RESEARCH REPORT

Institute of Transportation Studies | UNIVERSITY OF CALIFORNIA

Assessing the Variation of Curbside Safety at the City Block Level

Aditya Medury, Ph.D., Postdoctoral Researcher, Safe Transportation Research and Education Center, University of California, Berkeley

Dimitris Vlachogiannis, Graduate Student Researcher, Department of Civil and Environmental Engineering, University of California, Berkeley

Offer Grembek, Ph.D., Co-Director, Safe Transportation Research and Education Center, University of California, Berkeley

June 2020

Berkeley Institute of Transportation Studies

Report No.: UC-ITS-2019-15 | DOI: 10.7922/G2Z899Q9

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
UC-ITS-2019-15	N/A	N/A

4. Title and Subtitle Assessing the Variation of Curbside Safety at the City Block Level	**5. Report Date** June 2020
	6. Performing Organization Code ITS Berkeley
7. Author(s) Aditya Medury, Ph.D., https://orcid.org/0000-0001-6472-8023 Dimitris Vlachogiannis, https://orcid.org/0000-0001-5486-0274 Offer Grembek, Ph.D., https://orcid.org/0000-0003-1869-9457	**8. Performing Organization Report No.** N/A
9. Performing Organization Name and Address Institute of Transportation Studies, Berkeley 109 McLaughlin Hall, MC1720 Berkeley, CA 94720-1720	**10. Work Unit No.** N/A
	11. Contract or Grant No. UC-ITS-2019-15
12. Sponsoring Agency Name and Address The University of California Institute of Transportation Studies www.ucits.org	**13. Type of Report and Period Covered** Final Report (July 2018 – June 2020)
	14. Sponsoring Agency Code UC ITS

15. Supplementary Notes
DOI:10.7922/G2Z899Q9

16. Abstract
Investigating the dynamics behind the likelihood of vehicle crashes has been a focal research point in the transportation safety field for many years. However, the abundance of data in today's world generates opportunities for deeper comprehension of the various parameters affecting crash frequency. This study incorporates data from many different sources including geocoded police-reported crash data, curbside infrastructure data and socio-demographic data for the city of San Francisco, CA. Findings revealed that the GFMNB model provides a better statistical fit than the FMNB and NB model in terms of AIC and log likelihood, while the NB model outperformed both mixture models in terms of BIC due to model complexity of the latter. Among the signicant variables, TNC pick-ups/dropoffs and duration of parked vehicles were positively associated with segment-level crashes.

17. Key Words	18. Distribution Statement
Traffic safety, crash data, ridesourcing, curbs, cities, crash risk forecasting, mathematical models	No restrictions.

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	36	N/A

Form Dot F 1700.7 (8-72) Reproduction of completed page authorized

Institute of Transportation Studies | UNIVERSITY OF CALIFORNIA

Assessing the Variation of Curbside Safety at the City Block Level

Aditya Medury, Ph.D., Postdoctoral Researcher, Safe Transportation Research and Education Center, University of California, Berkeley

Dimitris Vlachogiannis, Graduate Student Researcher, Department of Civil and Environmental Engineering, University of California, Berkeley

Offer Grembek, Ph.D., Co-Director, Safe Transportation Research and Education Center, University of California, Berkeley

June 2020

Berkeley Institute of Transportation Studies

Report No.: UC-ITS-2019-15 | DOI: 10.7922/G2Z899Q9

Table of Contents

Contents

Acknowledgements

The researchers would like to thank the financial support received from the Road Repair and Accountability Act of 2017 (SB 1) bill. We would also like to thank the San Francisco Department of Public Health (SFDPH), San Francisco Metropolitan Transportation Authority (SFMTA) and San Francisco County Transportation Authority (SFCTA) for their support in providing access to traffic volume data.

Executive Summary

Executive Summary

Investigating the dynamics behind the likelihood of vehicle crashes has been a focal research point in the transportation safety field for many years. However, the abundance of data in today's world generates opportunities for deeper comprehension of the various parameters affecting crash frequency. This study incorporates data from many different sources including geocoded police-reported crash data, curbside infrastructure data and socio-demographic data for the city of San Francisco, CA.

Specifically, we analyzed 5 years of crashes (2013–2017) occurring at the segment level of roads using explanatory variables pertaining to mode-specific traffic volume estimates, curbside infrastructure (e.g., on-street parking, bicycle lanes), block-level transportation network company pick-up/drop-off estimates, and socio-economic data (e.g., percentage of zero vehicle households). We also excluded segments that were less relevant to this research, such as, freeways, parks, private and pedestrian streets.

