

1
2
3 **Analysis of Crash Proportion by Vehicle Type at Traffic Analysis Zone Level:**
4 **A Mixed Fractional Split Multinomial Logit Modeling Approach with Spatial**
5 **Effects**
6
7
8
9

10 **Jaeyoung Lee***
11 **Shamsunnahar Yasmin**
12 **Naveen Eluru**
13 **Mohamed Abdel-Aty**
14 **Qing Cai**
15
16

17 * Corresponding Author
18
19
20

21 Department of Civil, Environmental and Construction Engineering
22 University of Central Florida
23 Orlando, Florida 32816-2450
24 (407) 823-2733
25 jaeyoung@knights.ucf.edu
26
27
28
29
30
31
32
33
34

35
36 May 2017
37
38

39 Revised for Possible Publication in Accident Analysis and Prevention
40
41

1 **ABSTRACT**

2

3 In traffic safety literature, crash frequency variables are analyzed using univariate count models or
4 multivariate count models. In this study, we propose an alternative approach to modeling multiple crash
5 frequency dependent variables. Instead of modeling the frequency of crashes we propose to analyze the
6 proportion of crashes by vehicle type. A flexible mixed multinomial logit fractional split model is employed
7 for analyzing the proportions of crashes by vehicle type at the macro-level. In this model, the proportion
8 allocated to an alternative is probabilistically determined based on the alternative propensity as well as the
9 propensity of all other alternatives. Thus, exogenous variables directly affect all alternatives. The approach
10 is well suited to accommodate for large number of alternatives without a sizable increase in computational
11 burden. The model was estimated using crash data at Traffic Analysis Zone (TAZ) level from Florida. The
12 modeling results clearly illustrate the applicability of the proposed framework for crash proportion analysis.
13 Further, the Excess Predicted Proportion (EPP) – a screening performance measure analogous to Highway
14 Safety Manual (HSM), Excess Predicted Average Crash Frequency is proposed for hot zone identification.
15 Using EPP, a statewide screening exercise by the various vehicle types considered in our analysis was
16 undertaken. The screening results revealed that the spatial pattern of hot zones is substantially different
17 across the various vehicle types considered.

18

19

20 **Key words:** multinomial logit fractional split model; traffic crash analysis; macroscopic crash analysis;
21 traffic analysis zones; vehicle type; screening

22

1 1. INTRODUCTION

2
3 The Federal, State, and Local government officials and transportation engineers have been working with
4 consistent efforts in reducing both road crash related fatalities and fatality rates. In the United States, traffic
5 collisions have steadily declined from 2003 to 2011. However, traffic fatalities and fatality rates rose in
6 2012 with slight drop in 2013 highlighting the challenges faced by the safety community. Overall 29,989
7 people were reported to be killed in the United States from traffic crashes in 2014 (NHTSA, 2016). These
8 facts highlight that there is a need for continued efforts to identify remedial measures to reduce crash
9 occurrences and crash consequences. Towards this end, the traffic safety literature has evolved along two
10 main dimensions: collision frequency analysis and collision severity analysis. The former group of studies
11 is focused on identifying factors that result in traffic collisions while the latter group is concentrated on
12 ameliorating the consequences of traffic crashes (conditional on their occurrences). The current study
13 contributes to traffic safety literature along the first dimension – identify factors that inform us about traffic
14 collision occurrences.

15
16 Collision frequency analysis is traditionally undertaken at the microscopic and macroscopic levels. The
17 microscopic safety analysis focuses on roadway entities such as segments, intersections, and corridors (Lee
18 et al., 2017). The studies broadly aim to identify contributing factors for traffic crashes from roadway
19 geometric design, traffic characteristics, and provide specific engineering countermeasures to alleviate
20 traffic collisions. On the other hand, the macroscopic safety analysis relates traffic crashes aggregated at a
21 spatial level (traffic analysis zone (TAZ), census tract or county) with demographic, socioeconomic, built
22 environment, traffic attributes and/or roadway characteristics at a study unit level. While microscopic level
23 analysis is more focused on the engineering design and evaluation, the macroscopic analysis provides a
24 broad spectrum for long-term policy based countermeasures such as enactments of traffic laws, police
25 enforcement, education, and area-wide engineering solutions (Lee et al., 2014b). There has been growing
26 recognition within the planning community to incorporate macroscopic models as part of long range
27 transportation plans. For example, the Moving Ahead for Progress in the 21st Century Act (MAP-21) and
28 Fixing America's Surface Transportation Act (FAST) have emphasized the role of macro-level safety
29 analysis in planning.

30
31 While total traffic fatalities and fatality rates show a downward trend, NHTSA statistics indicate that the
32 proportions of motorcycles, bicycles, and pedestrians in fatal crashes have considerably increased whereas
33 the proportion of passenger cars has decreased from 2005 to 2014 (NHTSA, 2016). The proportions of
34 motorcyclists and non-motorists involved in fatal crashes has risen from 11% to 14% and 13% to 18%,
35 respectively, since the last decade. Despite these increases, earlier studies in the safety area have
36 predominantly focused either on total crashes or crashes involving passenger cars/trucks or crashes
37 involving non-motorists. These studies provide important information in improving safety situation for
38 different road user groups separately. However, it is also important to examine critical factors contributing
39 to crash occurrences including all road user groups in a single framework, which would allow stakeholders
40 to devise a more general safety conscious planning. Towards that end, the main objective of this study is to
41 explore the proportions of traffic crashes at TAZ level across different vehicle types including both
42 motorized and non-motorized group of road users. Specifically, the current study considers the proportion
43 of crashes by vehicle type as the dependent variable and estimates a TAZ level mixed multinomial fractional
44 split model. The proposed approach will assist transportation planners and engineers in devising safety

1 conscious plans. Specifically, traffic engineers and planners can understand the factors affecting the
2 proportions of crashes by vehicle types from the modeling results, and use these findings to design long-
3 term transportation plans. Furthermore, the predicted proportions of crashes from the proposed model can
4 be used to identify hotspots for each vehicle type. Thus, traffic engineers and planners can proactively
5 provide effective safety countermeasures for the zones with excessive high proportions for specific vehicle
6 types. The vehicle types considered in our analysis include: passenger car¹, van, light truck, medium and
7 heavy truck, bus, motorcycle, bicycle, and pedestrian. The reader would note that the model employed is
8 not similar to the traditional multinomial logit model because the dependent variable in our case is
9 proportion by vehicle type whereas it is a single chosen alternative in the multinomial logit model.

11 *1.1. Literature Review*

13 The macro-level safety studies have been conducted by total crashes, crashes by severity levels (such as no
14 injury, minor injury, severe injury, and fatal injury) and crashes by vehicle type (such as motor vehicle,
15 pedestrian, and bicycle). It is beyond the scope of this paper to exhaustively review all the studies in
16 frequency modeling (see Lord and Mannering (2010) and Yasmin and Eluru (2016) for a detailed review).
17 In our study, we group literature in the context of our research effort along two main groups: 1) independent
18 frequency models for a single dependent variable or multiple dependent variables; and 2) multivariate count
19 models for the multiple dependent variables are estimated.

21 In the first group of studies, usually either total number of crashes in the study unit or crashes by vehicle
22 type or severity level are investigated. In some studies, multiple dependent variables are considered in the
23 analysis while ignoring the relationship across the dependent variables. For example, Noland and Quddus
24 (2004) developed fixed-effect negative binomial models for bicycle and pedestrian crashes for severe and
25 minor injury. Nevertheless, these models did not account for the possible correlation among dependent
26 variables. Lee et al. (2013) estimated a series of negative binomial models for total, severe, driving under
27 the influence (DUI), pedestrian and bicycle crashes based on zero-inflated Poisson (ZIP) framework while
28 ignoring the relationships across the dependent variables. Abdel-Aty et al. (2013) analyzed the contributing
29 factors for total, and severe pedestrian crashes, using negative binomial models without considering
30 common unobserved factors across different crash types.

32 In the second group of studies, multivariate models that recognize the dependencies between multiple
33 dependent variables are estimated. For example, Song et al. (2006) analyzed intersection, intersection-
34 related, driveway access, and non-intersection crashes at the county level. The authors developed Bayesian
35 multivariate conditional autoregressive models that can account for the dependencies between the variables
36 and spatial effects. (Narayanamoorthy et al., 2013) explored pedestrian and bicycle crash frequencies by
37 injury severity levels. The authors adopted a multivariate model to accommodate jointness in the dependent
38 variables, while considering spatial dependence effects. Furthermore, Lee et al. (2015b) estimate
39 multivariate Poisson lognormal models to analyze motor-vehicle, pedestrian, and bicycle crashes. The
40 authors found that the multivariate model accounting for unobserved common factors across the dependent
41 variables outperforms the univariate model that does not take the dependencies between dependent

¹ Passenger cars include sedans and sport utility vehicles (SUV).

