based Unordered Response Choice Models," *Transportation Research Part B*, Vol. 45, No. 7, pp. 923-939
19. Ma, J., and Kockelman, K. (2006). Bayesian multivariate Poisson regression for models of injury count, by severity. *Transportation Research Record* 1950, pp.24-34

20. Park, E.S., Lord, D., 2007. Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. Transportation Research Record 2019, 1–6.
21. El-Basyouny, K., and T. Sayed. Collision Prediction Models Using Multivariate Poisson-Lognormal Regression. *Accident Analysis and 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. (2006). The Significance of endogeneity problems in crash models: An examination of left-turn lanes in intersection crash models. *Accident Analysis and Prevention*, Forthcoming.
24. X. Ye, R.M. Pendyala, S.P. Washington, K. Konduri, J. Oh “A simultaneous equations model of crash frequency by collision type for rural intersections” *Saf. Sci.*, 47 (3) (2009), pp. 443–452

25. Modeling the Differences in Counted Outcomes using Bivariate Copula Models: with Application to Mismeasured Counts" (with Tong Li, Pravin K. Trivedi and David M. Zimmer), *Econometrics Journal, December 2004*, 7, 566-584.

26. Eluru, N., R. Paleti, R.M. Pendyala, and C.R. Bhat (2010), "Modeling Multiple Vehicle Occupant Injury Severity: A Copula-Based Multivariate Approach," *Transportation Research Record*, Vol. 2165, 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 2147, TRB, National Research Council, Washington*, D.C., pp. 75–87. 2010.

28. Yasmin. S., N. Eluru, A. R. Pinjari and R. Tay (2014), "Examining Driver Injury Severity in Two Vehicle Crashes - A Copula Based Approach," *Accident Analysis and Prevention* Vol. 66, 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., 201*

30. Bhat, C.R., and N. Eluru (2009), "A Copula-Based Approach to Accommodate Residential Self- Selection Effects in Travel Behavior Modeling", *Transportation Research Part B*, Vol. 43, No. 7, pp. 749-765

31. Trivedi, P. K., Zimmer, D. M., 2007. Copula modeling: An introduction for practitioners. Foundations and Trends in Econometrics 1(1), Now Publishers.

32. Sklar, A., 1973. Random variables, joint distribution functions, and copulas. *Kybernetika* 9, 449- 460.

33. Quinn, C., 2007. The health-economic applications of copulas: Methods in applied econometric research. Health, Econometrics and Data Group (HEDG) Working Paper 07/22, Department of Economics, University of York

34. Elvik, R. (2009). The non-linearity of risk and the promotion of environmentally sustainable transport. *Accident Analysis & Prevention*, 41(4), 849-855.
35. Ukkusuri, S., Miranda-Moreno, L. F., Ramadurai, G., & Isa-Tavarez, J. (2012). The role of built environment on pedestrian crash frequency. *Safety Science*, 50(4), 1141-1151.
36. LaScala, E. A., Gerber, D., & Gruenewald, P. J. (2000). Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis. *Accident Analysis & Prevention*, 32(5), 651-658.
37. Loukaitou-Sideris, A., Liggett, R., & Sung, H.-G. (2007). Death on the crosswalk A study of pedestrian-automobile collisions in loss Angeles. *Journal of Planning Education and Research*, 26(3), 338-351.
38. Cottrill, C. D., & Thakuriah, P. V. (2010). Evaluating pedestrian crashes in areas with high low-income or minority populations. *Accident Analysis & Prevention*, 42(6), 1718-1728.
39. Siddiqui, C., Abdel-Aty, M., & Choi, K. (2012). Macroscopic spatial analysis of pedestrian and bicycle crashes. *Accident Analysis & Prevention*, 45, 382-391.
40. Martinez, Ricardo, and Richard A. Veloz. "A Challenge in Injury Prevention-The Hispanic Population." *Academic Emergency Medicine* 3.3 (1996): 194-197.

41. Wier, M., Weintraub, J., Humphreys, E. H., Seto, E., & Bhatia, R. (2009). An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. *Accident Analysis & Prevention*, 41(1), 137-145.
42. Givoni, M., & Rietveld, P. (2007). The access journey to the railway station and its role in passengers' satisfaction with rail travel. *Transport Policy*, 14(5), 357-365.
43. 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.
44. Pulugurtha, S. S., & Thakur, V. (2015). Evaluating the effectiveness of on-street bicycle lane and assessing risk to bicyclists in Charlotte, North Carolina. *Accident Analysis & Prevention*, 76(0), 34-41.
45. Qin, X., & Ivan, J. (2001). Estimating Pedestrian Exposure Prediction Model in Rural Areas. *Transportation Research Record: Journal of the Transportation Research Board*, 1773, 89-96.
46. Kim, K., & Yamashita, E. (2002). Motor Vehicle Crashes and Land Use: Empirical Analysis from Hawaii. *Transportation Research Record: Journal of the Transportation Research Board*, 1784, 73-79.
47. Lee, J., Abdel-Aty, M., & Jiang, X. (2014). Development of zone system for macro-level traffic safety analysis. *Journal of Transport Geography*, 38, 13-21.
48. Eluru, N., and C.R. Bhat (2007), "A Joint Econometric Analysis of Seat Belt Use and Crash-Related Injury Severity", *Accident Analysis & Prevention*, Vol. 39, No. 5, pp. 1037-1049