In order to handle over-dispersion in crash data, we estimated negative binomial (NB) models. In addition, to capture additional unobserved heterogeneity, two-component finite mixture negative mixture models were formulated, one with fixed priors (FMNB) and another with varying priors (GFMNB).

Findings revealed that the GFMNB model provides a better statistical fit than the FMNB and NB model in terms of AIC and log likelihood, while the NB model outperformed both mixture models in terms of BIC due to model complexity of the latter. Among the significant variables, TNC pick-ups/drop-offs and duration of parked vehicles were positively associated with segment-level crashes.

Contents

Chapter 1

Introduction

The surge in demand for shared-mobility services has made the presence of transportation network companies (TNCs), such as Uber and Lyft, ubiquitous across California in a short span of time. In 2016, TNCs were serving 15% of all intra-San Francisco vehicle trips, which was equal to 12 times the number of taxi trips [5], while in New York in 2016, TNC ridership equaled that of yellow cab. This number went on to double annually between 2014 and 2016 [29]. Given their rapid rise, cities currently lack metrics to assess their impact on the transportation system and appropriately regulate and plan for them moving forward. The need for effective system-level policy guidance to address operational and safety concerns is most noticeable at roadside curbs, which are in constant demand from a variety of multimodal road users, TNC and non-TNC drivers, buses, cyclists, delivery vehicles, etc. The unfettered consumption of this shared resource leads to a tragedy of commons, and hinders a city's ability to develop a safe system on which no one can be severely or fatality injured on.

In order to improve the operational effectiveness of curbside usage, several efforts are being made across Unites States. In 2011, San Francisco Municipal Transportation Authority (SFMTA) launched SFpark, a smart parking initiative, which uses thousands of computerized meters to reduce cruising through demand-responsive pricing and thereby increasing parking availability and improving the operational efficiency of SFMTA's buses and trolleys [8].Zalewski, Buckley, and Weinberger [35] provide a summary of various policy solutions being adopted by cities manage curbside demand, ranging from smart pricing, curbcuts, loading zone restrictions, bike share systems on roadbeds, parklets, etc. However, the impact of these curbside management solutions on segment-level safety is not well understood.

This study aims to quantify the crash risk along the city block segment level as a function of different infrastructure and traffic-related parameters. What distinguishes this study from previous work in the field of safety is the adjustment to the recently established urban transportation reality dominated by TNCs. More specifically, in order to explore the impact of TNC pick-ups/drop-offs and detailed vehicle turnovers and parking spaces, a case study is undertaken for the city of San Francisco, combining crash data along street segments with curbside infrastructure and TNC mobility data to parse out how crash occurrence varies across block segments.

1.1 Literature Review

1.1.1 Work on Segment Level Modeling

Lord and Mannering [19] underline the methodological issues arising in crash frequency analysis. They emphasize upon potential over-dispersion (and on occasion under-dispersion) in crash data, which if present restrict the application of Poisson models as they assume variance and mean to be equal. In contrast, Negative Binomial (or else Poisson Gamma) models can account for over-dispersion although they can be adversely influenced by the low sample-mean and small sample size bias.

Negative binomial models are more commonly used to estimate or predict the number of crashes based on information such as geometric, demographic, or infrastructural characteristics. Sawalha and Sayed [28] investigated accident prediction models for estimating the safety performance of urban arterial roadways in the Greater Vancouver Regional District, British Columbia, Canada. Using negative binomial error structure models to account for overdispersion in their data, they concluded the variables with significant effect on accident occurrence were section length, traffic volume, unsignalized intersection density, driveway density, pedestrian crosswalk density, number of traffic lanes, type of median and type of land use. The study estimated that conversion from an undivided arterial to one with a raised-curb median could result, on average, in a 10% accident reduction. Das and Abdel-Aty [10] observed that higher ADT increases crash frequency while sensitivity analysis revealed ADT as the major factor for the variation in crash counts on urban arterials. Absence of on-street parking can lead to injuries of lower severity as the fixed road side objects are replaced with the much softer parked vehicles. Park et al. [26] showed that the addition of a bike lane on urban arterials significantly improves safety, specifically for bike crashes.