1 variables into consideration. Recently, Nashad et al. (2016) adopted a bivariate copula based modeling
2 approach to examine pedestrian and bicyclist crashes simultaneously.

3
4 Based on earlier literature, the factors that are likely to affect motor vehicle crashes are socio-demographic
5 factors, land-use characteristics, roadway-related variables, and traffic characteristics (Kim et al., 2006; Lee
6 et al., 2015b). Based on 0.1 mi² grid structure, Kim et al. (2006) estimated various regression models based
7 on the grid. The authors identified that population, total job, total economic output, and commercial area
8 had positive relationship with motor vehicle crashes. Lee et al. (2015b) analyzed motor vehicle crashes
9 based on TAZs. The authors found that population, total employment, total economic output, commercial
10 area, roadway types, proportion of households without a vehicle, number of accommodation facilities per
11 square mile, and number of traffic signals per mile have a positive effect on motor vehicle crashes. Very
12 few researchers have explored truck-involved crashes at a macro-level. Pasupuleti and Pulugurtha (2013)
13 identified several zonal characteristics related to truck crashes. The authors revealed that truck crashes are
14 positively correlated to industrial areas and areas with large residential lots but negatively correlated with
15 highly populated areas.

16
17 Several factors that tend to increase bicycle crashes have been identified by researchers. Noland and Quddus
18 (2004) found that national health service staff per population, percentage of motorway, percentage of trunk
19 road density, percentage of older vehicle, percentage of households without cars, per capita expenditure on
20 alcohol, population, percentage of population aged 65 or over have a positive relationship with severe
21 bicycle crashes. On the other hand, length of inpatient stay in the hospital, income level, percentage of
22 population aged 45-64 have a negative relationship with severe bicycle crashes. Kim et al. (2006) attempted
23 several candidate variables including population, total job, total economic output, hospital, park,
24 commercial area, and school for bicycle crashes but only population was found significant and positively
25 related to bicycle crashes. Lee et al. (2013) investigated the residence of bicyclists who were involved in
26 traffic crashes. It was revealed that median age, average travel time to work, household income, and workers
27 in the primary industry field were negatively associated with the number of crash involved bicyclists. On
28 the other hand, Hispanic people, workers commuting by bicycle, urban area, and older buildings were
29 positively associated with the number of bicyclists who were involved in traffic crashes. Lee et al. (2015b)
30 developed a multivariate model for motor vehicle-to-vehicle, bicycle-to-vehicle, and pedestrian-to-vehicle
31 crashes. The authors uncovered that vehicle-miles-traveled, population, commuters using bicycle,
32 hotel/motel/timeshare rooms per square mile, employments and school enrollments per square mile, number
33 of traffic signals per mile are likely to increase bicycle crashes whereas proportion of roadway with speed
34 limit of 20 mph or less tends to decrease bicycle crashes. The above mentioned literature suggest that the
35 factors that tend to increase bicycle crashes are population, distance to urban location, employment, school
36 enrollment density, roadway types, number of traffic signals per mile, proportion of households without a
37 vehicle, and household income.

38
39 Lastly, some factors that increase the propensity of pedestrian crashes have been found. Noland and Quddus
40 (2004) found that total population and percentage of population aged 65 or over are positively correlated
41 whereas percentage of other roads, income, percentage of population aged 45-64 are negatively correlated
42 with severe pedestrian crashes. Loukaitou-Sideris et al. (2007) discovered that population, employment
43 density, high traffic volume, commercial/retail land-use, and multifamily residential land-use have a
44 positive effect on pedestrian crashes. In the study of Wier et al. (2009), traffic volume, arterials without

1 transit, proportion of mixed land-use (commercial and residential), proportion of commercial land-use,
2 employee and resident populations, and proportion of people below poverty have a positive relationship
3 while land areas, and proportion of population aged 65 or over have a negative relationship with pedestrian
4 crashes. Lee et al. (2015a) used the product of population and VMT as an exposure variable in their model.
5 The authors found that proportion of households below poverty level, rail and bus station density,
6 accommodation density, marina/ferry terminal density, and K-12 school density tend to increase pedestrian
7 crashes; but proportion of high-speed roads have a tendency to decrease pedestrian crashes. To the best
8 authors' knowledge, none of studies have explored motorcycle or bus related crashes at a macro-level. In
9 the data preparation process, we consider the independent variables that have been found significant in the
10 previous studies.

11 *1.2. Current Study*

12
13
14 In earlier research, the impact of exogenous variables is quantified through the propensity component of
15 count models. The main interaction across different count variables is either ignored (first group) or sought
16 through unobserved effects (studies from second group) i.e. there is no interaction of observed effects across
17 the multiple count models. While this might not be a limitation per se, it might be beneficial to evaluate the
18 impact of exogenous variables in framework that directly relates a single exogenous variable to all count
19 variables of interest simultaneously i.e. a framework where the observed propensities of crashes by vehicle
20 type interact directly. In the traditional count modeling approaches this is not feasible. In this study, an
21 alternative approach to macro-level crash modeling is proposed. Specifically, as opposed to modeling the
22 number of crashes, we adopt a fractional split modeling approach to study the fraction of crashes by each
23 vehicle type for a zone. So for example, in a three-count variable case, the traditional approach would be to
24 adopt a trivariate count model framework with three count equations. In the proposed approach, we adopt
25 a multinomial fractional split model that examines the proportion of crashes (not frequency) by count type
26 with three equations representing the three crash types in a single probabilistic model system. So for a
27 zone, the dependent variable could take the following form – crash type 1: 0.30, crash type 2: 0.25 and
28 crash type 3: 0.45. The fractional split model is used to analyze the proportion by vehicle type across zones
29 as a function of exogenous variables. To be sure, the approach is not a replacement for traditional count
30 based approaches. We believe that the fractional split modeling approach would serve complementary to
31 existing traditional approaches in providing more insights on the impact of exogenous variables on crash
32 proportions.

33
34 The fractional split model also provides another advantage. From the review of earlier literature, it is evident
35 that vehicle crashes are usually grouped under one category. However, it is possible that the occurrence of
36 crashes might vary across zones by vehicle type (such as passenger cars, light trucks, medium and heavy
37 trucks, and buses). It is worthwhile to investigate the factors that influence crash occurrence by each vehicle
38 type. However, the addition of different vehicle types would add additional computational burden for count
39 modeling and hence such fine resolution of vehicle types is rarely considered. Within a fractional split
40 model, the additional computation burden associated with adding the fine vehicle type resolution is minimal
41 and thus facilitates considering detailed vehicle type resolution.

42
43 The proposed methodology is based on earlier work in econometrics undertaken by Papke and Wooldridge
44 (1993). The authors proposed a quasi-likelihood estimation method for binary probit model with a fractional

1 dependent variable. The authors explored 401(K) plan participation rates in two portfolios using their
2 proposed method. The approach was extended to multinomial fractional model by Sivakumar and Bhat
3 (2002). The authors analyzed statewide interregional commodity-flow volumes in Texas using the proposed
4 model. Eluru et al. (2013) extended the binary probit model by Papke and Wooldridge (1993) and proposed
5 a panel mixed ordered probit fractional split model to analyze vehicle operating speed on urban roads in
6 Montreal. Yasmin et al. (2016) employed an ordered version of the fractional split model to investigate
7 crash proportion by injury severity. To be sure, there has been earlier work in safety literature exploring the
8 multinomial fractional split model at the microscopic level. Milton et al. (2008) developed a mixed
9 multinomial fractional split model to study injury-severity distribution of crashes on highway segments
10 by using highway-injury data from Washington State. Also several other researchers adopted a fractional
11 split model in the respective fields (Nam, 2012; Witter et al., 2012; Wang & Wolman, 2014).

12
13 Furthermore, we consider the influence of observed spatial effects in our analysis. In the macro-level studies,
14 traffic crashes occurring in a geographic unit are aggregated. The aggregation process might create errors
15 in identifying exogenous variables for the geographic unit. For instance, a crash occurring near or on the
16 boundary of the geographic unit might be strongly related to the neighboring zone than the actual zone
17 where the crash happened, which is the result of arbitrarily defining boundaries. In order to alleviate such
18 geographic unit induced bias, the following two methodologies have been utilized to account for spatial
19 autocorrelations: 1) spatial error correlation effects (unobserved exogenous variables at one location affect
20 dependent variable at the targeted and adjacent zones); and 2) spatial spillover effects (observed exogenous
21 variables at one location impact the dependent variable at both the targeted and adjacent zones)
22 (Narayanamoorthy et al., 2013). Some research efforts have accommodated for spatial random error in
23 safety literature (Huang et al., 2010; Dong et al., 2014; Lee, 2014; Dong et al., 2015; Lee et al., 2015a; Lee
24 et al., 2015b; Dong et al., 2016; Huang et al., 2016; Xu et al., 2017). Nevertheless, using such spatially
25 lagged dependent variable models, specifically for prediction, is of limited use because observed crash at
26 adjacent geographic unit is required as an independent variable in the model. Thus, we adopted a method
27 considering exogenous variables from adjacent zones for accounting for spatial dependency, which was
28 recently suggested (Cai et al., 2016). To summarize, in this research we employ a zonal level mixed
29 multinomial fractional split model to investigate the impact of exogenous factors with spatial spillover
30 effects on the proportion of vehicle types in traffic crashes for the state of Florida.

