Incorporating the Influence of Vehicle Mix on Crash Frequency and Severity

Shahrior Pervaz*
Post-doctoral Scholar
Department of Civil, Environmental and Construction Engineering, University of Central Florida
Tel: 1-407-561-0298; Email: shahrior.pervaz@ucf.edu

Manmohan Joshi
Graduate Research Assistant
Department of Civil and Environmental Engineering, University of Connecticut
Tel: 860-465-6834; Email: manmohan.joshi@uconn.edu

Tanmoy Bhowmik
Assistant Professor
Department of Civil and Environmental Engineering, Portland State University
Tel: 1-407-927-6574; Email: tbhowmik@pdx.edu

Dewan Ashraful Parvez
Graduate Research Assistant
Department of Civil, Environmental and Construction Engineering, University of Central Florida
Tel: 407-437-2587; Email: daparvez@ucf.edu

Kai Wang
Assistant Research Professor
Connecticut Transportation Institute, University of Connecticut
Tel: 860-486-1587; Email: kai.wang@uconn.edu

John N. Ivan
Professor
Department of Civil and Environmental Engineering, University of Connecticut
Tel: 860-486-0352; Email: john.ivan@uconn.edu

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

Keywords: Crash frequency, Crash severity, Vehicle mix variables, Pooled model, Unobserved effects.

*Corresponding author

ABSTRACT

The current approaches for crash frequency and severity prediction in the Highway Safety Manual (HSM) do not employ vehicle mix information. In this research effort, we build advanced alternatives to HSM methods while incorporating vehicle mix information. Two model systems: (a) multivariate Poisson-lognormal model (MVPLN) and (b) negative binomial – ordered probit fractional split model (NB-OPFS) are estimated by incorporating vehicle mix variables. The developed model systems can also capture the influence of observed and unobserved heterogeneity of different independent variables including vehicle mix variables. We estimate the models for three facility types including Urban Arterial 4-Lane Divided segments, Rural 3-Leg STOP Controlled and Rural 4-Leg STOP Controlled intersections using data from four Highway Safety Information System (HSIS) states including California, Illinois, Minnesota, Washington, and three Non-HSIS states including Connecticut, Florida and Texas. For modeling crashes at each facility level, we adopt a pooled modeling technique that accounts for state specific observed and unobserved heterogeneity in the pooled datasets. A comprehensive set of independent variables including traffic volume, vehicle mix indicators, roadway characteristics and state specific indicators are considered in the analysis. The model comparison exercise is conducted based on a comprehensive set of quantitative and qualitative metrics. The study highlights how different methodological approaches perform better for different facilities. The study findings also underscore how capturing the observed and unobserved impacts of vehicle mix variables improves model performance in crash frequency and severity dimensions across the facility types.

Keywords: Crash frequency, Crash severity, Vehicle mix variables, Pooled model, Unobserved effects.

1. BACKGROUND

Transportation safety literature employs statistical or econometric models to examine crash occurrences and their consequences at various spatial levels such as site level, corridor level and area level. The site level and corridor level analysis are conducted to identify geometric design specific and/or engineering solutions to reduce the impact of crashes for the examined road entities (segment, intersection or network) while the area level (state, zone or block) studies facilitate the identification of regional hotspots, and adoption of area-wide planning and remedial solutions. The different types of crash models employed include univariate count models (where a single count variable such as total crashes is examined for a spatial unit (*1–3*)), simulation based multivariate and/or unobserved heterogeneity incorporated count models (where multiple crash count variables by crash type and/or severity are analyzed for a spatial unit in multivariate models including means and variances approaches (*4–13*), latent class models to incorporate class-specific heterogeneity (*14*, *15*)), analytically closed form based count models (where multivariate distributions or approximations of multivariate distributions with an analytical closed form probability expression are employed (*16*, *17*)), count-fractional split models (where the count component models total crashes and the fractional split component models fraction of crashes by severity/crash type (*18–20*)), and integrated multi-resolution crash frequency models (where crash data from multiple observational resolutions are considered simultaneously within a unified system (*21–24*)).

The findings from these research studies traditionally form the basis for safety planning and guidance provided by transportation agencies across the country. The American Association of State Highway and Transportation Officials (AASHTO) released the first edition of the Highway Safety Manual (HSM) in 2010 that provides a uniform guidance documenting methods and procedures for estimating total crashes, crashes by type and crashes by severity at the site level, project level and corridor level (*25*). While the HSM approaches are widely employed in transportation agencies, researchers are continuing to develop enhanced approaches that are practical and reliable for application across transportation jurisdictions in the country. Several research studies identified vehicle mix information as a relevant variable for inclusion in applied crash frequency and severity models (*19*, *20*, *22*, *26*, *27*). Vehicle mix, in this context, is defined as traffic volume (AADT) by vehicle type. The vehicle type information can be considered at a coarser resolution such as passenger car and truck AADT (or percentage). A finer resolution vehicle mix variable can include detailed information such as types of buses, trucks, utility vehicles, SUV and other vehicle classes (see (*1*, *2*, *28–32*) for studies employing this resolution for modeling).

In the NCHRP project titled "The Effect of Vehicle Mix on Crash Frequency and Crash Severity", we developed a practical approach to systematically incorporate the impact of vehicle mix on crash occurrence and severity (*33*). In this project, we considered the impact of different vehicle mix variables (coarse and fine resolution) on crash frequency and severity analysis. While the negative binomial model system is the most commonly incorporated framework in HSM, several competing frameworks have emerged in recent years. Eluru et al. (2024) tested two emerging methods: (a) multivariate Poisson-lognormal model and (b) negative binomial – ordered probit fractional split model (*33*). The model estimation procedures were implemented for a large number of facilities using data from multiple states and a user guidebook was developed. The current study builds on the NCHRP project effort along the following ways. *First*, the methodological frameworks developed in the NCHRP project were limited by practical considerations. Hence, the model building process was limited to a smaller set of variables with few interactions i.e., limited observed heterogeneity. Further, the models estimated did not account for random parameters and/or common unobserved factors affecting the dependent variables. Thus, in our current research effort, we developed advanced variants of the modeling frameworks that account for additional observed and unobserved heterogeneity while accounting for the impact of vehicle mix. *Second*, the study builds on the pooled modeling approach employed in NCHRP project by incorporating additional interactions of jurisdiction-specific variables with other independent variables. For example, we examine how the impact of independent variables such as AADT vary by jurisdiction. The approach allows for custom development of jurisdiction specific models without the disadvantages of partitioning data by jurisdiction. Thus, the proposed approach accommodates state specific observed and unobserved heterogeneity. *Finally*, we recognize that a single model structure cannot outperform all alternatives for all facility types. Hence, in the current study we employ a detailed model comparison exercise based on a comprehensive set of quantitative and qualitative metrics to identify the most appropriate model system for each facility type. We compare the two novel frameworks with the current state of the art models employed in practice through the HSM model.

For our analysis, we consider data from four Highway Safety Information System (HSIS) states including California, Illinois, Minnesota, Washington and three Non-HSIS states including Connecticut, Florida, and Texas. Finally, the guidance exercise is undertaken for different facility types to illustrate how there is no universal model system that offers enhanced fit across different facility types. In our analysis, we developed models for three different facility types based on HSM facility guidelines (see (*33*)). In this paper, we select Urban Arterial 4-Lane Divided segment (UA4LD) facility, Rural 3-Leg STOP Controlled (R3ST) and Rural 4-Leg STOP Controlled (R4ST) intersection facilities for our comparison exercise. We wanted to select different facility types to examine if and how the impact of vehicle mix varies by location (urban and rural) and facility type (segment and intersection). Further, we considered three facility types to highlight how a single framework does not necessarily offer improvement for all facility types. The comparison exercise allows us to see how different model systems might offer enhanced performance across facility types.

2. METHODOLOGY

In this study, we consider two advanced frameworks: a) negative binomial-fractional split framework (NB-OPFS), and b) multivariate Poisson-lognormal (MVPLN) model. The equation systems for NB-OPFS and MVPLN models are discussed in the following sections.

2.1 Negative Binomial-Ordered Probit Fractional Split (NB-OPFS) Model

In the NB-OPFS framework, NB component models the total crashes and the OPFS component estimates the fraction of crashes by severity levels.

2.1.1 Count component (NB model)

For a spatial unit l (where l is segment s or intersection i), negative binomial (NB) model can be employed to estimate total crash count. The probability density function of NB model can be written as,

$$P(c_l) = \frac{\Gamma\left(c_l + \frac{1}{\alpha_l}\right)}{\Gamma(c_l + 1)\Gamma\left(\frac{1}{\alpha_l}\right)} \left(\frac{1}{1 + \alpha_l \mu_l}\right)^{\frac{1}{\alpha_l}} \left(1 - \frac{1}{1 + \alpha_l \mu_l}\right)^{c_l} \tag{1}$$

where, c_l be the index for crashes occurring over a period of time in a spatial unit l (segment or intersection). $P(c_l)$ is the probability that unit l has c_l number of crashes. $\Gamma(\cdot)$ is the gamma function, α_l is negative binomial overdispersion parameter and μ_l is the expected number of crashes occurring in the unit l over a given time period. The equation for μ_l can be written as follows,

$$\mu_l = E(c_l|\boldsymbol{Y}_l) = exp((\boldsymbol{\delta}_l + \boldsymbol{\zeta}_l)Y_l + \varepsilon_l + \eta_l) \tag{2}$$

where, $\boldsymbol{Y}_l$ is a vector of explanatory variables associated with the analysis unit l. $\boldsymbol{\delta}_l$ is a vector of coefficients to be estimated. $\boldsymbol{\zeta}_l$ a vector of unobserved factors on crash count propensity for unit l. ε_l is a gamma distributed error term with mean 1 and variance α_l. η_l captures the influence of common unobserved factors that impact the total number of crashes and proportion of crashes by severity for unit l.