Crash frequency modeling should also properly account for unobserved heterogeneity [19]. Unobserved reasons that cannot be attributed to observable features and characteristics of the road conditions may affect crash outcomes. When evaluating the safety of road segments, prior identification of underlying groups of segments with similar dynamics and response can significantly improve the analysis. The employment of models with fixed parameters for crash modeling is an example of not accounting for unobserved heterogeneity. Constraining the effect of an exploratory variable to be the same for all observations, the resulting parameter estimates may be biased and the inferences drawn erroneous. Unobserved heterogeneity can also be referred to as omitted variable bias when the unobserved characteristics are correlated with a predictor that is included in a model.

One way to account for unobserved effects like spatio-temporal considerations are random effect models [16]. Another approach to tackle the problem of unobserved heterogeneity is modeling with random parameters which allows each estimated parameter of the model to vary across each individual observation in the dataset [3, 4]. Alarifi et al. [2] explored multilevel Poison-lognormal (MPLN) joint models with spatial corridor and sub-corridor random effects terms and MPLN joint models with random parameters, varying across corridors and sub corridors. In terms of the roadway segments-related variables, AADT, driveway density, one-way roads versus two-way roads and roadway classification (principal arterial versus others) were found to be significant. Although such models can significantly improve the statistical fit, random-parameter models are complex and sometimes hard to estimate.

A third method often used to account for unobserved heterogeneity is finite mixture modeling which assumes the presence of latent groups of subjects responding in similar manner to the explanatory variables within a given population [33]. Finite mixture models are semi-parametric, and therefore do not require any distributional assumptions for the mixing variable[11]. A generalization of the finite mixture models with the parameterization of prior component probability distribution was presented by Zou, Zhang, and Lord [36].

1.1.2 Accounting for Intersection Influence Areas

The influence area of an intersection has been defined in many different ways throughout literature. Wang et al. [32] collected data for a sample of 177 four-legged signalized intersections from the state of Florida and showed that variable safety influence areas for intersection approaches improve safety analysis over the 250 ft boundary which is the default used in many states. In Bindra, Ivan, and Jonsson [6], the authors consider intersection related crashes those occurring within 250 ft of the center of the intersection on any leg. Finally, another approach [9] examined intersection safety along state roads in Utah by estimating intersection safety influence areas based on the stopping sight distance for an average approach speed of 40 mph (around 500 ft). Yet,investigating the influence area of 35 hazardous intersections results showed that a 100-ft radius was applicable to about 25 intersections and only two of the intersections appeared to truly have a 500-ft radius of influence area, pointing out that the use of a large radius tended to overestimate the crash risk.

Chapter 2

Segment Level Modeling

2.1 Modeling Methodology

2.1.1 Negative Binomial model

Given the non-negative integer nature of crashes, negative binomial (NB) models are extensively used for crash frequency modeling. Using the negative binomial distribution also accounts for the over-dispersion in the data generated by unobserved heterogeneity. The negative binomial regression arises from a two-stage model for the distribution of number of crashes, n_i, which follows a Poisson distribution with the mean λ_i, and,λ_i follows the Gamma distribution with shape parameter ϕ and scale parameter, μ_i/ϕ:

$$\begin{aligned} n_i|\lambda_i &\sim \text{Poisson}\left(\lambda_i\right), && (2.1)\\ \lambda_i &\sim \text{Gamma}\left(\phi, \mu_i/\phi\right) && (2.2)\\ p(n_i) &= \int_0^\infty p(n_i|\lambda_i) f(\lambda_i) d\lambda_i && (2.3)\\ &= \frac{\Gamma(\phi+n_i)}{\Gamma\left(\phi\right)\Gamma\left(n_i+1\right)} \left(\frac{\phi}{\phi+\mu_i}\right)^{\phi} \left(\frac{\mu_i}{\phi+\mu_i}\right)^{n_i}. && (2.4) \end{aligned}$$

Herein, $\Gamma()$ is the Gamma function and $f(n_i)$ is the marginal probability of observing n_i crashes. Given a vector of explanatory variables, $\mathbf{X_i}$, the mean, $\mathrm{E}(n_i)$, and variance, $\mathrm{VAR}(n_i)$, for a given segment, i, can be expressed as:

$$\begin{aligned} \mathrm{E}_{NB}(n_i) &= \mu_i = exp(\beta \mathbf{X_i}) && (2.5)\\ \mathrm{VAR}_{NB}(n_i) &= \left(1+\frac{\mu_i}{\phi}\right)\mu_i && (2.6) \end{aligned}$$