31
32 The rest of the paper is organized as follows: Section 2 provides a description of the mixed multinomial
33 fractional split model. Section 3 describes the data collection and sample preparation steps. Section 4
34 discusses the modeling results and the elasticity effects. A new performance measure for hot zone
35 identification is proposed and screening results are provided in Section 5. Lastly, Section 6 summarizes and
36 concludes the paper.

37 38 **2. STATISTICAL FRAMEWORK**

39
40 The dependent variable in this study is defined as the proportion of vehicle type in traffic crashes by TAZ.
41 The sum of the proportions across a TAZ is equal to unity and each proportion of vehicle types in traffic
42 crashes ranges between zero and one. Let y_{mn} be the fraction of crashes by vehicle type m ($m= 1,2, \dots, M$;
43 $M=8$) in TAZ n ($n=1, 2, \dots, N$). In this paper, the eight vehicle types correspond to passenger car, van, light
44 truck, medium and heavy truck, bus, motorcycle, bicycle, and pedestrian.

$$0 \leq y_{mn} \leq 1 \quad \sum_{m=1}^M y_{mn} = 1 \quad (1)$$

Let the fraction y_{mn} be a function of a vector x_{mn} of relevant explanatory variables associated with attributes of TAZ n .

$$E[y_m|x] = G_m(x; \beta) \\ 0 < G_m(\cdot) < 1 \quad \sum_{m=1}^M G(\cdot) = 1 \quad (2)$$

where $G_m(\cdot)$ is a predetermined function. The properties specified in Equation (2) for $G_m(\cdot)$ warrant that the predicted fractional crash vehicle types will range between 0 and 1, and will add up to 1 for each TAZ. In this study, a mixed multinomial logit functional form for G_m in the fractional split model of Equation 2. Then Equation 2 is rewritten as:

$$E(y_m|x) = G_m(x; \beta) = \frac{\exp(x\beta_m)}{\sum_{m=1}^M \exp(x\beta_m)}, \quad m=1, \dots, M \quad (3)$$

Given the probability expression above, the quasi likelihood function is written as follows:

$$L_q(\beta) = \prod_{m=1}^M G_m(x_n; \beta)^{y_{mn}} \quad (4)$$

The quasi log-likelihood function for the sample is defined as: $\mathcal{L}(\beta) = \sum_{n=1}^N \ln(L_q(\beta))$

The model estimation is undertaken by maximizing the quasi log-likelihood function based on a routine in Gauss matrix programming language. The readers would note that the coefficient vector includes both mean parameters and standard deviation parameters following a normal distribution.

3. DATA PREPARATION

The number of road users involved in crashes for eight vehicle types (passenger car, van, light truck², medium and heavy truck, bus, motorcycle, bicycle and pedestrian) were acquired from the Florida Department of Transportation (FDOT) Crash Analysis Reporting System (CARS) for the year 2010 through 2012. The collected crash data by vehicle types were further aggregated based on statewide TAZs (N=8,518) and corresponding crash proportion were computed. Among 8,518 TAZs, there were no crashes reported for 389 TAZs for the study years and hence were excluded from the analysis of our study. The numbers of units by vehicle type in total crashes were aggregated based on TAZ-level. Then the TAZ-based unit counts by vehicle type were converted to the proportions. Table 1 summarizes the descriptive statistics of traffic crash related variables. From the table, we can observe that the proportion of passenger cars (57.5%) is the highest while the proportion of buses (1.1%) is the smallest among all vehicle types considered. The data for the dependent variable is further augmented by socio-demographic, traffic, roadway and commuter travel data from multiple sources such as FDOT CARS/Systems Planning Office/Roadway Characteristics Inventory, Florida Department of Revenue, and U.S. Census Bureau. The prepared independent variables were considered based on previous studies as discussed in the literature review section.

² The reader would note that not only passenger car but also van and light truck represent vehicles used for passenger transport.

1

Table 1: Descriptive statistics of traffic crash related variables

Variable (N=8,129)	Mean	Stdev	Min	Max	% of zero proportion
Total number of crash-involved units based on a TAZ	146.433	214.156	1	2773	-
Proportion of passenger cars	0.575	0.157	0.000	1.000	1.8%
Proportion of vans	0.061	0.056	0.000	1.000	18.3%
Proportion of light trucks	0.264	0.139	0.000	1.000	4.2%
Proportion of medium & heavy trucks	0.032	0.065	0.000	1.000	33.8%
Proportion of buses	0.011	0.030	0.000	1.000	53.6%
Proportion of motorcycles	0.029	0.062	0.000	1.000	33.0%
Proportion of bicycles	0.013	0.031	0.000	1.000	49.4%
Proportion of pedestrians	0.015	0.030	0.000	0.667	45.6%

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

Socio-demographic data such as population, family vehicle ownership, hotel/motel/timeshare rooms, employment, and school enrollment were obtained from the Systems Planning Office of the FDOT. Four employment related variables were processed: total employment density, and proportion by employment type - industrial, commercial, and service. Roadway/traffic data were compiled from the FDOT Roadway Characteristics Inventory (RCI) and they were processed at a TAZ level using geographical information systems (GIS). The roadway data include proportion of roadway length by functional classifications (i.e., arterial, collector, and local road), signals per mile, VMT density (VMT normalized by TAZ area), proportion of heavy vehicles, bike lane length density (per square mile), and sidewalk length density (per square mile). Two types of urban classification data were processed. Urban area polygon data (2010) was collected from the U.S. Census Bureau. The urban area data were processed into: 1) proportion of urban areas; and 2) distance to the nearest urban area. Commuting travel data were acquired from the U.S. Census Bureau and the proportions of commuters using specific travel modes (e.g., public transportation, bicycle, walking) were computed for each TAZ. Table 2 exhibits the descriptive statistics of the candidate explanatory variables.

1

Table 2: Descriptive statistics of candidate explanatory variables

Variable (N=8,129)		Mean	Stdev	Min	Max
Socio-demographic characteristics	Population density (per sqmi)	2574.39	4032.83	0	63069
	Proportion of young people (15-24 years)	0.132	0.083	0	1.000
	Proportion of elderly people (65 years and older)	0.171	0.116	0	0.938
	Hotel, motel, and timeshare room density (per sqmi)	173.89	951.75	0	32610
	Total employment density (per sqmi)	1190.3	1745.9	0	31932
	Proportion of industrial employments	0.176	0.232	0	1.000
	Proportion of commercial employments	0.299	0.235	0	1.000
	Proportion of service employments	0.492	0.259	0	1.000
	Proportion of families with no available vehicle	0.095	0.123	0	1.000
	School enrollment density (per sqmi)	803.4	6115.3	0	255147
Median household income (in USD)	57389	24714	0	215192	
Roadway /traffic characteristics	Arterial proportion of public road mileage	0.223	0.272	0	1.000
	Collector proportion of public road mileage	0.188	0.239	0	1.000
	Local proportion of public road mileage	0.572	0.324	0	1.000
	Number of signals per mile	3.032	88.134	0	6347.7
	Vehicle-miles-traveled density (per sqmi)	54304	187263	0	11469720
	Proportion of trucks	0.068	0.051	0	0.428
	Bike lane length density (per sqmi)	0.533	3.207	0	153.183
	Sidewalk length density (per sqmi)	3.542	10.245	0	183.373
	Presence of transit system (dummy variable)	0.597	0.491	0	1
	Lengths of transit system (mi)	1.474	2.857	0	59.580
	The existence of SIS (Strategic Intermodal System) network	0.310	0.463	0	1
	Number of bus stations	0.005	0.071	0	1
	Number of cargo centers	0.010	0.435	0	36
Number of commuter rail stations	0.009	0.107	0	3	
Number of train stations	0.002	0.048	0	1	
Land-use attributes	Proportion of urban areas	0.732	0.424	0	1.000
	Distance to the nearest urban area (mi)	1.991	5.209	0	44.101
	Mixed land-use area (sqmi)	0.004	0.017	0	0.636
	Residential area (sqmi)	1.265	6.020	0	280.114
	Commercial area (sqmi)	0.100	0.225	0	7.052
	Industrial area (sqmi)	0.045	0.428	0	23.902
	Agriculture area (sqmi)	3.428	13.707	0	227.568
	Institutional area (sqmi)	0.033	0.260	0	11.514
	Governmental area (sqmi)	1.756	15.665	0	748.162
	Miscellaneous area (sqmi)	0.279	4.098	0	338.046
Vacant area (sqmi)	1.495	13.261	0	702.884	
Number of nightclubs, cocktail lounges and bars	0.234	0.757	0	18	
Commuting characteristics	Proportion of commuters using public transportation	0.024	0.043	0	0.549
	Proportion of commuters using bicycle	0.009	0.023	0	0.309
	Proportion of commuters who walk	0.025	0.041	0	0.449
Indicator variables	District dummy variables (1 to 7)	Percentages			
	District 1	11.8%			
	District 2	17.1%			
	District 3	13.6%			
	District 4	12.1%			
	District 5	19.3%			
	District 6	10.5%			
District 7	15.6%				