2.1.2 Fractional split component (OPFS model)

The modeling of crash proportions by severity levels is undertaken using the ordered probit fractional split model (OPFS). In the ordered outcome framework, the actual injury severity proportions (y_{lk}) are assumed to be associated with an underlying continuous latent variable (y_l^*) as follows:

$$y_l^* = ((\boldsymbol{\beta_l} + \boldsymbol{\rho_l})\boldsymbol{F}_l + \xi_l \pm \eta_l),\ y_{lk} = k \ if \ \tau_{l(k-1)} < y_l^* < \tau_{lk} \tag{3}$$

The latent propensity y_l^* is mapped to the actual severity proportion categories y_{lk} by τ_l thresholds ($\tau_{l0} = -\infty \ and \ \tau_{lK} = +\infty$). $\boldsymbol{F}_l$ is a vector of attributes (not including a constant) that influences the propensity associated with severity proportion categories for unit l. $\boldsymbol{\beta_l}$ is the corresponding vector of mean effects. $\boldsymbol{\rho_l}$ a vector of unobserved factors on severity proportion propensity for unit l. ξ_l is an idiosyncratic error term assumed to be identically and independently standard normally distributed across unit l. η_l term generates the correlation between equations for total number of crashes and crash proportions by severity levels and also allows for considering the influence of various unobserved factors affecting the frequency and proportion variables. The $\pm$ sign in front of η_l indicates that the correlation in unobserved individual factors between total crashes and crash proportions by severity levels may be positive or negative. A positive sign implies that facilities with higher number of crashes are intrinsically more likely to incur higher proportions for severe crashes. On the other hand, negative sign implies that facilities with higher number of crashes intrinsically incur lower proportions for severe crashes. To determine the appropriate sign one can empirically test the models with both $'+'$ and $'-'$ signs independently. The model structure that offers the superior data fit is considered as the final model.

It is important to note here that the unobserved heterogeneity between total number of crashes and crash proportions by severity levels can vary across facilities. Therefore, in the current study, the correlation parameter η_l is parameterized as a function of observed attributes as follows:

$$\eta_l = G_l\boldsymbol{Q}_l \tag{4}$$

where, Q_l is a vector of exogenous variables, $\boldsymbol{G}_l$ is a vector of unknown parameters to be estimated (including a constant).

To estimate the model presented in equation 3, we assume that:

$$E(y_{lk}|F_l) = H_{lk}(\beta_l, \tau_l), 0 \leq H_{lk} \leq 1, \sum_{k=1}^{K} H_{lk} = 1 \quad (5)$$

H_{lk} in our model takes the ordered probit probability (Λ) form for the severity category k.

Given these relationships across different parameters, the resulting probability (Λ) for the ordered probit fractional split model takes the following form:

$$\Lambda(y_{lk} = k) = \varphi\{\tau_{lk} - y_l^*\} - \varphi\{\tau_{l(k-1)} - y_l^*\} \quad (6)$$

where, $\varphi(\cdot)$ is the standard normal cumulative distribution function.

2.1.3 Model estimation

In examining the model structure of total crash count and proportions of crashes by severity level, it is necessary to specify the structure for the unobserved vectors $\boldsymbol{G}, \boldsymbol{\zeta}, \boldsymbol{\rho}$ represented by Ω. In this study, it is assumed that the elements are drawn from independent realization from normal population: $\Omega \sim N(0, (\boldsymbol{\sigma}_1^2, \boldsymbol{\sigma}_2^2, \boldsymbol{\sigma}_3^2))$. Thus, conditional on Ω, the likelihood function for the integrated probability can be expressed as:

$$L_l = \int_{\Omega} P(c_l) \times \prod_{k=1}^{K} \left(\Lambda(y_{lk} = k)\right)^{\varpi_l d_{lk}} d\Omega \quad (7)$$

where, w_l is a dummy variable taking a value of 1 if the corresponding unit l has at least one crash over the study period and 0 otherwise. d_{lk} is the proportion of crashes in severity category k for unit l. Finally, the log-likelihood function is:

$$LL = \sum_l Ln(L_l) \quad (8)$$

All the parameters in the model are estimated by maximizing the logarithmic function LL presented in equation 8. To estimate the proposed model, we apply Quasi-Monte Carlo simulation techniques based on the scrambled Halton sequence to approximate this integral in the likelihood function and maximize the logarithm of the resulting simulated likelihood function across individuals (please see (*34*, *35*) for details). We use the GAUSS matrix programming software to run the models (*36*).

2.2 Multivariate Poisson-Lognormal (MVPLN) Model

Multivariate Poisson-lognormal (MVPLN) model estimates the factors affecting crashes across severity levels. Let *n* be the number of observations in facility (segments or intersections), *J* be the number of severity levels, and ***Y*** be a matrix of crash counts, with Y_{ij} be the number of crashes at location *i* with severity *j*. The crash count of the j^{th} severity type at the i^{th} entity, y_{ij}, follows a Poisson distribution with parameter λ_{ij}, as shown in equations below (*37*).

$$Y_{ij}|\lambda_{ij} \sim Poisson(\lambda_{ij}) \quad (9)$$

$$\lambda_{ij} = \exp\left(X_{ij}'\beta_j + \varepsilon_{ij}\right) \quad (10)$$

In this model, X_{ij} is a k-dimensional matrix of covariates, and β_j is a vector of parameters. Notably, the parameter for some selected covariates X_t, for example $\beta_{j,t}$, are allowed to vary according to a multivariate normal distribution across all severity levels, while the other parameters remain constant.

$$\beta_{j,t} \sim N_J(\mu_t, \Omega) \tag{11}$$

Where μ_t is a mean vector of coefficients of covariate X_t across all severities and Ω is corresponding variance-covariance matrix. All other parameters $\beta_{j,k}$ (for $k \neq t$) are constant. The random effects ε_{ij} are assumed to follow a multivariate normal distribution as:

$$\varepsilon_i | \Sigma \sim N_J(0, \Sigma) \tag{12}$$

The unrestricted covariance matrix $\sum$ captures the correlation between severity levels that is modeled using a J-dimensional multivariate normal distribution N_J. A full Bayesian approach is adopted for estimation of parameters, and this involves solving multi-dimensional integrals without a closed form solution and hence Markov Chain Monte Carlo (MCMC) simulation approach is used to determine parameter estimates. The MCMC algorithm is implemented using Just Another Gibbs Sampler (JAGS) to estimate posterior distributions of parameters (*38*). For all the regression coefficients, noninformative normal prior with zero mean and very high variance is used; and the inverse of variance-covariance is estimated using Wishart distributed prior as:

$$\Omega^{-1} \sim Wishart(I, J),\ \Sigma^{-1} \sim Wishart(I, J);\ I = J\ dimensional\ Identity\ matrix \tag{13}$$

3. DATA PREPARATION

This study considers data from four Highway Safety Information System (HSIS) states including California, Illinois, Minnesota, Washington and three Non-HSIS states including Connecticut, Florida, and Texas for the analysis. The reader will note that we attempted to collect data from other HSIS and non-HSIS states such as Michigan, Maine, Utah, North Carolina, Ohio and South Dakota. However, data obtained from some states were outdated or had missing information for very important variables such as lane width, shoulder type, shoulder width, and median width. Hence, we did not include those states for our analysis. Based on the data availability by facility types, we considered data from California, Florida, Illinois, Minnesota, Texas, and Washington states for segment facilities and California, Connecticut, Florida, and Minnesota states for intersection facilities. This study considers Urban Arterial 4-Lane Divided (UA4LD) segment facility, Rural 3-Leg STOP Controlled (R3ST) and Rural 4-Leg STOP Controlled (R4ST) intersection facilities for the empirical analysis. For the analysis, we spatially assign the crashes for segment and intersection facilities by using ArcGIS tools. In this process, for assigning intersection-related crashes, a 250 feet buffer around the center of each intersection was considered and the crashes were spatially assigned (see earlier studies that adopted this approach (*21*, *39–41*)). The information of the facility types and crash statistics across the facility types are shown in Table 1.

TABLE 1 Information of the Facility Types

State	Crash Data (Year)	Urban 4-Lane Divided Arterial		Rural 3-Leg STOP Controlled		Rural 4-Leg STOP Controlled	
		Number of Sites	Total Crashes	Number of Sites	Total Crashes	Number of Sites	Total Crashes
HSIS							
California	2013 – 2017	1,549	17,488	6,197	977	1,850	850
Illinois	2013 – 2017	21,600	160,786	--[a]	--	--	--
Minnesota	2011 – 2015	3,281	9,959	1,699	5,900	1,953	9,525
Washington	2014 – 2018	1,374	3,753	--	--	--	--
Non-HSIS							
Connecticut	2015 - 2019	--[b]	--	198	446	46	200
Florida	2015 - 2019	1,533	258,375	746	5,192	185	1,908
Texas	2015 - 2019	6,342	25,822	--	--	--	--
Total		35,679	476,183	8,840	12,115	4,034	12,483
Estimation Samples		7,500	93,832	6,500	8,991	3,000	9,310
Validation Samples		20,000	279,778	2,340	3,524	1,034	3,173

Note: a) The intersection file is only available for 2 HSIS states including California and Minnesota and 2 non HSIS states including Connecticut and Florida. b) For Connecticut state, the crash counts for the selected segment facility were very low in addition to the missing information of some important variables. Hence, we excluded Connecticut state from segment facility.

For modeling crashes at each facility level by using data from all the analysis states, we adopt a pooled modeling technique. In this technique, for each facility, we gather the datasets from all the analysis states and prepare a single dataset for model estimation process. This single pooled dataset is then split into estimation dataset (used for the model development) and validation dataset (used for the model performance assessment) by randomly sampling the data. For instance, for UA4LD segment facility, the data from all six states resulted in a pooled dataset of 35,679 segments. From these segments, 7,500 segments were randomly drawn for model estimation while drawing 20,000 different segments for model validation. A similar procedure was followed for the two intersection facilities.