In order to correct for regression-to-the-mean, an Empirical Bayes (EB) adjustment is typically applied to observed crash frequencies. The EB method combines a site's historical crash data with the expected number of crashes estimated based on the site characteristics (e.g., $\mu_i = exp(\beta\mathbf{X_i})$ The EB-adjustment is estimated using the posterior mean, whose distribution is informed by the Bayes' rule:

$$p(\lambda_i|n_i) = \frac{p(n_i, \lambda_i)}{p(n_i)} \propto p(n_i|\lambda_i)p(\lambda_i) \tag{2.7}$$

Using the posterior distribution information, the EB-adjusted crash frequency is estimated as follows:

$$\mathrm{E}_{NB}(\lambda_i|n_i) = \left(\frac{\mu_i}{\mu_i + \phi}\right) n_i + \left(\frac{\phi}{\mu_i + \phi}\right) \mu_i. \tag{2.8}$$

Finally, a metric similar to EB that is also considered for prioritization is the potential for safety improvement (PSI), which is defined as excess expected average crash frequency with EB adjustment, and is calculated as follows:

$$\mathrm{PSI}_{NB,i} = \mathrm{E}_{NB}(\lambda_i|n_i) - \mu_i \tag{2.9}$$

2.1.2 Generalized Finite Mixture Negative Binomial model

A finite mixture negative binomial model utilizes a finite number (K) of unobserved categories/latent classes to capture the unobserved heterogeneity in crash data [13, 24]. The crash count at a location, n_i, follows the Poisson distribution with the mean crash rate λ_i, and the λ in turn follows a K-component finite mixture of gamma distribution [37]:

$$n_i|\lambda_i \sim \mathrm{Poisson}(\lambda_i), \tag{2.10}$$

$$p(\lambda_i) = \sum_{k=1}^{K} \pi_{ik} p_k(\lambda_{ik}), \tag{2.11}$$

where, $\pi_{ik} = \pi_k(\gamma, z_i)$ is the prior probability of component k. For the component weights π_{ik}, it holds that

$$\sum_{k=1}^{K} \pi_{ik} = 1$$
$$\pi_{ik} > 0, \forall k.$$

The marginal distribution of n_i follows a mixture of NB distributions with probability density function, mean and variance defined as follows:

$$p(n_i|\mu_i,\mathbf{\Theta}) = \sum_{k=1}^{K} \pi_{ik} \left(\frac{\Gamma(\phi_k + n_i)}{\Gamma(\phi_k)\Gamma(n_i+1)} \left(\frac{\phi_k}{\phi_k + \mu_{ik}} \right)^{\phi_k} \left(\frac{\mu_{ik}}{\phi_k + \mu_{ik}} \right)^{n_i} \right) \tag{2.12}$$

$$E(n_i|\mathbf{X_i},\mathbf{\Theta}) = \sum_{k=1}^{K} \pi_{ik}\mu_{ik} \tag{2.13}$$

$$VAR(n_i|\mathbf{X_i},\mathbf{\Theta}) = E(n_i|\mathbf{X_i},\mathbf{\Theta}) + (\sum_{k=1}^{K} \pi_{ik}\mu_{ik}^2(1+\frac{1}{\phi_k}) - E(n_i|\mathbf{X_i},\mathbf{\Theta})^2) \tag{2.14}$$

where μ_{ik} is the mean value of crash frequency of component k, modeled as $\mu_{ik} = exp(\beta_k \mathbf{X_i})$ and β_k is a vector of the regression coefficients for component k.

The weight $\pi_k(\gamma, z)$ indicates the a-priori probability for an observation to come from this component and may depend on further variables. In this work, the prior probability in modeled using multinomial/binary logit framework using explanatory variables z and coefficients γ. $\mathbf{\Theta} = \{(\beta_1, ..., \beta_K), (\phi_1, ..., \phi_K), \gamma\}$ is the vector of all parameters.

The posterior probability of an observation, i belong to component, k is given by:

$$w_{ik} = \pi_{ik} p_k(n_i)/p(n_i), \tag{2.15}$$

where, $p_k(n_i)$ represents marginal probability of observing n_i crashes conditional on the observation belonging to component k.