2

3

4. RESULTS

4.1. Modeling Results

Table 3 presents the modeling results of the mixed multinomial logit fractional split model. The model estimation results revealed the absence of any significant unobserved parameters. Hence, the model collapsed to a multinomial logit fractional split model. Within the fractional split model, each alternative has a propensity equation. However, for the sake of identification, one of the alternative propensities has to be selected to serve as the base. In our analysis, the model estimation is undertaken considering that the passenger car option serves as the base alternative. Hence, there are no coefficients specific to passenger car in Table 3.

Overall, the model results are intuitive and offer credence to our hypothesis that a fractional split approach to crash proportion modeling is of value. A brief examination of the results highlights substantial differences in the number of significant parameters across alternatives. The propensity for light truck and medium and heavy truck alternatives has more number of significant parameters while pedestrian and bicyclist alternatives have the fewest significant parameters in current study context. Given that the multinomial logit fractional split approach is compensatory in nature (difference of propensities affects proportion), alternatives with smaller proportions (as is the case with pedestrian and bicyclist crashes) are likely to have fewer parameters. In terms of data fit, the additional exogenous variables substantially improve the quasi log-likelihood at convergence (-9738.217) compared to the quasi log-likelihood at zero (-16903.780) and quasi log-likelihood at constants (-9930.663). A log-likelihood ratio test with respect to model at constants yields a test statistic of 384.886 ($=2*(-9738.217 - (-9930.663))$). The test statistic is larger than the corresponding chi-square distribution value for 25 additional parameters at any level of significance. The results from the model are discussed subsequently by variable groups.

The reader would note that parameter values in Table 3 correspond to the impact of exogenous variables on the alternative relative to passenger car. Hence, a positive (negative) sign indicates increase (decrease) in the proportion of the alternative relative to the proportion of passenger cars. An insignificant effect implies the variable does not have any differential impact on the alternative relative to the passenger car alternative. The authors also discussed the explanatory variables used in the previous macro-level safety studies. At a first glance, some may think that the findings from prior studies are contradictory to those from this study. However, all the previous studies developed crash count models (by crash types or vehicle types) whereas this study estimated the proportions of vehicle types in traffic crashes, and thus they are not actually conflicting.

Constants

The constants in the model are negative for all alternatives with respect to the base alternative passenger cars. This is expected as the passenger car proportion is substantially larger than other alternatives.

Socio-demographic characteristics

Several sociodemographic variables influence the zonal level proportion of crashes by vehicle types. An increase in the variable, logarithm of population density indicates is associated with a reduction in the proportion of light truck and medium and heavy trucks involved crashes with the magnitude being higher

1 for medium and heavy trucks. In high population density zones, the exposure to light trucks is likely to be
2 much lower (relative to zones with lower population density). The lower exposure might result in lower
3 proportion of light truck involved crashes. Similarly, in high population density zones, the exposure to
4 medium and heavy trucks is lower (relative to other zones) and more importantly the average speed of
5 medium and heavy trucks is also likely much lower thus possibly reducing the proportion of their crash
6 involvement. No prior studies have utilized the population density for the proportion of trucks in traffic
7 crashes at the macro-level. Instead, some studies used the population density as an exploratory variable for
8 total crashes (Lovegrove & Sayed, 2007; Lee et al., 2014b; Xu et al., 2014), property damage only (PDO)
9 crashes (Ladron de Guevara et al., 2004), injury crashes (Ladron de Guevara et al., 2004; Lee et al., 2014b;
10 Xu et al., 2014), and fatal crashes (Ladron de Guevara et al., 2004), bicycle crashes and pedestrian crash
11 count models (Siddiqui et al., 2012; Cai et al., 2016). It was commonly found that the variable had positive
12 effects on those crashes.

13
14 The modeling results indicate that a higher proportion of elderly population tends to increase the crash
15 proportion of bicycles and motorcycles (among the crash involved units). Florida has the highest percentage
16 of senior people aged 65 or older and it has experienced consistent in-migration of retired people from other
17 states (Sperazza et al., 2012). The large number of retired people might escalate demand for recreational
18 and leisure activities including cycling and motorcycling. Thus, the higher percentage of elderly population
19 may be associated with the larger crash proportion of both bicycles and motorcycles. The elderly population
20 factor has been considered in the several macro-level studies. Huang et al. (2010) found that the larger
21 elderly population tends to decrease total and severe crash counts. Also, Lee et al. (2014a) accommodated
22 the proportion of elderly population in their study and the variable has a negative relationship with total
23 crash counts.

24
25 The variable corresponding to proportion of families with no available vehicle offers interesting results.
26 The variable, on its own, is associated with a reduction in the crash proportion of motorcycles. Motorcycle
27 is mainly a recreational vehicle in Florida and expensive to maintain, hence in such zones the likelihood of
28 ownership as well as the use of motorcycles is likely to be lower. In addition to the TAZ variable, a spatial
29 variable based on the proportion of families with no available vehicle from neighboring TAZs is also
30 considered in our model. The variable is associated with an increase in the crash proportion of buses (among
31 the crash involved units). Families without access to vehicles are captive to public transportation and
32 possibly live in zones with higher access to public transit. Hence, it is not surprising that such zones have
33 higher proportion of bus crashes. On the other hand, some research efforts employed vehicle ownership
34 variable in their studies for other crash types. The higher proportion of households without available vehicle
35 has a propensity to increase bicycle (Noland & Quddus, 2004), pedestrian (Noland & Quddus, 2004; Lee
36 et al., 2015a), and total/severe crashes (Lee et al., 2014b). Siddiqui et al. (2012) used the percentage of
37 households with 0 or 1 vehicle and identified that the variable was positively associated with bicycle and
38 pedestrian crashes. Quddus (2008) showed that the logarithm of households without vehicle has a positive
39 impact on fatal and serious injury crashes.

40
41 Tourism is an important industry in Florida and it is important to consider tourist presence in traffic safety
42 planning. A surrogate measure for tourist activity is hotel, motel and timeshare facility density. The variable
43 highlights a negative association with the proportion of light trucks in traffic crashes alluding to the
44 possibility that tourist areas are less likely to have larger number of light trucks in the zone. Although no

1 studies have considered the accommodation factor for the proportion of light trucks, some studies found
2 that the hotel units have a positive impact on pedestrian (Siddiqui et al., 2012; Lee et al., 2015a; Lee et al.,
3 2015b; Cai et al., 2016), bicycle (Lee et al., 2015b), total, severe (Lee et al., 2014b), and vehicle-to-vehicle
4 crashes (Lee et al., 2015b). On the contrary, Ng et al. (2002) showed that the hotel units have a negative
5 effect on fatal crashes.

6
7 The employment density variable is associated with a lower likelihood of motorcycle crashes. As
8 motorcycle is mainly used for recreational purposes, it is expected that the use of such vehicles in zones
9 with high employment density is unlikely. In terms of industrial employment, the result indicates that zones
10 with higher industrial employment is likely to involve higher medium and heavy truck crash proportions.
11 The variable is a reflection of increased exposure to medium and heavy trucks in these zones. The findings
12 are further substantiated based on the result of the proportion of trucks estimates. From the previous studies,
13 many researchers have found that employment-related factors have a significant effect on various crash
14 types. An area with larger employment has a tendency to exhibit a greater number of vehicle-to-vehicle and
15 severe crashes (Hadayeghi et al., 2006), pedestrian and bicycle crashes (Siddiqui et al., 2012; Cai et al.,
16 2016), and pedestrian crashes only (Loukaitou-Sideris et al., 2007). Levine et al. (1995) attempted diverse
17 employment variables. The authors found that manufacturing, retail trade, and service employment
18 positively affect crash counts while financial and military employment are negatively related to total crash
19 counts.