3.1 Variables Considered

In this study, a five-point severity scale KABCO is considered for the crash analysis by severity type. KABCO is a widely used injury severity scale where K = fatal crashes (crashes which result in at least a death within 30 days of crashes), A = incapacitating crashes (non-fatal crashes which result in disabling injuries, such as broken bones, severed limbs, skull/chest/abdominal injuries, etc. and usually require hospitalization and transport to medical facility), B = non-incapacitating crashes (non-fatal crashes which result in non-disabling but evident injuries, such as lacerations, scrapes, bruises, etc.), C = possible injury crashes (non-fatal crashes which result in no visible signs of injury but complaint of pain, momentary unconsciousness, nausea or hysteria), and O = no injury crashes (*42–44*). For NB-OPFS model framework, total crash counts and crash proportion by each severity class are considered as dependent variables while for multivariate Poisson-lognormal modeling approach, crash counts by each severity level are considered. The severity proportion in the NB-OPFS model for a specific severity class is defined by crash counts by that severity class divided by total number of crashes (total of all severity classes). The severity proportions are: 1) proportion of fatal crashes (KP), 2) proportion of incapacitating crashes (AP), 3) proportion of non-incapacitating crashes (BP), 4) proportion of possible injury crashes (CP), and 5) proportion of no injury crashes (OP).

In terms of independent variables, a comprehensive set of variables including traffic volume information (aggregate level AADT), vehicle mix indicators (such as truck percentage, single unit truck percentage, and high truck zone indicator (a detailed definition of this variable is provided in Section 3.1.1 Approaches and assumptions of vehicle mix data)), roadway characteristics (such as lane width, median width, speed limit, shoulder type and width) and state specific indicators (variables reflecting the state specific differences caused by the driver behavior, roadway design and operations) are considered for the crash frequency analysis. The reader would note that we observed varying speed limit distributions across facility types in different states (see Table 2). The variation in speed limit distributions could be attributed to several factors such as state-specific geography, land-use, roadway environment, traffic intensity, and state-specific regulations.

TABLE 2 Distribution of Speed Limits Across the States

State	UA4LD			R3ST			R4ST		
	SL<=40 mph	SL 41-55 mph	SL>55 mph	SL<=40 mph	SL 41-55 mph	SL>55 mph	SL<=40 mph	SL 41-55 mph	SL>55 mph
CA	15.19	33.23	51.58	15.20	39.43	45.37	9.62	20.19	70.19
CT	--	--	--	55.48	44.52	0.00	54.76	45.24	0.00
FL	49.53	40.19	10.28	23.64	59.09	17.27	26.52	62.88	10.61
IL	51.52	35.36	13.12	--	--	--	--	--	--
MN	0.00	100.00	0.00	3.98	90.46	5.56	2.29	92.17	5.54
TX	14.35	22.89	62.76	--	--	--	--	--	--
WA	0.00	100.00	0.00	--	--	--	--	--	--

Note: The values in the table indicate percentages; -- indicates that the state is excluded from the analysis for that facility type due to data unavailability.

3.1.1 Approaches and assumptions of vehicle mix data

We explored the vehicle mix data availability across our study states. We used the observed vehicle mix data in the model estimation process for the states where data are available. Alternatively, if vehicle mix data was not available, we adopted the Quasi-induced exposure (QIE) technique for generating the vehicle mix data across each facility type within the state, and then used the generated vehicle mix data in crash frequency and severity models for the corresponding facility (please see (*45*, *46*) for a detailed discussion on the QIE approach). By exploring the vehicle mix data availability and resolution of the vehicle classification (coarse and fine) across the seven states, we found that the data for passenger cars and trucks are available for the majority of the states, at least at the coarser resolution. At a finer resolution, we did not obtain any data on passenger cars. However, categorization of trucks is available for five states including California, Illinois, Minnesota, Washington, and Texas. Among these states, 4 states have available data for single unit trucks. Other finer resolutions such as combination unit or multi-unit are available for 1 or 2 states only. Therefore, in the current study, we used total truck percentage and single unit truck percentage as the vehicle mix information variables. To examine the additional impacts of truck traffic, we tested the impact of trucks in locations with high truck volume, referred to as high truck zone. These locations are defined as having truck percentage ≥ 85th percentile of truck traffic percentage for the corresponding facility type. We considered 85th percentile value as it is a commonly used metric in transportation engineering, such as 85th percentile speed. In addition to the direct vehicle mix variables, we incorporated several interaction variables between high truck zone and other geometric attributes to capture the non-linear effect of truck percentage on crash frequency and severity.

In estimating the model, several functional forms, and combination of variables are considered and those that provide the best fit are retained in the final specification. The final specification of the models is based on removing the statistically insignificant variables in a systematic process based on 90% confidence level. The summary statistics of the variables considered for the final model estimation across the facilities are presented in Table A4 to Table A6 in the Appendix.

4. EMPIRICAL ANALYSIS

4.1 Model Specification and Overall Measure of Fit

In this study, a negative binomial-ordered probit fractional split (NB-OPFS) model framework and a multivariate Poisson-lognormal (MVPLN) model framework are employed to estimate crash frequency for Urban Arterial 4-Lane Divided (UA4LD) segment facility, Rural 3-Leg STOP Controlled, and Rural 4-Leg STOP Controlled intersection facilities while incorporating vehicle mix information. We also estimate HSM predictive methods for three facilities by following the equations described in Chapter 11 and Chapter 12 of part C of HSM 2010 (*25*). The HSM methods provide the benchmark for newly developed model systems. Since three model systems are different, instead of depending on log-likelihood and Bayesian Information Criterion (BIC) value, we evaluate model predictive performance by employing two statistical measures of fit: mean absolute deviation (MAD) and mean squared prediction error (MSPE) (please see (*23*, *27*) for a detailed definition of these measures). The model with the lower values of MAD and MSPE provides better predictions for the observed data. The MAD and MSPE values for all severity levels by facility for the three models are presented in Table 3.

Table 3 presents the MAD and MSPE values for total crash frequency and crash frequency by five severity classes. For each facility type, we obtained the values from both estimation and validation datasets for the three model frameworks. The reader would note that a single framework might not perform consistently better across all the dimensions for a facility type. Therefore, identifying the best model for each facility type is challenging. To this end, we adopt two approaches that consider the model performances across all the estimated dimensions. The <u>first approach</u>, defined as total crash approach considers MAD and MSPE values from total crash frequency predictions (sum of crash frequency across all severity levels). Alternatively, the <u>second approach,</u> defined as severity level scoring approach considers MAD and MSPE values from five severity levels. The following sections discuss both approaches.

4.1.1 Total crash approach

In this approach, we identify the model that provides the lowest MAD and MSPE with respect to total crash frequency. The values of the MAD and MSPE measures are presented in the 10th column and the selected models based on this approach are presented in 12th column in Table 3.

4.1.2 Severity level scoring approach

In this approach, the model that performs better for the severity level is awarded a point and the total score for each model across the severity levels is aggregated.

TABLE 3 Predictive Performance and Model Selection Process

Facility	Dataset	Measures	Models	O	C	B	A	K	Total	Severity Score	Total Crash App.	Severity Scoring App.	Final Model
UA4LD	Estimation	MAD	HSM	9.977	1.507	1.087	0.898	0.075	12.581	1	NB-OPFS	MVPLN	NB-OPFS
			NB-OPFS	8.909	1.450	1.080	1.057	0.086	11.479	3			
			MVPLN	9.976	1.346	1.006	0.788	0.065	12.662	4			
		MSPE	HSM	1,335.919	33.522	9.507	4.340	0.078	2,002.542	1	NB-OPFS	NB-OPFS	
			NB-OPFS	1,127.443	31.901	8.041	7.023	0.077	1,680.627	3			
			MVPLN	2,315.543	56.700	10.661	3.146	0.066	3,566.045	2			
	Validation	MAD	HSM	10.743	1.705	1.213	0.891	0.078	13.639	0	NB-OPFS	MVPLN	
			NB-OPFS	9.764	1.651	1.205	1.086	0.090	12.669	2			
			MVPLN	11.063	1.528	1.006	0.788	0.062	14.671	4			
		MSPE	HSM	2,463.751	56.808	17.231	4.748	0.102	3,574.123	2	NB-OPFS	MVPLN	
			NB-OPFS	2,352.022	63.135	16.816	3.010	0.100	3,427.745	2			
			MVPLN	2,963.729	78.095	10.661	3.146	0.068	4,651.209	3			
R3ST	Estimation	MAD	HSM	1.094	0.372	0.279	0.111	0.039	1.677	1	NB-OPFS	MVPLN	NB-OPFS
			NB-OPFS	0.751	0.287	0.211	0.091	0.043	1.110	4			
			MVPLN	0.817	0.289	0.223	0.091	0.039	1.579	5			
		MSPE	HSM	10.338	0.921	0.512	0.161	0.027	23.462	1	NB-OPFS	NB-OPFS	
			NB-OPFS	6.969	0.688	0.377	0.135	0.027	15.110	5			
			MVPLN	7.860	0.680	0.451	0.126	0.026	22.873	3			
	Validation	MAD	HSM	1.175	0.402	0.294	0.110	0.044	1.799	1	NB-OPFS	NB-OPFS /MVPLN	
			NB-OPFS	0.809	0.308	0.234	0.097	0.047	1.214	4			
			MVPLN	0.891	0.314	0.248	0.097	0.042	1.334	4			
		MSPE	HSM	8.387	0.873	0.392	0.113	0.027	17.944	1	NB-OPFS	NB-OPFS	
			NB-OPFS	6.236	0.746	0.343	0.102	0.027	13.177	5			
			MVPLN	7.801	0.759	0.403	0.103	0.027	15.958	3			
R4ST	Estimation	MAD	HSM	2.039	0.939	0.555	0.177	0.126	3.339	1	MVPLN	MVPLN	MVPLN
			NB-OPFS	1.640	0.771	0.472	0.176	0.142	2.637	3			
			MVPLN	1.601	0.799	0.462	0.157	0.122	2.583	5			
		MSPE	HSM	328.781	60.992	3.149	0.332	0.091	780.322	3	NBOPFS	MVPLN	
			NB-OPFS	323.821	59.991	2.816	0.309	0.092	765.033	4			
			MVPLN	327.824	60.333	2.785	0.304	0.079	773.041	5			
	Validation	MAD	HSM	1.828	0.832	0.574	0.229	0.109	3.038	1	MVPLN	MVPLN	
			NB-OPFS	1.473	0.682	0.480	0.198	0.136	2.327	4			
			MVPLN	1.427	0.710	0.469	0.183	0.115	2.274	5			
		MSPE	HSM	27.342	4.153	1.355	0.245	0.067	64.858	2	MVPLN	MVPLN	
			NB-OPFS	22.833	4.070	1.105	0.206	0.073	54.896	4			
			MVPLN	22.973	3.976	1.096	0.204	0.062	54.889	5			

Specifically, we identify the model with the lowest measures (MAD/MSPE) by severity level in the dataset and assigned a value of 1 while a value of 0 is given to the other models. In this process, two or all models are considered as similar performing models if the difference in the predictive measures are less than 10% relative to the top performing model. The final scores for each model are computed by adding the score across severity levels for each facility type (as shown in 11th column in Table 3). The model with the highest score at a facility type is considered as the top performing model for that facility type (as shown in 13th column).