The EB estimate for the GFMNB model is given by:

$$E_{GFMNB}(\lambda_i|n_i) = \sum_{k=1}^{K} w_{ik} \left\{ \left(\frac{\mu_{ik}}{\mu_{ik} + \phi_k} \right) n_i + \left(\frac{\phi_k}{\mu_{ik} + \phi_k} \right) \mu_{ik} \right\} \tag{2.16}$$

$$\tag{2.17}$$

Two different functions for the derivation of the priors are implemented in the context of this research. One for constant component weights without a class membership model (FMNB),i.e., $\pi_{ik} = \pi_k$ and one for a finite mixture model with a class membership model estimated using multinomial logit models (GFMNB).

In this paper, the finite mixture model was estimated using a expectation maximization (EM) algorithn, and implemented using the R package, *flexmix* [15]. For a more detailed derivation of the EB estimates for a GFMNB approach, the readers are encouraged to refer to Zou et al. [37].

2.2 Data Collection and Processing

This study incorporated data from various sources aiming to achieve a deep understanding of crash dynamics at the city block level.

2.2.1 Crash data

Geocoded crash data were provided for this study by the San Francisco Department of Public Health for 2013–2017. Given the aim of this study is to explore the in-block safety related dynamics, crashes related to the main intersection area were removed from the database. In addition, a 12 meter buffer around the main intersection area was also removed to avoid the inclusion of crashes occurring at crosswalks. While we acknowledge that this represents a relatively small influence area for an intersection, it allows a majority of the block to be analyzed, especially since bus stops and parking spaces near the edge of the block would otherwise get excluded. .

2.2.2 Traffic volumes

The motor vehicle AADT and peak hour bicycle volume estimates were obtained from the San Francisco County Transportation Authority. The counts represent typical weekday travel calibrated for 2015 using San Francisco's CHained Activity Modeling Process (SF-CHAMP) model [21].

2.2.3 Curbside infrastructure data

The parking data utilized for the analysis apart from the availability and type of parking spots also included the parking meter transactions. This variable can reveal crash risk and constitutes a surrogate measure for traffic demand in the area. Moreover, the analysis incorporated geocoded transit stop and traffic signs and signals data, as well as bicycle infrastructure and speed limit data.

2.2.4 TNC pick-up/drop-off

This study incorporated pick-ups and drop-offs data from Uber and Lyft APIs using data collected by the San Francisco County Transportation Authority [5] at the Transportation Analysis Zone (TAZ) level from several weeks during Fall 2016. In the absence of more fine-grained data with yearly activity of pick-ups and drop-offs, we assume that these metrics are approximately representative for the entire time period of the study. In particular, the core assumption is that the relative level of TNC activity remained the same during the time period, as a uniform growth in TNC activity across the entire city would be equivalent to a constant factor adjustment, and would not affect the estimation process. Furthermore, to restrict the scope of the analysis to locations with meaningful TNC activity, the analysis was restricted to "Transportation Analysis Zones" (TAZs) with at least 40 daily pick-ups and drop-offs.

2.2.5 On-street parking data

The shapefiles of all the on-street parking meters (made available through San Francisco's open data initiative) were utilized for this study. Furthermore, individual parking transactions for 2018–2019 were also obtained from the open data repository. In order to explore the influence of vehicle turnover at meter parking space on traffic safety, it is assumed that the relative trends for the average daily transactions at the parking meters remained the same during the study period of 2013–2017.

2.2.6 Socio-economic data

The Smart Location Database [27] summarizes more than 90 different indicators associated with the built environment and location efficiency. Indicators include density of development, diversity of land use, street network design, and accessibility to destinations as well as various demographic and employment statistics.

Collectively, the data sources discussed above provide a variety of traffic, infrastructure and demographic variables to analyze segment-level crashes. Table 2.1 provide a summary of the primary variables considered in the study.