20 21 Roadway /traffic characteristics

22 The increased vehicle mileage in the zone is negatively associated with medium and heavy trucks and bus
23 crash proportions. The increased vehicle mileage reflects suburban zones where the exposure to medium
24 and heavy trucks or buses is very low. Hence, the trend is expected. In zones with increased exposure to
25 trucks, a higher incidence of van, light truck and medium and heavy truck crashes is likely with a
26 substantially larger impact on the proportion of medium and heavy trucks. It is also found that the increased
27 vehicle mileages in the neighboring zone have a negative effect on medium and heavy truck proportions.
28 While there have been no macro-level studies exploring the share of trucks in traffic crashes, Golob and
29 Regan (2003) uncovered the negative relationship between annual average daily traffic (AADT) level and
30 the probability of truck involvement on urban freeways at the micro-level. This result is quite consistent
31 with the finding from this study.

32 33 Land-use attributes

34 In terms of spatial location of the zone, proportion of urban areas and distance to the nearest urban location
35 exert significant impact on crash proportions. The zones with larger proportion of urban areas, as expected,
36 are likely to have higher incidence of bicycle crashes and lower proportion of light truck and motorcycle
37 crashes. The zones that are farther from urban areas are likely to have a higher proportion of light trucks as
38 these vehicles are more likely to be used in these zones. It was shown that several land-use characteristics
39 have substantial effects on crashes in prior macro-level research studies. Huang et al. (2010) showed that
40 the higher level of urbanization is positively related to total and severe crashes, and Siddiqui et al. (2012)
41 also showed that an urbanized zone experience the larger number of bicycle crashes.

42
43 Moreover, the agriculture area in the neighboring zone has a positive effect on the proportion of crash-
44 involved light trucks. Previous macro-level safety studies have not explicitly shown that the agriculture

1 area per se has a significant impact on traffic safety but some macroscopic safety studies showed that fatal
2 crashes are more likely to occur in rural areas (Blatt & Furman, 1998; Stamatiadis & Puccini, 2000; Clark,
3 2003).

4
5 Other than urban and land-use variables, we considered FDOT districts as an independent variable in the
6 model. The operations of the FDOT are organized into seven districts and their locations are shown in
7 Figure 1. The result indicates that District 2 has higher crash proportion of medium and heavy trucks while
8 District 4 has a propensity of less light truck crash proportion. District 2 is predominantly characterized by
9 vast rural areas (90.4%) whereas 41.1% area in District 4 is urbanized area.

10 11 Commuting characteristics

12 The zonal commuting characteristics exhibit influence on crash proportions. The zones with high proportion
13 of public transportation commuters are likely to have lower proportion of light truck crashes. The result is
14 an indication of lower light truck ownership in zones with prevalence of public transit ridership. In zones
15 with increased walking commuters, the proportion of pedestrian crashes increases (among the crash
16 involved units). The result is a classic case of higher number of pedestrians in the zone resulting in more
17 pedestrian crashes. In addition, the proportion of commuters using bicycle in adjacent zones has a tendency
18 to increase motorcycle crashes. Several studies have used commuter variables in their studies. Abdel-Aty
19 et al. (2013) found that both commuters by public transportation and walking commuters have considerable
20 impacts on total, severe, and pedestrian crashes. Lee et al. (2014a) found that zones with higher proportion
21 of commuters using non-motorized modes is more likely to have more total crashes *ceteris paribus*. Lee et
22 al. (2015b) exhibited that the logarithm of commuters using bicycle is positively related to the number of
23 bicycle crashes.

24 25 **4.2. Elasticity Effects**

26
27 The model results from Table 3 provide an indication of how the exogenous variables affect the proportion
28 of crashes involving different vehicle types. However, the exact magnitude of the impact on all alternatives
29 is not easily available. Hence, to evaluate the impact of exogenous variables on all crash proportions, we
30 resort to the computation of elasticity effects. The elasticity effects in our study are computed by evaluating
31 the change in crash proportions in response to increasing the value of significant exogenous variables by
32 10% (see Eluru and Bhat (2007) for more details on elasticity calculations). The results from the elasticity
33 exercise are presented in Table 4. The numbers presented in the table represent the percentage change in
34 the proportion of alternatives in response to a 10% increase in the exogenous variable. For example, the
35 value of elasticity for the variable logarithm of population density of 0.696 represents an increase in
36 passenger car crash proportion with a 10% increase in zonal population density. All the other numbers can
37 be interpreted similarly.

38 39 **4.3. Summary**

40
41 Based on the results, the following interesting observations may be made. First, the logarithm of population
42 density affects most significantly and negatively the proportion of medium and heavy truck crashes. Second,
43 the proportion of elderly people has a positive effect on bicycle proportions. Third, the proportion of
44 households with no available vehicle affects negatively the proportion of motorcycles whereas the

1 proportion of households with no available vehicle in adjacent zones has the only and positive effect on the
2 crash proportion of buses. Fourth, the logarithm of total employment density has the most significant and
3 negative influence on the motorcycle crash proportion. Fifth, an increase in vehicle-miles-traveled density
4 has a substantial negative impact on the proportion of medium and heavy trucks as well as buses. Sixth, the
5 proportion of trucks has a significant positive association with crash proportion of medium and heavy trucks.
6 Finally, the proportion of commuters who walk has a strong impact on crash proportion of pedestrians.

1 **Table 3: fractional split multinomial model results for proportion of vehicle types involved in traffic crashes**

Variables		Van	Light truck	Medium & heavy truck	Bus	Motorcycle	Bicycle	Pedestrian
Constant		-2.316 (-39.382)	-0.441 (-3.365)	-1.383 (-3.429)	-3.607 (-10.955)	-2.187 (-9.634)	-4.842 (-14.810)	-3.710 (-35.120)
Socio-demographic characteristics	Log of population density (per sqmi)	-	-0.031 (-1.865) [#]	-0.117 (-3.671)	-	-	-	-
	Proportion of population aged 65 and more	-	-	-	-	1.356 (2.625)	1.906 (2.891)	-
	Proportion of families with no available vehicle	-	-	-	-	-1.484 (-1.987)	-	-
	<i>Spatial</i> -Proportion of families with no available vehicle of neighboring TAZs	-	-	-	2.221 (3.487)	-	-	-
	Log of hotel, motel, and timeshare rooms density (per sqmi)	-	-0.023 (-1.912) [#]	-	-	-	-	-
	Log of total employment density (per sqmi)	-	-	-	-	-0.071 (-1.991)	-	-
	Proportion of industrial employment	-	-	0.572 (2.259)	-	-	-	-
Roadway /traffic characteristics	Log of VMT density (per sqmi)	-	-	-0.063 (-2.486)	-0.068 (-2.023)	-	-	-
	Proportion of heavy vehicles in VMT	1.074 (2.001)	1.074 (2.001)	6.832 (6.718)	-	-	-	-
	<i>Spatial</i> -Log of VMT density of neighboring TAZs (per sqmi)	-	-	-0.082 (-2.565)	-	-	-	-
Land-use attributes	Proportion of urban areas	-	-0.201 (-1.960)	-	-	-0.795 (-5.043)	0.848 (2.622)	-
	Distance to the nearest urban area (mi)	-	0.010 (1.700) [#]	-	-	-	-	-
	<i>Spatial</i> -Log of agriculture area of neighboring TAZs (sqmi)	-	0.028 (1.935) [#]	-	-	-	-	-
	District 2	-	-	0.306 (2.004)	-	-	-	-
	District 4	-	-0.228 (-2.652)	-	-	-	-	-
Commuting characteristics	Proportion of commuters using public transportation	-	-2.392 (-3.404)	-	-	-	-	-
	Proportion of commuters who walk	-	-	-	-	-	-	2.923 (1.763) [#]
	<i>Spatial</i> -Proportion of commuters using bicycle of neighboring TAZs	-	-	-	-	5.149 (1.743) [#]	-	-
Log-likelihood	Pseudo log-likelihood at 0	-16903.78						
	Pseudo log-likelihood at constant	-9930.663						
	Pseudo log-likelihood at convergence	-9738.217						

2 # significant at 90% confidence interval, and all other variables are significant at 95% confidence interval.