4.1.3 Final model selection process

Based on the two approaches discussed above, the model that performs better across the measures and datasets is considered as the final model for the respective facility type. The final selected models across facility types are shown in the Final Model Column in Table 3. The results show that both NB-OPFS and MVPLN frameworks with vehicle mix data performed better than HSM predictive model that does not consider vehicle mix data. Within these advanced frameworks, the NB-OPFS models are selected for UA4LD segment and R3ST intersection facility while the MVPLN model is selected for R4ST intersection facility type.

4.2 Accommodating Unobserved Heterogeneity in the Final Selected Models

To capture the parameter variability across the sample, we estimate the random parameters in our selected model system. We compare the random parameter models with their independent counterparts (fixed parameter models) in terms of log-likelihood (LL) and BIC values (as shown in Figure 1) to see the improvement of the models. Figure 1 shows that models that capture random parameter effects perform better than fixed parameter models across facility types.

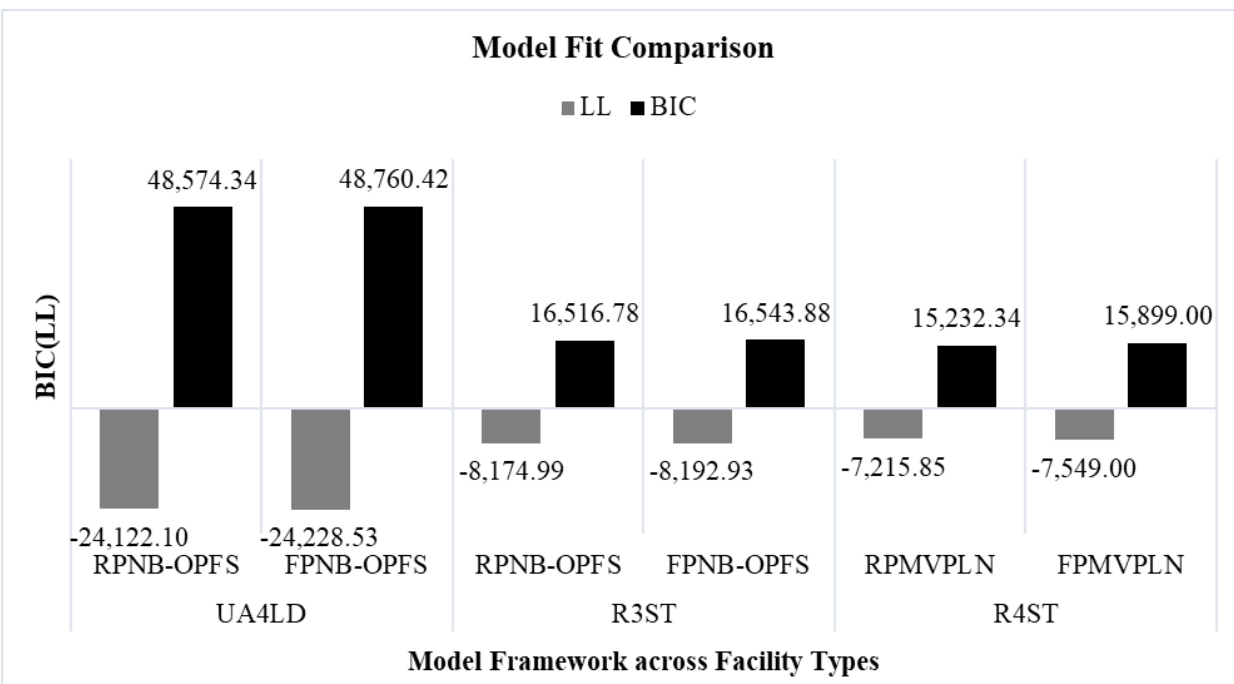

Figure 1 Comparison between random parameters and fixed parameters model frameworks

4.3 Model Estimation Results

This section provides a detailed discussion of the factors affecting crash count by severity levels across the facility types considered in the analysis. Since random parameters models are found to have improved data fit as evidenced by lower LL and BIC values in Figure 1, we discuss the results of random parameter NB-OPFS and random parameter MVPLN models. Tables 4, 5 and 6 represent the final selected models with random parameter effects. The reader would note that a positive (negative) sign for a variable in Tables indicates that an increase in the variable is likely to result in more (less) crashes as well as exhibits a higher (lower) impact on severity. We discuss the variables effects by facility types in the following sections. The results of the fixed parameter models are presented in Table A1 to Table A3 in the Appendix.

4.3.1 Urban arterial 4-lane divided (UA4LD) segment facility

In the crash count component of UA4LD segment facility, the model constant does not have any substantive interpretation. However, our model shows that the model constant value varies across the jurisdictions (states). For Minnesota and Washington states, the values are different from the other states as shown by state indicator variables in Table 4. For instance, for all the states other than Minnesota and Washington, the constant value is -1.767, and for Minnesota and Washington, the constants are -9.538 (-1.767-7.771) and -7.968 (-1.767-6.201) respectively. These differences highlight the region-specific influences on the estimates. It is to be noted that the segment length and number of years (5) are used as an offset variable in the NB model specification.

With respect to the traffic characteristics, several variables were found to be significant in our model. As expected, the parameter associated with AADT shows a positive impact on the likelihood of total crashes. AADT serves as a surrogate for exposure for traffic volume and therefore, with higher exposure, the likelihood of crash risk increases (see (*21*, *22*, *47*) for similar results). The interactions of AADT variables and state variables indicate that the effect of the AADT varies for Minnesota and Washington states. Further, the parameter associated with AADT also exhibits significant variation across segments as evidenced by the significant random parameter (indicated by the standard deviation variable in Table 4). Among the vehicle mix variables, a higher percentage of truck traffic and percentage of single unit truck traffic are found to decrease the crash risk in the UA4LD segments (see (*48*) for similar finding). The results possibly are the effects of cautious driving with the increased percentage of heavy vehicles on the UA4LD segments. Additionally, the effect of the percentage of single unit truck traffic on crash count propensity varies significantly across the segments as indicated by the standard deviation parameter in Table 4.

With regards to the roadway characteristics, the results show that segments with wider lane (>12 feet) and wider median width (> 20 feet) show a negative impact on crash count compared to the lane width ≤12 feet and median with ≤ 20 feet, respectively (as found in (*49*)). Interestingly, a wider lane width (>12 feet) in the high truck zone is found to further reduce the crash risk. Further, narrow outside and inside shoulder width (<8 feet) are found to increase crash risk in the UA4LD segment facility compared to the wider shoulder width (≥8 feet) (see (*49*) for similar results). Again, the results show that the narrow outside shoulder (<8 feet) in the high truck zone further increases the crash risk. The paved shoulder type is found to reduce crash risk in the segments and the effect gets moderated for California state as indicated by the interaction variable. The speed limit >55 mph variable also shows similar effect, perhaps reflecting the better roadway facilities and design conditions on sections with higher speed limits (*48*, *50*).

TABLE 4 Model Estimation Results of Random Parameters NB-OPFS Model for UA4LD Segments (N=7,500)

Variable Names	Count Component		Severity Proportion Component	
	Estimates	t-stat	Estimates	t-stat
Constant	-1.767	-5.424	--	--
State-Minnesota	-7.771	-6.146	--	--
State-Washington	-6.201	-3.379	--	--
Ln (Year = 5)	1.000	--	--	--
Threshold Parameters				
Threshold between OP-CP	--	--	0.708	4.648
Threshold between CP-BP	--	--	1.010	6.610
Threshold between BP-AP	--	--	1.455	9.511
Threshold between AP-KP	--	--	2.664	16.620
Traffic Characteristics				
Ln (AADT)	0.436	13.860	0.033	2.279
Standard Deviation	0.061	18.194	--	--
Ln (AADT)* State-Florida	--	--	-0.013	-3.797
Ln (AADT)* State-Minnesota	0.722	5.712	--	--
Ln (AADT)* State-Washington	0.510	2.832	--	--
%Truck	-0.026	-4.204	-0.007	-3.120
%Single unit truck	-0.043	-4.128	--	--
Standard Deviation	0.028	1.657	--	--
Roadway Characteristics				
Ln (Segment length, miles)	1.000	--	0.036	3.136
Lane width (base: ≤12 feet)				
LW>12	-0.171	-2.231	--	--
HTZ*LW>12	-0.982	-4.089	--	--
Median width (base: ≤20 feet)				
MW>20	-0.293	-5.187	-0.065	-2.252
Outside shoulder width (base: ≥8 feet)				
OSW<8	0.449	8.226	--	--
HTZ*OSW<8	0.433	4.453	--	--
Inside shoulder width (base: ≥8 feet)				
ISW<8	0.372	3.661	-0.133	-2.680
Shoulder type (base: unpaved)				
Paved	-0.716	-8.899	--	--
Paved* State-California	0.442	3.015	--	--
Speed limit (base: ≤55 mph)				
SL>55	-0.725	-5.484	0.173	3.723
SL>55* State-California	0.430	2.467	--	--
Overdispersion Parameter				
Constant	1.810	32.108	--	--
State-California	-0.950	-7.456	--	--
State-Washington	-1.679	-17.085	--	--
Unobserved Heterogeneity (Correlation between crash count and severity component)				
%Truck	0.001	1.972	0.001	1.972
Log-Likelihood: -24,122.100; BIC: 48,574.338; Number of Parameters: 37				

Note: -- denotes that the variable is not significant at 90% significant level.