3 Numbers in parentheses are t-values of the estimated coefficients.

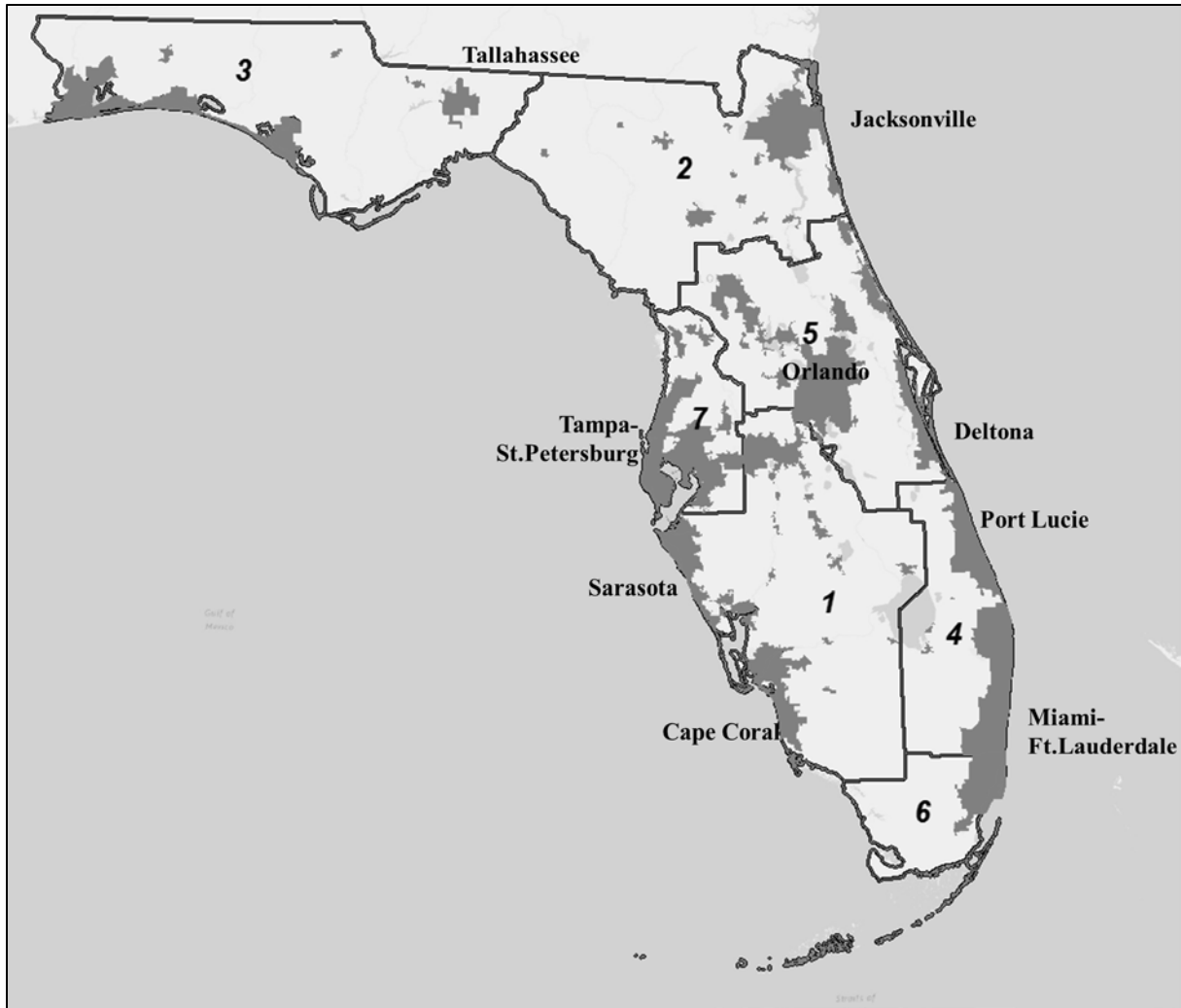


Figure 1: Districts boundaries (1-7) and urban areas (dark colored) in Florida

1
2
3

Table 4: Elasticity effects

Variables		Passenger car	Van	Light truck	Medium & heavy truck	Bus	Motorcycle	Bicycle	Pedestrian
Socio-demographic characteristics	Log of population density (per sqmi)	0.696	0.698	-1.201	-5.448	0.694	0.688	0.696	0.695
	Proportion of population aged 65 and more	-0.127	-0.126	-0.125	-0.118	-0.122	2.459	3.815	-0.127
	Proportion of families with no available vehicle	0.032	0.032	0.031	0.031	0.037	-1.082	0.032	0.034
	<i>Spatial</i> - Proportion of families with no available vehicle of neighboring TAZs	-0.036	-0.035	-0.029	-0.031	3.075	-0.025	-0.038	-0.042
	Log of hotel, motel, and timeshare rooms density (per sqmi)	0.086	0.085	-0.235	0.062	0.088	0.071	0.096	0.089
	Log of total employment density (per sqmi)	0.111	0.111	0.115	0.113	0.106	-3.780	0.111	0.110
	Proportion of industrial employment	-0.038	-0.040	-0.051	1.272	-0.038	-0.050	-0.028	-0.037
Roadway /traffic characteristics	Log of VMT density (per sqmi)	0.221	0.226	0.236	-4.876	-5.640	0.232	0.211	0.222
	Proportion of heavy vehicles in VMT	-0.415	0.262	0.237	5.991	-0.407	-0.536	-0.334	-0.407
	<i>Spatial</i> - Log of VMT density of neighboring TAZs (per sqmi)	0.262	0.270	0.313	-8.434	0.267	0.311	0.228	0.258
Land-use attributes	Proportion of urban areas	0.031	0.031	-0.062	0.031	0.028	-0.361	0.405	0.030
	Distance to the nearest urban area (mi)	-0.058	-0.062	0.178	-0.147	-0.061	-0.109	-0.028	-0.056
	<i>Spatial</i> -Log of agriculture area of neighboring TAZs (sqmi)	-0.063	-0.068	0.196	-0.179	-0.066	-0.132	-0.020	-0.060
	District 2	-0.981	-1.009	-1.147	31.370	-1.019	-1.146	-0.874	-0.972
	District 4	5.625	5.659	-15.828	6.344	5.539	6.112	5.332	5.574
Commuting characteristics	Proportion of commuters using public transportation	0.126	0.126	-0.349	0.113	0.157	0.097	0.128	0.134
	Proportion of commuters who walk	-0.016	-0.016	-0.013	-0.012	-0.021	-0.013	-0.017	0.953
	<i>Spatial</i> -Proportion of commuters using bicycle of neighboring TAZs	-0.013	-0.013	-0.011	-0.010	-0.014	0.425	-0.014	-0.015

1

2

6. HOT ZONE IDENTIFICATION FOR SPECIFIC VEHICLE TYPE

To facilitate the application of multinomial fractional split model for hot zone identification, we propose a measure based on the predicted proportions. The measure is analogous to the Highway Safety Manual (HSM) (AASHTO, 2010) performance measure for identifying crash hotspots. The HSM approach employs Excess Predicted Average Crash Frequency Using Safety Performance Functions. The measure is calculated by subtracting the predicted crash frequency from the observed crash frequency. When the excess predicted average crash frequency is greater than zero, a zone experiences more traffic crashes than predicted. On the other hand, when the excess predicted average crash frequency is less than zero, a zone experience fewer traffic crashes than predicted.

Similar to this method, we propose the Excess Predicted Proportion (EPP) for a macroscopic screening performance measure, which is the difference between the observed and predicted proportion of each vehicle type for a TAZ.

$$EPP_{mn} = P(obs)_{mn} - P(prd)_{mn} \quad (5)$$

where, EPP_{mn} is the Excess Predicted Proportion of crash vehicle type m at TAZ n . $P(obs)_{mn}$ is the observed proportion of crash vehicle type; and $P(prd)_{mn}$ is the predicted proportion of crash vehicle type estimated from the fractional split multinomial logit model (Table 3). When EPP exceeds zero, the type proportion for that TAZ is higher than predicted. In contrast, when EPP is smaller than zero, the type proportion for that TAZ is lower than predicted.

The EPP approach is slightly different from the earlier count based approach. Because we deal with proportions that add up to 1 for a TAZ, a positive EPP for one vehicle type automatically causes a negative EPP for at least another vehicle type. Hence, directly identifying the zones with positive EPP as hot zones will not be appropriate. Because every zone will be a hot zone for at least one mode (unless EPP is exactly 0 for all modes). Hence, after computing the measure across the entire sample for all vehicle types, a vehicle type specific hot zone (H) is identified based on the top 10 percentile ranking of TAZs for that vehicle type. The other zones are labeled as normal (N). An illustration of EPP computation for the passenger car mode is provided in Table 5. The TAZs are arranged in descending order of EPP values. Clearly, TAZs #5762, #6944, and #7882 were classified as a hot zone for passenger cars since their EPPs are within top 10% in the study area. Therefore, it is recommended to focus on passenger car-involved crashes at these zones, since the proportion of crash-involved passenger cars far exceeded the expected proportion of crashes. The EPP measure is a useful tool for policy makers and practitioners to understand potential safety risks by various vehicle types, and provide an adequate safety countermeasure solely for the problematic vehicle types. The zones identified as risky might be very different from zones identified as risky using the HSM hot zone identification process. We believe that this provides an alternative paradigm for safety analysis. The reader would note that analogous measures to excess expected proportion can also be computed using the proposed model.