In the severity proportion component, interestingly the parameter associated with the AADT shows positive effect. While this finding is counterintuitive, it requires further investigation. The results also show that the effect of AADT is different for Florida state as indicated by interaction variable. Among the vehicle mix variables, a higher percentage of truck traffic is found to reduce the severity risk. In contrast, the longer urban arterial segments are found to contribute to increased severity. Usually, drivers in the longer segments with no/little change in geometry tend to drive at a higher speed than usual, which might increase the risk of severe crashes. Wider median width (> 20 feet) and narrow inside shoulder width (<8 feet) are found to decrease severity of crashes. A wider median may provide additional safety zone in crashes while narrow shoulders may discourage higher operating speed. Alternatively, as expected, urban arterial segments with speed limit higher than 55 mph have higher probability of severe crashes.

The proposed model system can capture the unobserved correlation between total crash count and crash proportion by severity levels. In our testing, we found the percentage of truck traffic exhibit significant unobserved correlation that affects both crash count and crash severity.

4.3.2 Rural 3-leg STOP controlled (R3ST) intersection facility

In the crash count component of R3ST intersections, the model constant does not have any substantive interpretation. However, our model shows that the model constant is not same across the states and the value is different for Florida and Minnesota states as indicated by state indicator variables in Table 5. These deviations highlight the region-specific influences on the estimates. We use the number of years (5) as an offset variable in the NB model specification.

Among the traffic characteristics, variables associated with both major road AADT and minor road AADT are found to have positive association with total crash count for R3ST intersections (as found in (*51*)). Further, the impact of major road AADT varies across R3ST intersections as indicated by significant standard deviation parameter in Table 5. In the case of Minnesota state, the net impact of major road AADT on crashes is further moderated. Among the vehicle mix indicators, the results show that a higher percentage of major road trucks is more likely to increase the number of intersection crashes (as found in (*6*, *51*)). This could be due to the visibility issues as wider space is required by heavy vehicles for turning movement and differential speeds with other vehicles at the intersections. Among the roadway attributes, intersections with major road speed limit higher than 55 mph are more likely to increase the total number of crashes (as found in (*52*)).

In the severity component for R3ST intersections, major road AADT is found to have a negative impact on crash severity. This could be due to the lower operating speed in the presence of higher volume of traffic at the intersection (*6*, *51*). Further, intuitively, intersections with major roads posted speed limit ≤ 40 mph are found to be associated with lower probability of severe crashes.

TABLE 5 Model Estimation Results of Random Parameters NB-OPFS Model for R3ST Intersection (N=6,500)

Variable Names	Count Component		Severity Proportion Component	
	Estimates	t-stat	Estimates	t-stat
Constant	-10.880	-34.999	--	--
State-Connecticut	--	--	-0.291	-2.838
State-Florida	2.795	34.291	--	--
State-Minnesota	2.989	5.844	--	--

Variable Names	Count Component		Severity Proportion Component	
	Estimates	t-stat	Estimates	t-stat
Ln (Year = 5)	1.000	--	--	--
Threshold Parameters				
Threshold between OP-CP	--	--	-0.124	-0.673
Threshold between CP-BP	--	--	0.451	2.429
Threshold between BP-AP	--	--	1.128	6.124
Threshold between AP-KP	--	--	1.633	8.686
Traffic Characteristics				
Ln (Major road AADT)	0.588	15.338	-0.041	-2.004
Standard Deviation	0.061	7.054	--	--
Ln (Major road AADT) * State-Minnesota	-0.125	-2.219	--	--
Ln (Minor road AADT)	0.502	25.117	--	--
% Major road truck	0.012	3.108	--	--
Roadway Characteristics				
Major road speed limit (base: 41-55 mph)				
Maj SL<=40	--	--	-0.242	-3.574
Maj SL >55	-0.237	-3.815	--	--
Overdispersion Parameter				
Constant	0.690	6.487	--	--
State-Florida	-0.251	-2.569	--	--
Log-Likelihood: -8,174.985; BIC: 16,516.782; Number of Parameters: 19				

Note: -- denotes that the variable is not significant at 90% significant level.

4.3.3 Rural 4-leg STOP controlled (R4ST) intersection facility

The results of the RPMVPLN model show that the model constant in the framework is not the same across the states and the values are different for California and Florida states as indicated by state indicator variables in Table 6. These deviations highlight the region-specific influences on the estimates. We use the number of years (5) as an offset variable in the model specification.

With respect to the traffic characteristics, the findings show that the parameters associated with AADT on both major and minor roads are positively correlated with crash counts across all severities (similar results are found in (*6*, *51*)). Interestingly, a higher percentage of trucks on major roads generally reduces crash counts, especially for more severe crashes, while the same percentage on minor roads increases fatal crash counts in R4ST intersections. The random parameter effects for the percentage of trucks on major roads show variability across intersections and severity levels, as indicated by the main-diagonal values. The non-diagonal values represent the covariance among severity levels due to these random parameter effects, suggesting interdependencies influenced by the truck percentage.

TABLE 6 Model Estimation Results of Random Parameters MVPLN Model for R4ST Intersection (N=3,000)

Variable Names	O	C	B	A	K
Constant	-10.447	-10.628	-10.683	-10.262	-8.873
State-California	-1.789	-2.026	-2.162	-1.797	-1.165
State-Florida	1.659	1.092	1.533	1.903	1.006
Ln (Year = 5)	1.000	1.000	1.000	1.000	1.000
Traffic Characteristics					

Variable Names	O	C	B	A	K
Ln (Major road AADT)	0.717	0.647	0.642	0.384	0.302
Ln (Minor road AADT)	0.469	0.481	0.408	0.458	0.352
% Major road truck	-0.045	-0.066	-0.062	-0.129	-0.196
% Minor road truck	--	--	--	--	0.138
Roadway Characteristics					
Major road speed limit (base: ≤55 mph)					
Maj SL > 55 mph	-0.545	-0.374	--	--	--
Maj SL >55 * State-California	0.536	--	--	--	--
Light (Base: No lighting)	--	--	--	--	-0.464
Variance-Covariance Matrix for Random Effects (% Major road truck)					
	O	C	B	A	K
O	0.007	0.002	0.002	--	--
C		0.009	0.002	--	--
B			0.010	--	--
A				0.022	--
K					0.026
Variance Covariance Matrix for Unobserved Heterogeneity					
O	0.679	0.676	0.617	0.677	0.631
C		0.778	0.657	0.737	0.683
B			0.694	0.679	0.649
A				0.916	0.689
K					0.873
Log-Likelihood: -7,215.850; BIC: 15,232.337; Number of Parameters: 100					

Note: -- denotes that the variable is not significant at 90% significant level.

With regards to the roadway characteristics, higher speed limit on major roads (>55 mph relative to ≤55mph) is associated with reduction of crash counts of lower severities but increase them in California for O crashes. The results align with the expectation because collisions occurring at high speed usually result in higher severity and there is slim possibility to those resulting in property damage only or just a possible injury. The presence of lighting is found to reduce fatal crashes significantly. Furthermore, the statistically significant variance-covariance matrix for unobserved heterogeneity further underscores significant correlations across severity levels, pointing to underlying factors affecting multiple crash severities at an intersection.

5. CONCLUSIONS

The vehicle mix information, defined as traffic volume by vehicle type, has been identified as a significant contributing factor to crash frequency analysis. However, the current version of the HSM predictive methods does not incorporate the vehicle mix information when estimating crashes. The current study estimates and compares the performance of crash frequency and severity systems with vehicle mix information incorporated in different methodological frameworks. Specifically, we build on the HSM approach with two model systems: (a) multivariate Poisson-lognormal model (MVPLN) and (b) negative binomial – ordered probit fractional split model (NB-OPFS). The MVPLN model estimates crash counts by each severity level while NB component estimates total crashes and OPFS component models crash proportion by severity class in the NB-OPFS model framework. We developed advanced variants that account for additional observed and unobserved heterogeneity while accounting for the impact of vehicle mix data. Further, using data from multiple jurisdictions, we develop pooled models that accommodate for jurisdiction-specific observed and unobserved heterogeneity. The models developed are compared with each other and the HSM benchmark model based on a comprehensive set of quantitative and qualitative metrics to identify the most appropriate model system for each facility type. The proposed models are estimated using data from multiple states that include four Highway Safety Information System (HSIS) states including California, Illinois, Minnesota, Washington and three Non-HSIS states including Connecticut, Florida, and Texas. We selected Urban Arterial 4-Lane Divided segment facility (UA4LD), Rural 3-Leg STOP Controlled (R3ST) and Rural 4-Leg STOP Controlled (R4ST) intersection facilities for our analysis.

In the current study, we consider five severity levels for crash frequency estimation including fatal, incapacitating, non-incapacitating, possible injury and no injury crashes. A comprehensive set of independent variables including traffic volume, vehicle mix indicators (truck percentage, single unit truck percentage, and high truck zone), roadway characteristics and state specific indicators are considered. We evaluate the model performance by employing two statistical measures of fit: mean absolute deviation (MAD) and mean squared prediction error (MSPE). A single framework might not perform best across all the dimensions at a facility type. Therefore, we adopt two approaches (total crash approach and severity level scoring approach) that consider the model performances across all the estimated dimensions for final model selection. The total crash approach considers MAD and MSPE values from total crash frequency predictions while the severity level scoring approach considers MAD and MSPE values from five severity levels. Based on the two approaches, the model that performs better across the measures is selected for the respective facility type. The results show that both NB-OPFS and MVPLN frameworks with vehicle mix data performed better than the HSM predictive models that do not consider vehicle mix data. Within these advanced frameworks, the NB-OPFS model performed better for UA4LD segment and R3ST intersection facility while the MVPLN model showed better performance for R4ST intersection facility. Further, within all the frameworks, vehicle mix variables show statistically significant observed and unobserved effects in crash frequency dimensions across the facility types.

The study found that higher truck traffic reduces crash frequency and severity at UA4LD segments. This may be due to factors like lower truck speeds, less disruptive flow, fewer abrupt lane changes, designated truck lanes, advanced traffic management systems (ITS), and strict law enforcement. Further, for R4ST, higher truck traffic decreased crashes, while for R3ST it increased crashes. This difference may be due to the larger intersection areas, more maneuvering space and wider turning radius at 4-leg intersections compared to 3-leg intersections. Taking these results into consideration, transportation engineers and safety planners may prioritize infrastructure investment and road safety initiatives at road segments and intersections with high truck traffic. These investments and initiatives could include dedicated truck lanes, wider lanes, curb adjustments, better signage, ITS, and stricter enforcement. In areas with heavy truck traffic, additional measures for vulnerable road users, such as improved crosswalks, pedestrian signals, and barriers, could also be considered. Additionally, promoting safe driving behaviors through education, driver training programs, and public awareness campaigns can further reduce crash frequency and severity at segments and intersections.