1

Table 5: Examples of screening results of passenger cars

Rank based on passenger car EPP	TAZ No	Passenger car				
		P_{obs}	P_{pred}	EPP	percentile	classification
1	5762	1	0.317	0.683	0.01%	H
2	6944	1	0.373	0.627	0.02%	H
:	:	:	:	:	:	:
812	7882	0.800	0.670	0.130	9.99%	H
813	1474	0.617	0.487	0.130	10.00%	N
:	:	:	:	:	:	:
8129	6241	0	0.649	-0.679	100.00%	N

2

3 To further illustrate the value of the proposed framework, we identify hot zones across the state of Florida
4 for all vehicle types. To clarify the presentation a figure that depicts the main urban areas in the State of
5 Florida is provided (see Figure 1). In Figure 2 statewide screening results for all 8 vehicle types are
6 provided. An observation of Figure 2 clearly shows that the spatial pattern of hot zones varies considerably
7 across the various vehicle types. It is interesting to note that the spatial pattern of hot zones for passenger
8 cars, vans, light trucks vary substantially across the region. For passenger cars the hot zones correspond to
9 urban areas. The spatial distribution of van specific hot zones does not follow any particular pattern. On the
10 other hand, the spatial distribution of light truck specific hot zones clearly indicates a predisposition for
11 rural areas. For medium and heavy trucks, the hot zones are concentrated in central and south rural areas.
12 The overall bus hot zones tend to be located in urban and suburban areas with minor exceptions. Motorcycle
13 hot zones seem typically located in rural areas. The hot zones for bicyclists are apparently found in large
14 metropolitan areas whereas rural areas are relatively safe for bicyclists. Lastly, the pedestrian hot zones are
15 generally located in urban and suburban areas, where many residential areas are located.



Figure 2: Statewide screening results for each vehicle type

1
2

7. SUMMARY AND CONCLUSION

In recent years, there is growing recognition that incorporating macroscopic traffic safety analysis into the long-term transportation planning process is beneficial. The macroscopic safety analysis relates traffic crashes aggregated at a spatial level (traffic analysis zone (TAZ), census tract or county) with demographic, socioeconomic, and zonal level traffic/roadway characteristics. The macroscopic analysis can provide a broad spectrum perspective and long-term policy suggestions. The macro-level safety studies have been conducted by total crashes, crashes by severity level (such as no injury, minor injury, severe injury, and fatal injury) and crashes by vehicle type (such as motor vehicle, pedestrian, and bicycle). The range of approaches employed in literature includes univariate count frequency approaches for single dependent variable and multivariate count frequency models for multiple dependent variables. It has been recognized that in the presence of multiple dependent variables, multivariate approaches are appropriate.

In this study, we propose an alternative approach to examining multiple crash dependent variables. Specifically, a mixed multinomial logit fractional split model that examines the proportion of crash by vehicle type is developed. In this model, each alternative proportion is associated with a propensity. The proportion allocated to an alternative is probabilistically determined based on the alternative propensity as well propensity of all other alternatives. Thus, in this approach, exogenous variable effects directly affect all alternatives. In count frequency models, such interactions are absent. Thus, the proposed approach allows an alternate mechanism to examine multiple crash dependent variables. Further, the approach is well suited to accommodate a large number of alternatives without a sizable increase in computational burden. On the contrary, in a count modeling approach developing multivariate models for large number of dependent variables is computationally and methodologically challenging. The advantage is illustrated in our study by considering eight vehicle types in our analysis – passenger car, van, light trucks, medium and heavy trucks, bus, motorcycle, bicycle and pedestrian. The modeling approach allows us to identify and quantify the factors affecting crash proportions at the macro-level.

The proposed mixed multinomial logit fractional split model was estimated using socio-demographic, traffic, land-use, and commuting data at a Traffic Analysis Zone level using data compiled from multiple sources in Florida. The modeling results clearly highlight the applicability of the proposed approach for crashes involving different vehicle type analysis. The modeling results revealed that the impact of explanatory variables varies significantly across different vehicle types. For example, medium and heavy truck crashes are influenced by both socio-demographic (such as population density and industrial employment) and traffic characteristics (such as vehicle-miles-traveled density and proportion of trucks). On the other hand, non-motorized vehicle types (i.e. bicycle and pedestrian) had only one variable related to urban location or commuting characteristics. Subsequently, to quantify the impact of variables across the alternatives, elasticity effects were computed and presented.

To illustrate the applicability of the proposed framework for screening purposes, we also proposed the Excess Predicted Proportion (EPP) measure that computes the difference between the observed and predicted proportion of each vehicle type involved crashes in a zone. Hence, EPP, analogous to the Highway Safety Manual (HSM) Excess Predicted Average Crash Frequency, allows us to identify unsafe or hot zones. Based on this measure, a statewide screening exercise by the various vehicle types considered in our analysis was undertaken. The screening results revealed that the spatial pattern of hot zones is substantially

1 different across the various vehicle type crashes. The screening exercise clearly illustrates the value of the
2 proposed approach.

3
4 Overall, the paper demonstrated the application of a mixed multinomial fractional split model for crash
5 proportion modeling as statewide screening purposes (using EPP). The findings from our study are useful
6 for policy makers and practitioners to understand potential safety risks by various transportation vehicle
7 types and provide appropriate and effective policy-based countermeasures by vehicle types. The proposed
8 approach is quite flexible and can be adopted to examine crash proportions by other crash attributes, such
9 as collision type, severity level and temporal classification.

10
11 To be sure, the study is not without limitations. The consideration of crash proportions (as opposed to crash
12 counts) might result in incorrect identification of problematic zones under rare circumstances (with very
13 low number of crashes). To address this limitation, future studies can consider coupling a crash count model
14 (for total crashes) with a fractional split model to ensure that proportion and frequency are simultaneously
15 considered. Finally, due to the compensatory nature of the model structure an increase (or decrease) in crash
16 proportion for an alternative is associated with decrease (or increase) in crash proportions from other
17 alternatives. This needs to be recognized while developing and employing fractional split models for crash
18 proportions.

19 20 **ACKNOWLEDGMENT**

21 The authors wish to thank the Florida Department of Transportation for providing data and funding for this
22 study.

23 24 **REFERENCES**

- 25 AASHTO. (2010). *Highway Safety Manual*. Washington, DC: American Association of State Highway
26 and Transportation Officials.
- 27 Abdel-Aty, M., Lee, J., Siddiqui, C., & Choi, K. (2013). Geographical unit based analysis in the context
28 of transportation safety planning. *Transportation Research Part A: Policy and Practice*, 49, 62-
29 75.
- 30 Blatt, J., & Furman, S. M. (1998). Residence location of drivers involved in fatal crashes. *Accident*
31 *Analysis & Prevention*, 30(6), 705-711.
- 32 Cai, Q., Lee, J., Eluru, N., & Abdel-Aty, M. (2016). Macro-level pedestrian and bicycle crash analysis:
33 incorporating spatial spillover effects in dual state count models. *Accident Analysis & Prevention*,
34 93, 14-22.
- 35 Clark, D. E. (2003). Effect of population density on mortality after motor vehicle collisions. *Accident*
36 *Analysis & Prevention*, 35(6), 965-971.
- 37 Dong, N., Huang, H., Lee, J., Gao, M., & Abdel-Aty, M. (2016). Macroscopic hotspots identification: a
38 Bayesian spatio-temporal interaction approach. *Accident Analysis & Prevention*, 92, 256-264.
- 39 Dong, N., Huang, H., Xu, P., Ding, Z., & Wang, D. (2014). Evaluating spatial-proximity structures in
40 crash prediction models at the level of traffic analysis zones. *Transportation Research Record:*
41 *Journal of the Transportation Research Board*(2432), 46-52.
- 42 Dong, N., Huang, H., & Zheng, L. (2015). Support vector machine in crash prediction at the level of
43 traffic analysis zones: assessing the spatial proximity effects. *Accident Analysis & Prevention*, 82,
44 192-198.
- 45 Eluru, N., & Bhat, C. R. (2007). A joint econometric analysis of seat belt use and crash-related injury
46 severity. *Accident Analysis & Prevention*, 39(5), 1037-1049.

- 1 Eluru, N., Chakour, V., Chamberlain, M., & Miranda-Moreno, L. F. (2013). Modeling vehicle operating
2 speed on urban roads in Montreal: A panel mixed ordered probit fractional split model. *Accident*
3 *Analysis & Prevention*, 59, 125-134.
- 4 Golob, T. F., & Regan, A. C. (2003). Truck-involved crashes and traffic levels on urban freeways.
5 *University of California Transportation Center*.
- 6 Hadayeghi, A., Shalaby, A. S., Persaud, B. N., & Cheung, C. (2006). Temporal transferability and
7 updating of zonal level accident prediction models. *Accident Analysis & Prevention*, 38(3), 579-
8 589.
- 9 Huang, H., Abdel-Aty, M., & Darwiche, A. (2010). County-level crash risk analysis in Florida: Bayesian
10 spatial modeling. *Transportation Research Record: Journal of the Transportation Research*
11 *Board*(2148), 27-37.
- 12 Huang, H., Song, B., Xu, P., Zeng, Q., Lee, J., & Abdel-Aty, M. (2016). Macro and micro models for
13 zonal crash prediction with application in hot zones identification. *Journal of Transport*
14 *Geography*, 54, 248-256.
- 15 Kim, K., Brunner, I., & Yamashita, E. (2006). Influence of land use, population, employment, and
16 economic activity on accidents. *Transportation Research Record: Journal of the Transportation*
17 *Research Board*(1953), 56-64.
- 18 Ladron de Guevara, F., Washington, S., & Oh, J. (2004). Forecasting crashes at the planning level:
19 simultaneous negative binomial crash model applied in Tucson, Arizona. *Transportation*
20 *Research Record: Journal of the Transportation Research Board*(1897), 191-199.
- 21 Lee, J. (2014). *Development of Traffic Safety Zones and Integrating Macroscopic and Microscopic Safety*
22 *Data Analytics for Novel Hot Zone Identification*. University of Central Florida, Orlando, Florida.
- 23 Lee, J., Abdel-Aty, M., & Choi, K. (2014a). Analysis of residence characteristics of at-fault drivers in
24 traffic crashes. *Safety science*, 68, 6-13.
- 25 Lee, J., Abdel-Aty, M., Choi, K., & Huang, H. (2015a). Multi-level hot zone identification for pedestrian
26 safety. *Accident Analysis & Prevention*, 76, 64-73.
- 27 Lee, J., Abdel-Aty, M., Choi, K., & Siddiqui, C. (2013). *Analysis of residence characteristics of drivers,*
28 *pedestrians, and bicyclists involved in traffic crashes*. Paper presented at the Transportation
29 Research Board 92nd Annual Meeting.
- 30 Lee, J., Abdel-Aty, M., & Jiang, X. (2014b). Development of zone system for macro-level traffic safety
31 analysis. *Journal of transport geography*, 38, 13-21.
- 32 Lee, J., Abdel-Aty, M., & Jiang, X. (2015b). Multivariate crash modeling for motor vehicle and non-
33 motorized modes at the macroscopic level. *Accident Analysis & Prevention*, 78, 146-154.
- 34 Lee, J., Abdel-Aty, M., & Cai, Q. (2017). Intersection crash prediction modeling with macro-level data
35 from various geographic units. *Accident Analysis & Prevention*, 102, 213-226.
- 36 Levine, N., Kim, K. E., & Nitz, L. H. (1995). Spatial analysis of Honolulu motor vehicle crashes: II.
37 Zonal generators. *Accident Analysis & Prevention*, 27(5), 675-685.
- 38 Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: a review and
39 assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*,
40 44(5), 291-305.
- 41 Loukaitou-Sideris, A., Liggett, R., & Sung, H.-G. (2007). Death on the crosswalk A study of pedestrian-
42 automobile collisions in los angeles. *Journal of Planning Education and Research*, 26(3), 338-
43 351.
- 44 Lovegrove, G., & Sayed, T. (2007). Macrolevel collision prediction models to enhance traditional
45 reactive road safety improvement programs. *Transportation Research Record: Journal of the*
46 *Transportation Research Board*(2019), 65-73.
- 47 Milton, J. C., Shankar, V. N., & Mannering, F. L. (2008). Highway accident severities and the mixed logit
48 model: an exploratory empirical analysis. *Accident Analysis & Prevention*, 40(1), 260-266.
- 49 Nam, S. (2012). Multiple Fractional Response Variables with Continuous Endogenous Explanatory
50 Variables.

- 1 Narayanamoorthy, S., Paleti, R., & Bhat, C. R. (2013). On accommodating spatial dependence in bicycle
2 and pedestrian injury counts by severity level. *Transportation research part B: methodological*,
3 55, 245-264.
- 4 Nashad, T., Yasmin, S., Eluru, N., Lee, J., & Abdel-Aty, M. A. (2016). *Joint Modeling of Pedestrian and*
5 *Bicyclist Crashes: A Copula Based Approach*. Paper presented at the Transportation Research
6 Board 95th Annual Meeting.
- 7 Ng, K.-s., Hung, W.-t., & Wong, W.-g. (2002). An algorithm for assessing the risk of traffic accident.
8 *Journal of safety research*, 33(3), 387-410.
- 9 NHTSA. (2016). Traffic Safety Facts 2014 (Vol. DOT HS 812 263).
- 10 Noland, R., & Quddus, M. (2004). Analysis of pedestrian and bicycle casualties with regional panel data.
11 *Transportation Research Record: Journal of the Transportation Research Board*(1897), 28-33.
- 12 Papke, L. E., & Wooldridge, J. (1993). Econometric methods for fractional response variables with an
13 application to 401 (k) plan participation rates: National Bureau of Economic Research
14 Cambridge, Mass., USA.
- 15 Pasupuleti, N., & Pulugurtha, S. S. (2013). Spatial extent and modeling intracity truck crashes. *Procedia-*
16 *Social and Behavioral Sciences*, 104, 1188-1197.
- 17 Quddus, M. A. (2008). Modelling area-wide count outcomes with spatial correlation and heterogeneity:
18 an analysis of London crash data. *Accident Analysis & Prevention*, 40(4), 1486-1497.
- 19 Siddiqui, C., Abdel-Aty, M., & Choi, K. (2012). Macroscopic spatial analysis of pedestrian and bicycle
20 crashes. *Accident Analysis & Prevention*, 45, 382-391.
- 21 Sivakumar, A., & Bhat, C. (2002). Fractional split-distribution model for statewide commodity-flow
22 analysis. *Transportation Research Record: Journal of the Transportation Research Board*(1790),
23 80-88.
- 24 Song, J. J., Ghosh, M., Miaou, S., & Mallick, B. (2006). Bayesian multivariate spatial models for
25 roadway traffic crash mapping. *Journal of multivariate analysis*, 97(1), 246-273.
- 26 Sperazza, L. J., Dauenhauer, J., & Banerjee, P. (2012). Tomorrow's Seniors: Technology And Leisure
27 Programming. *The Journal of Community Informatics*, 8(1).
- 28 Stamatiadis, N., & Puccini, G. (2000). Socioeconomic descriptors of fatal crash rates in the Southeast
29 USA. *Injury control and safety promotion*, 7(3), 165-173.
- 30 Wang, Z., & Wolman, A. L. (2014). Payment Choice and the Future of Currency: Insights from Two
31 Billion Retail Transactions *Working Paper Series: The Federal Reserve Bank of Richmond*.
- 32 Wier, M., Weintraub, J., Humphreys, E. H., Seto, E., & Bhatia, R. (2009). An area-level model of
33 vehicle-pedestrian injury collisions with implications for land use and transportation planning.
34 *Accident Analysis & Prevention*, 41(1), 137-145.
- 35 Witter, L. A., Johnson, C. J., Croft, B., Gunn, A., & Gillingham, M. P. (2012). Behavioural trade-offs in
36 response to external stimuli: time allocation of an Arctic ungulate during varying intensities of
37 harassment by parasitic flies. *Journal of Animal Ecology*, 81(1), 284-295.
- 38 Xu, P., Huang, H., Dong, N., & Abdel-Aty, M. (2014). Sensitivity analysis in the context of regional
39 safety modeling: Identifying and assessing the modifiable areal unit problem. *Accident Analysis*
40 *& Prevention*, 70, 110-120.
- 41 Xu, P., Huang, H., Dong, N., & Wong, S. (2017). Revisiting crash spatial heterogeneity: A Bayesian
42 spatially varying coefficients approach. *Accident Analysis & Prevention*, 98, 330-337.
- 43 Yasmin. S., and N. Eluru (2016), "Latent Segmentation Based Count Models: Analysis of Bicycle Safety
44 in Montreal and Toronto," *Accident Analysis & Prevention*, Volume 95, Part A, October 2016,
45 Pages 157-171
- 46 Yasmin. S., N. Eluru, J. Lee and M. Abdel-Aty (2016), "Ordered Fractional Split Approach for Aggregate
47 Injury Severity Modeling," *Transportation Research Record* Vol. 2583, pp. 119-126