This study is not without limitations. The study considers crash data from multiple years. Due to the lack of detailed multi-year independent variables information, temporal heterogeneity could not be accommodated in the current model estimation. However, the characteristics of the explanatory variables may change over time, affecting the crash frequency/severity along different sections of a facility (*53*). For example, driver behavior may change over the years due to the advancement of the vehicle's technological features, traffic management dynamics, and infrastructure improvements impacting crash patterns. Future research efforts could address this issue by incorporating data from additional states while also accounting for temporal effects with multi-year independent variables information to obtain more precise inference and enhanced predictive power. Further, it would be interesting to explore the effects of the factors including technological advancements, changes in driver behavior, infrastructure development, and societal responses in response to the emergence of electric and autonomous vehicles (EVs and AVs) in addition to more finer resolution vehicle mix information on crash frequency and severity.

ACKNOWLEDGMENTS

The authors would like to gratefully acknowledge all the agencies that provided data including California DOT, Connecticut DOT, Florida DOT, Illinois DOT, Minnesota DOT, Washington State DOT, Highway Safety Information System Database, Signal-4-Analytics. Specifically, the data collection effort of this research is part of the NCHRP project 22-49, "The Effect of Vehicle Mix on Crash Frequency and Crash Severity", which is part of the National Cooperative Highway Research Program (NCHRP). NCHRP is administered by the Transportation Research Board (TRB) and funded by participating member states of the American Association of State Highway and Transportation Officials (AASHTO). NCHRP also receives critical technical support from the Federal Highway Administration (FHWA), United States Department of Transportation. However, the opinions and conclusions expressed or implied in this paper are those of the authors. They are not necessarily those of the TRB, NCHRP, the National Academies, or the program sponsors.

AUTHORS CONTRIBUTION STATEMENT

The authors confirm contribution to the paper as follows: study conception and design:, Naveen Eluru, Tanmoy Bhowmik, Shahrior Pervaz, Dewan Ashraful Parvez, John N. Ivan, Kai Wang, Manmohan Joshi; data collection: Shahrior Pervaz, Manmohan Joshi, Tanmoy Bhowmik; model estimation and validation: Shahrior Pervaz, Manmohan Joshi, Tanmoy Bhowmik, Dewan Ashraful Parvez, Kai Wang; analysis and interpretation of results: Shahrior Pervaz, Manmohan Joshi, Tanmoy Bhowmik, Dewan Ashraful Parvez, Naveen Eluru, John N. Ivan, Kai Wang; draft manuscript preparation: Shahrior Pervaz, Manmohan Joshi, Tanmoy Bhowmik, Dewan Ashraful Parvez, Naveen Eluru, John N. Ivan, Kai Wang. All authors reviewed the results and approved the final version of the manuscript.

DECLARATION OF CONFLICTING INTERESTS

Naveen Eluru is a member of Transportation Research Record's editorial board. The other authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

1. Zeng, Q., H. Wen, H. Huang, J. Wang, and J. Lee. Analysis of Crash Frequency Using a Bayesian Underreporting Count Model with Spatial Correlation. *Physica A: Statistical Mechanics and its Applications*, Vol. 545, 2020, p. 123754. https://doi.org/https://doi.org/10.1016/j.physa.2019.123754.
2. Wen, H., X. Zhang, Q. Zeng, and N. N. Sze. Bayesian Spatial-Temporal Model for the Main and Interaction Effects of Roadway and Weather Characteristics on Freeway Crash Incidence. *Accident Analysis & Prevention*, Vol. 132, 2019, p. 105249.
3. Hasan, T., M. Abdel-Aty, and N. Mahmoud. Freeway Crash Prediction Models with Variable Speed Limit/Variable Advisory Speed. *Journal of transportation engineering, Part A: Systems*, Vol. 149, No. 3, 2023, p. 04022159.
4. Ma, X., S. Chen, and F. Chen. Multivariate Space-Time Modeling of Crash Frequencies by Injury Severity Levels. *Analytic Methods in Accident Research*, Vol. 15, 2017, pp. 29–40. https://doi.org/https://doi.org/10.1016/j.amar.2017.06.001.
5. Huang, H., F. Chang, H. Zhou, and J. Lee. Modeling Unobserved Heterogeneity for Zonal Crash Frequencies: A Bayesian Multivariate Random-Parameters Model with Mixture Components for Spatially Correlated Data. *Analytic Methods in Accident Research*, Vol. 24, 2019, p. 100105. https://doi.org/https://doi.org/10.1016/j.amar.2019.100105.
6. Dong, C., D. B. Clarke, X. Yan, A. Khattak, and B. Huang. Multivariate Random-Parameters Zero-Inflated Negative Binomial Regression Model: An Application to Estimate Crash Frequencies at Intersections. *Accident Analysis & Prevention*, Vol. 70, 2014, pp. 320–329. https://doi.org/https://doi.org/10.1016/j.aap.2014.04.018.
7. Faden, A., M. Abdel-Aty, N. Mahmoud, T. Hasan, and H. Rim. Multivariate Poisson-Lognormal Models for Predicting Peak-Period Crash Frequency of Joint on-Ramp and Merge Segments on Freeways. *Transportation research record*, Vol. 2678, No. 3, 2024, pp. 133–147.
8. Venkataraman, N., V. Shankar, G. F. Ulfarsson, and D. Deptuch. A Heterogeneity-in-Means Count Model for Evaluating the Effects of Interchange Type on Heterogeneous Influences of Interstate Geometrics on Crash Frequencies. *Analytic Methods in Accident Research*, Vol. 2, 2014, pp. 12–20. https://doi.org/https://doi.org/10.1016/j.amar.2014.01.001.
9. Huo, X., J. Leng, Q. Hou, and H. Yang. A Correlated Random Parameters Model with Heterogeneity in Means to Account for Unobserved Heterogeneity in Crash Frequency Analysis. *Transportation Research Record*, Vol. 2674, No. 7, 2020, pp. 312–322. https://doi.org/10.1177/0361198120922212.
10. Huo, X., J. Leng, Q. Hou, L. Zheng, and L. Zhao. Assessing the Explanatory and Predictive Performance of a Random Parameters Count Model with Heterogeneity in Means and Variances. *Accident Analysis & Prevention*, Vol. 147, 2020, p. 105759. https://doi.org/https://doi.org/10.1016/j.aap.2020.105759.
11. Chuanyun, F., and S. Tarek. Random-Parameter Bayesian Hierarchical Extreme Value Modeling Approach with Heterogeneity in Means and Variances for Traffic Conflict–Based

Crash Estimation. *Journal of Transportation Engineering, Part A: Systems*, Vol. 148, No. 9, 2022, p. 04022056. https://doi.org/10.1061/JTEPBS.0000717.

12. Alnawmasi, N., and F. Mannering. The Impact of Higher Speed Limits on the Frequency and Severity of Freeway Crashes: Accounting for Temporal Shifts and Unobserved Heterogeneity. *Analytic Methods in Accident Research*, Vol. 34, 2022, p. 100205. https://doi.org/https://doi.org/10.1016/j.amar.2021.100205.

13. Heydari, S., L. Fu, L. Thakali, and L. Joseph. Benchmarking Regions Using a Heteroskedastic Grouped Random Parameters Model with Heterogeneity in Mean and Variance: Applications to Grade Crossing Safety Analysis. *Analytic Methods in Accident Research*, Vol. 19, 2018, pp. 33–48. https://doi.org/https://doi.org/10.1016/j.amar.2018.06.003.

14. Bhowmik, T., S. Yasmin, and N. Eluru. Accommodating for Systematic and Unobserved Heterogeneity in Panel Data: Application to Macro-Level Crash Modeling. *Analytic Methods in Accident Research*, Vol. 33, 2022, p. 100202. https://doi.org/https://doi.org/10.1016/j.amar.2021.100202.

15. Yasmin, S., and N. Eluru. Latent Segmentation Based Count Models: Analysis of Bicycle Safety in Montreal and Toronto. *Accident Analysis & Prevention*, Vol. 95, 2016, pp. 157–171. https://doi.org/https://doi.org/10.1016/j.aap.2016.07.015.

16. Yasmin, S., S. U. Momtaz, T. Nashad, and N. Eluru. A Multivariate Copula-Based Macro-Level Crash Count Model. *Transportation Research Record*, Vol. 2672, No. 30, 2018, pp. 64–75. https://doi.org/10.1177/0361198118801348.

17. Bhowmik, T., M. Rahman, S. Yasmin, and N. Eluru. Exploring Analytical, Simulation-Based, and Hybrid Model Structures for Multivariate Crash Frequency Modeling. *Analytic Methods in Accident Research*, Vol. 31, 2021, p. 100167. https://doi.org/https://doi.org/10.1016/j.amar.2021.100167.

18. Yasmin, S., and N. Eluru. A Joint Econometric Framework for Modeling Crash Counts by Severity. *Transportmetrica A Transport Science*, Vol. 14, No. 3, 2018, pp. 230–255. https://doi.org/https://doi.org/10.1080/23249935.2017.1369469.

19. Bhowmik, T., S. Yasmin, and N. Eluru. Do We Need Multivariate Modeling Approaches to Model Crash Frequency by Crash Types? A Panel Mixed Approach to Modeling Crash Frequency by Crash Types. *Analytic Methods in Accident Research*, Vol. 24, 2019, p. 100107. https://doi.org/https://doi.org/10.1016/j.amar.2019.100107.

20. Jahan, M. I., T. Bhowmik, and N. Eluru. Enhanced Aggregate Framework to Model Crash Frequency by Accommodating Zero Crashes by Crash Type. *Transportation research record*, Vol. 2678, No. 2, 2024, pp. 506–519.

21. Cai, Q., M. Abdel-Aty, J. Lee, and H. Huang. Integrating Macro- and Micro-Level Safety Analyses: A Bayesian Approach Incorporating Spatial Interaction. *Transportmetrica A: Transport Science*, Vol. 15, No. 2, 2019, pp. 285–306. https://doi.org/10.1080/23249935.2018.1471752.

22. Pervaz, S., T. Bhowmik, and N. Eluru. Integrating Macro and Micro Level Crash Frequency Models Considering Spatial Heterogeneity and Random Effects. *Analytic Methods in*

Accident Research, Vol. 36, 2022, p. 100238. https://doi.org/https://doi.org/10.1016/j.amar.2022.100238.

23. Pervaz, S., T. Bhowmik, and N. Eluru. An Econometric Framework for Integrating Aggregate and Disaggregate Level Crash Analysis. *Analytic Methods in Accident Research*, Vol. 39, 2023, p. 100280. https://doi.org/https://doi.org/10.1016/j.amar.2023.100280.
24. Pervaz, S., T. Bhowmik, and N. Eluru. An Integrated Multi-Resolution Framework for Jointly Estimating Crash Type and Crash Severity. *Analytic Methods in Accident Research*, Vol. 42, 2024, p. 100321. https://doi.org/https://doi.org/10.1016/j.amar.2024.100321.
25. American Association of State Highway Transportation Officials (AASHTO). *Highway Safety Manual*. Washington, DC, 2010.
26. Anastasopoulos, P. Ch. Random Parameters Multivariate Tobit and Zero-Inflated Count Data Models: Addressing Unobserved and Zero-State Heterogeneity in Accident Injury-Severity Rate and Frequency Analysis. *Analytic methods in accident research*, Vol. 11, 2016, pp. 17–32.
27. Bhowmik, T., S. Yasmin, and N. Eluru. A Joint Econometric Approach for Modeling Crash Counts by Collision Type. *Analytic Methods in Accident Research*, Vol. 19, 2018, pp. 16–32. https://doi.org/https://doi.org/10.1016/j.amar.2018.06.001.
28. Shin, E. J. Factors Associated with Different Types of Freight Crashes: A Macro-Level Analysis. *Journal of Safety Research*, Vol. 88, 2024, pp. 244–260. https://doi.org/https://doi.org/10.1016/j.jsr.2023.11.012.
29. Roy, U., A. Farid, and K. Ksaibati. Effects of Pavement Friction and Geometry on Traffic Crash Frequencies: A Case Study in Wyoming. *International Journal of Pavement Research and Technology*, Vol. 16, No. 6, 2023, pp. 1468–1481.
30. Gu, X., X. Yan, L. Ma, and X. Liu. Modeling the Service-Route-Based Crash Frequency by a Spatiotemporal-Random-Effect Zero-Inflated Negative Binomial Model: An Empirical Analysis for Bus-Involved Crashes. *Accident Analysis & Prevention*, Vol. 144, 2020, p. 105674. https://doi.org/https://doi.org/10.1016/j.aap.2020.105674.
31. Chiou, Y.-C., C. Fu, and H. Chih-Wei. Incorporating Spatial Dependence in Simultaneously Modeling Crash Frequency and Severity. *Analytic Methods in Accident Research*, Vol. 2, 2014, pp. 1–11. https://doi.org/https://doi.org/10.1016/j.amar.2013.12.001.
32. Chiou, Y.-C., and C. Fu. Modeling Crash Frequency and Severity with Spatiotemporal Dependence. *Analytic Methods in Accident Research*, Vol. 5–6, 2015, pp. 43–58. https://doi.org/https://doi.org/10.1016/j.amar.2015.03.002.
33. Eluru, N., T. Bhowmik, S. Pervaz, D. A. Parvez, L. Hoover, M. Abdel-Aty, K. Wang, J. N. Ivan, S. Zhao, and M. Joshi. *The Effect of Vehicle Mix on Crash Frequency and Crash Severity*. 2024.
34. Bhat, C. R. Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model. *Transportation Research Part B: Methodological*, Vol. 35, No. 7, 2001, pp. 677–693. https://doi.org/10.1016/S0191-2615(00)00014-X.

35. Yasmin, S., and N. Eluru. Evaluating Alternate Discrete Outcome Frameworks for Modeling Crash Injury Severity. *Accident Analysis & Prevention*, Vol. 59, 2013, pp. 506–521. https://doi.org/https://doi.org/10.1016/j.aap.2013.06.040.
36. Aptech. *Aptech Systems Inc*. 2015.
37. Serhiyenko, V., S. A. Mamun, J. N. Ivan, and N. Ravishanker. Fast Bayesian Inference for Modeling Multivariate Crash Counts. *Analytic Methods in Accident Research*, Vol. 9, 2016, pp. 44–53. https://doi.org/https://doi.org/10.1016/j.amar.2016.02.002.
38. Plummer, M. JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. 2003.
39. Sharafeldin, M., A. Farid, and K. Ksaibati. Examining the Risk Factors of Rear-End Crashes at Signalized Intersections. *Journal of Transportation Technologies*, Vol. 12, No. 4, 2022, pp. 635–650.
40. Ingle, A., and T. J. Gates. Safety Performance of Rural Offset-T Intersections. *Transportation research record*, Vol. 2675, No. 2, 2021, pp. 40–52.
41. Yue, H. Investigating the Influence of Streetscape Environmental Characteristics on Pedestrian Crashes at Intersections Using Street View Images and Explainable Machine Learning. *Accident Analysis & Prevention*, Vol. 205, 2024, p. 107693. https://doi.org/https://doi.org/10.1016/j.aap.2024.107693.
42. Tarko, A. P., H. Bar-Gera, J. Thomaz, and A. Issariyanukula. Model-Based Application of Abbreviated Injury Scale to Police-Reported Crash Injuries. *Transportation research record*, Vol. 2148, No. 1, 2010, pp. 59–68.
43. Burch, C., L. Cook, and P. Dischinger. A Comparison of KABCO and AIS Injury Severity Metrics Using CODES Linked Data. *Traffic injury prevention*, Vol. 15, No. 6, 2014, pp. 627–630.
44. Administrations, F. H. KABCO Injury Classification Scale and Definitions. *Online at: https://safety. fhwa. dot. gov/hsip/spm/conversion_tbl/pdfs/kabco_ctable_by_state. pdf. Accessed on November*, Vol. 30, 2019.
45. Jiang, X., R. W. Lyles, and R. Guo. A Comprehensive Review on the Quasi-Induced Exposure Technique. *Accident Analysis & Prevention*, Vol. 65, 2014, pp. 36–46. https://doi.org/https://doi.org/10.1016/j.aap.2013.12.008.
46. Sharmin, S., J. N. Ivan, S. Zhao, K. Wang, M. J. Hossain, N. Ravishanker, and E. Jackson. Incorporating Demographic Proportions into Crash Count Models by Quasi-Induced Exposure Method. *Transportation Research Record*, Vol. 2674, No. 9, 2020, pp. 548–560. https://doi.org/10.1177/0361198120930230.
47. Alarifi, S. A., M. A. Abdel-Aty, J. Lee, and J. Park. Crash Modeling for Intersections and Segments along Corridors: A Bayesian Multilevel Joint Model with Random Parameters. *Analytic Methods in Accident Research*, Vol. 16, 2017, pp. 48–59. https://doi.org/https://doi.org/10.1016/j.amar.2017.08.002.
48. Liu, J., A. J. Khattak, and B. Wali. Do Safety Performance Functions Used for Predicting Crash Frequency Vary across Space? Applying Geographically Weighted Regressions to

Account for Spatial Heterogeneity. *Accident Analysis & Prevention*, Vol. 109, 2017, pp. 132–142. https://doi.org/https://doi.org/10.1016/j.aap.2017.10.012.

49. Russo, B. J., and P. T. Savolainen. A Comparison of Freeway Median Crash Frequency, Severity, and Barrier Strike Outcomes by Median Barrier Type. *Accident Analysis & Prevention*, Vol. 117, 2018, pp. 216–224. https://doi.org/https://doi.org/10.1016/j.aap.2018.04.023.

50. Islam, M. S., J. N. Ivan, N. E. Lownes, R. A. Ammar, and S. Rajasekaran. Developing Safety Performance Function for Freeways by Considering Interactions between Speed Limit and Geometric Variables. *Transportation Research Record*, Vol. 2435, No. 1, 2014, pp. 72–81. https://doi.org/10.3141/2435-09.

51. Dong, C., S. H. Richards, D. B. Clarke, X. Zhou, and Z. Ma. Examining Signalized Intersection Crash Frequency Using Multivariate Zero-Inflated Poisson Regression. *Safety Science*, Vol. 70, 2014, pp. 63–69. https://doi.org/https://doi.org/10.1016/j.ssci.2014.05.006.

52. Tay, R. A Random Parameters Probit Model of Urban and Rural Intersection Crashes. *Accident Analysis & Prevention*, Vol. 84, 2015, pp. 38–40. https://doi.org/10.1016/J.AAP.2015.07.013.

53. Mannering, F. Temporal Instability and the Analysis of Highway Accident Data. *Analytic Methods in Accident Research*, Vol. 17, 2018, pp. 1–13. https://doi.org/https://doi.org/10.1016/j.amar.2017.10.002.

APPENDIX

TABLE A1 Model Estimation Results of Fixed Parameter NB-OPFS Model for UA4LD Segments (N=7,500)

Variable Names	Count Component		Severity Proportion Component	
	Estimates	t-stat	Estimates	t-stat
Constant	-1.836	-5.487	--	--
State-Minnesota	-6.926	-5.345	--	--
State-Washington	-6.963	-3.598	--	--
Ln (Year = 5)	1.000	--	--	--
Threshold Parameters				
Threshold between OP-CP	--	--	0.711	4.646
Threshold between CP-BP	--	--	1.013	6.601
Threshold between BP-AP	--	--	1.458	9.490
Threshold between AP-KP	--	--	2.667	16.574
Traffic Characteristics				
Ln (AADT)	0.461	14.415	0.033	2.280
Ln (AADT)* State-Florida	--	--	-0.013	-3.796
Ln (AADT)* State-Minnesota	0.622	4.822	--	--
Ln (AADT)* State-Washington	0.582	3.058	--	--
%Truck	-0.030	-4.548	-0.007	-3.114
%Single unit truck	-0.033	-2.486	--	--
Roadway Characteristics				
Ln (Segment length, miles)	1.000	--	0.036	3.133
Lane width (base: ≤12 feet)				
LW>12	-0.167	-1.835	--	--
HTZ*LW>12	-1.015	-4.248	--	--
Median width (base: ≤20 feet)				
MW>20	-0.299	-3.863	-0.065	-2.242
Outside shoulder width (base: ≥8 feet)				
OSW<8	0.418	6.388	--	--
HTZ*OSW<8	0.422	3.583	--	--
Inside shoulder width (base: ≥8 feet)				
ISW<8	0.402	3.557	-0.133	-2.675
Shoulder type (base: unpaved)				
Paved	-0.632	-6.538	--	--
Paved* State-California	0.350	2.088	--	--
Speed limit (base: ≤55 mph)				
SL>55	-0.650	-4.336	0.173	3.728
SL>55* State-California	0.339	1.683	--	--
Overdispersion Parameter				
Constant	2.208	38.717	--	--
State-California	-1.028	-7.531	--	--
State-Washington	-1.686	-13.198	--	--
Log-Likelihood: -24,228.525; BIC: 48,760.420; Number of Parameters: 34				

Note: -- denotes that the variable is not significant at 90% significant level.

TABLE A2 Model Estimation Results of Fixed Parameters NB-OPFS Model for R3ST Intersection (N=6,500)

Variable Names	Count Component		Severity Proportion Component	
	Estimates	t-stat	Estimates	t-stat
Constant	-10.930	-36.493	--	--
State-Connecticut	--	--	-0.291	-2.838
State-Florida	2.764	33.497	--	--
State-Minnesota	2.917	5.534	--	--
Ln (Year = 5)	1.000	--	--	--
Threshold Parameters				
Threshold between OP-CP	--	--	-0.124	-0.675
Threshold between CP-BP	--	--	0.451	2.438
Threshold between BP-AP	--	--	1.128	6.145
Threshold between AP-KP	--	--	1.633	8.715
Traffic Characteristics				
Ln (Major road AADT)	0.611	16.896	-0.041	-2.011
Ln (Major road AADT) * State-Minnesota	-0.118	-2.049	--	--
Ln (Minor road AADT)	0.504	24.418	--	--
% Major road truck	0.012	2.853	--	--
Roadway Characteristics				
Major road speed limit (base: 41-55 mph)				
Maj SL <=40	--	--	-0.242	-3.576
Maj SL >55	-0.245	-3.852	--	--
Overdispersion Parameter				
Constant	1.077	16.012	--	--
State-Florida	-0.439	-4.519	--	--
Log-Likelihood: -8,192.925; BIC: 16,543.882; Number of Parameters: 18				

Note: -- denotes that the variable is not significant at 90% significant level.

TABLE A3 Model Estimation Results of Fixed Parameters MVPLN Model for R4ST Intersection (N=3,000)

Variable Names	O	C	B	A	K
Constant	-10.421	-10.568	-10.62	-10.026	-8.978
State-California	-1.708	-1.865	-2.096	-1.839	-1.003
State-Florida	1.670	1.061	1.530	1.716	0.890
Ln (Year = 5)	1.000	1.000	1.000	1.000	1.000
Traffic Characteristics					
Ln (Major road AADT)	0.703	0.610	0.620	0.308	0.247
Ln (Minor road AADT)	0.458	0.491	0.401	0.481	0.379
% Major road truck	--	--	--	--	--
% Minor road truck	--	--	--	--	--
Roadway Characteristics					
Major road speed limit (base: ≤55 mph)					
Maj SL > 55 mph	-0.525	-0.311	-0.348	--	--
Maj SL >55 * State-California	0.548	--	0.704	--	--
Light (Base: No Lighting)	--	--	--	--	-0.417
Variance-Covariance Matrix					
	O	**C**	**B**	**A**	**K**
O	0.793	0.771	0.685	0.747	0.662
C		0.886	0.729	0.805	0.714
B			0.778	0.739	0.703
A				0.986	0.690
K					0.947
Correlation					
O	--	0.920	0.873	0.845	0.763
C		--	0.878	0.861	0.780
B			--	0.845	0.819
A				--	0.714
K					--
Log-Likelihood: -7,549.000; BIC: 15,899.000					

Note: -- denotes that the variable is not significant at 90% significant level.

TABLE A4 Descriptive Statistics of the Variables for UA4LD Segment Facility

Variables	Min.	Max.	Sum	Mean	Std. Dev.
State- California (1 if yes, 0 otherwise)	0.000	1.000	316.000	0.042	0.201
State- Florida (1 if yes, 0 otherwise)	0.000	1.000	321.000	0.043	0.202
State- Illinois (1 if yes, 0 otherwise)	0.000	1.000	4,567.000	0.609	0.488
State- Minnesota (1 if yes, 0 otherwise)	0.000	1.000	692.000	0.092	0.289
State- Texas (1 if yes, 0 otherwise)	0.000	1.000	1,324.000	0.177	0.381
State- Washington (1 if yes, 0 otherwise)	0.000	1.000	280.000	0.037	0.190
Total segment crashes	0.000	1,078.000	93,832.000	12.511	53.140
Proportion of fatal crashes (KP)	0.000	1.000	24.035	0.003	0.033
Proportion of incapacitating crashes (AP)	0.000	1.000	396.040	0.053	0.137
Proportion of non-incapacitating crashes (BP)	0.000	1.000	461.118	0.061	0.151
Proportion of possible injury crashes (CP)	0.000	1.000	448.120	0.060	0.163
Proportion of no injury crashes (OP)	0.000	1.000	3,812.687	0.508	0.413
Ln (AADT)	2.303	12.459	73,150.499	9.753	0.770
%Truck (Truck AADT*100/AADT)	0.000	49.793	39,064.819	5.209	6.204
%Single Unit Truck (Single Unit Truck AADT*100/AADT)	0.000	24.638	18,270.740	2.436	3.231
Ln (Segment length, mile)	-4.605	2.199	-20,148.390	-2.686	1.150
Paved shoulder type (1 if yes, 0 otherwise)	0.000	1.000	2,019.000	0.269	0.444
Lane width >12 ft (1 if yes, 0 otherwise)	0.000	1.000	608.000	0.081	0.273
Median width >20 ft (1 if yes, 0 otherwise)	0.000	1.000	2,015.000	0.269	0.443
Outside shoulder width < 8 ft (1 if yes, 0 otherwise)	0.000	1.000	4,612.000	0.615	0.487
Inside shoulder width < 8 ft (1 if yes, 0 otherwise)	0.000	1.000	6,724.000	0.897	0.305
Speed limit >55 mph (1 if yes, 0 otherwise)	0.000	1.000	573.000	0.076	0.266
Hight Truck Zone (HTZ) (1 if yes, 0 otherwise)	0.000	1.000	1,134.000	0.151	0.358

TABLE A5 Descriptive Statistics of the Variables for R3ST Intersection Facility

Variables	Min.	Max.	Sum	Mean	Std. Dev.
State- California (1 if yes, 0 otherwise)	0.000	1.000	4,598.000	0.707	0.455
State- Connecticut (1 if yes, 0 otherwise)	0.000	1.000	146.000	0.022	0.148
State- Florida (1 if yes, 0 otherwise)	0.000	1.000	550.000	0.085	0.278
State- Minnesota (1 if yes, 0 otherwise)	0.000	1.000	1,206.000	0.186	0.389
Total intersection crashes	0.000	199.000	8,991.000	1.383	5.164
Proportion of fatal crashes (KP)	0.000	1.000	41.625	0.006	0.065
Proportion of incapacitating crashes (AP)	0.000	1.000	82.370	0.013	0.083
Proportion of non-incapacitating crashes (BP)	0.000	1.000	259.456	0.040	0.151
Proportion of possible injury crashes (CP)	0.000	1.000	373.153	0.057	0.185
Proportion of no injury crashes (OP)	0.000	1.000	1,174.397	0.181	0.343
Ln (Major road AADT)	4.248	11.478	53,624.442	8.250	1.085
Ln (Minor road AADT)	0.000	10.235	30,744.337	4.730	1.655
%Major road truck (Major road truck AADT*100/Major road AADT)	0.000	72.196	26,126.607	4.019	6.031
Speed limit ≤ 40 mph (1 if yes, 0 otherwise)	0.000	1.000	958.00	0.147	0.354
Speed limit >55 mph (1 if yes, 0 otherwise)	0.000	1.000	2,248.000	0.346	0.476

TABLE A6 Descriptive Statistics of the Variables for R4ST Intersection Facility

Variables	Min.	Max.	Sum	Mean	Std. Dev.
State- California (1 if yes, 0 otherwise)	0.000	1.000	1,382.000	0.461	0.499
State- Connecticut (1 if yes, 0 otherwise)	0.000	1.000	42.000	0.014	0.118
State- Florida (1 if yes, 0 otherwise)	0.000	1.000	132.000	0.044	0.205
State- Minnesota (1 if yes, 0 otherwise)	0.000	1.000	1,444.000	0.481	0.500
Total intersection crashes	0.000	1,525.000	9,310.000	3.103	28.428
Proportion of fatal crashes (KP)	0.000	1.000	57.561	0.019	0.103
Proportion of incapacitating crashes (AP)	0.000	1.000	66.920	0.022	0.113
Proportion of non-incapacitating crashes (BP)	0.000	1.000	230.068	0.077	0.194
Proportion of possible injury crashes (CP)	0.000	1.000	339.911	0.113	0.238
Proportion of no injury crashes (OP)	0.000	1.000	835.539	0.279	0.373
Ln (Major road AADT)	4.700	11.446	25,360.259	8.453	0.986
Ln (Minor road AADT)	0.000	10.077	16,606.961	5.536	1.582
%Major road truck (Major road truck AADT*100/Major road AADT)	0.000	48.847	19,360.128	6.453	6.741
%Minor road truck (Minor road truck AADT*100/Minor road AADT)	0.000	23.913	2,789.775	0.930	1.672
Speed limit >55 mph (1 if yes, 0 otherwise)	0.000	1.000	1,064.000	0.355	0.478
Presence of light (1 if yes, 0 otherwise)	0.000	1.000	514.000	0.171	0.377