Table 2.1: Summary statistics of variables considered

Variable	**Min**	$\mathbf{q_1}$	$\widetilde{\mathbf{x}}$	$\bar{\mathbf{x}}$	$\mathbf{q_3}$	**Max**
No. of collisions (2013-2017)	0.00	0.00	0.00	0.42	0.00	17.00
Length	0.17	57.20	86.02	99.57	134.93	1242.03
AADT	0.00	240.60	1592.56	4433.23	5465.72	62287.93
Peak hour bicycle volumes	0.00	3.82	27.04	189.02	129.51	7392.41
Median presence (0/1)	0.00	0.00	0.00	0.10	0.00	1.00
No. of lanes	0.00	2.00	2.00	2.20	2.00	6.00
On-street parking spaces	0.00	0.00	0.00	3.09	0.00	110.00
Fraction of streets that are local	0.00	0.00	1.00	0.71	1.00	1.00
Presence of bicycle network	0.00	0.00	0.00	0.20	0.00	1.00
No. of bus stops	0.00	0.00	0.00	0.30	0.00	6.00
Presence of bus lines	0.00	0.00	0.00	0.35	1.00	1.00
Adjoining intersections, signalized	0.00	0.00	0.00	0.37	1.00	1.00
Adjoining intersections, 4-legged	0.00	0.00	0.00	0.38	1.00	1.00
Off-street parking presence (0/1)	0.00	0.00	0.00	0.09	0.00	1.00
Presence of regional transit stations	0.00	0.00	0.00	0.12	0.00	1.00
Ave. daily park meter transaction (in 1000 minutes), multi-space	0.00	0.00	0.00	0.09	0.00	42.62
Ave. daily park meter transaction (in 1000 minutes), single-space	0.00	0.00	0.00	0.18	0.00	14.13
Speed limit $\geq 35mph$	0.00	0.00	0.00	0.03	0.00	1.00

One-way street (0/1)	0.00	0.00	0.00	0.18	0.00	1.00
Ave. daily TNC activity	3.68	5.51	6.07	6.03	6.59	7.79
Percentage of zero vehicle households	0.00	0.12	0.23	0.29	0.41	0.98

There were cases of segments with missing speed limit data and some cases of segments which were impertinent to this research (e.g., street types, like freeways, parks, private and pedestrian streets), all of which were removed. The final dataset included 8213 segments 2.1.

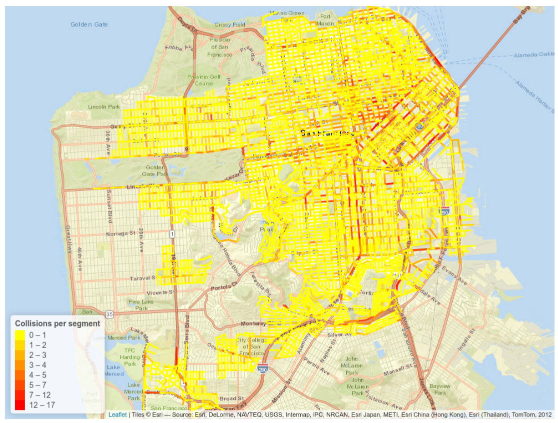


Figure 2.1: Illustration of road segments in scope along with their associated number of collisions per segment

2.3 Model estimation results and findings

Table 2.2 provides a summary of different model selection criteria (Log-Likelihood (LL), Akaike Informaiton Criterion (AIC) and Bayesian Information Criterion (BIC)) for three nested negative binomial models. The first NB considered only segment-level models, while the second NB model also incorporated adjoining intersection information. The third NB model alternative incorporated TAZ-level variables on TNC pick-ups/drop-offs and demographic information such as percentage of households with no vehicles. Based on all the three likelihood-based model selection criteria, we conclude that the third model best explains the segment-level crashes among the NB model specifications. The Superior performance of models including adjoining intersection variables can potentially be attributed to the smaller influence area chosen for excluding main intersection-related crashes. Subsequently, the improved performance from the inclusion of macroscopic variables (which are discussed in more detail below) is consistent with more focused research efforts that have demonstrated the value of macroscopic explanatory variables in microscopic crash models (see [18]).

Table 2.2: Alternative negative binomial model specifications

ID	Model Type	Number of Parameters	LL	AIC	BIC
1	NB with segment variables only	17	-5588.0	11212.1	11338.3
2	NB with segment and adjoining intersection variables	19	-5518.8	11077.6	11217.9
3	NB with segment, intersection, and TAZ-level TNC and demographic variables	21	**-5472.5**	**10988.9**	**11143.2**

Once the final NB model was identified, two-component FMNB and GFMNB models were estimated using the same variables as in the NB model. Since the finite mixture model estimation using EM converges to a local minimum, we estimated FMNB and GFMNB models for 20 iterations with random initial states. The coefficients associated with the best NB model, along with those for the two-component FMNB and GFMNB models are shown in Table 2.3. The coefficients for the class membership model, estimated using a binary logit model, are shown in 2.4. Finally, a summary of the component-level crash, built environment and demographic characteristics are shown in Table 2.6.

Table 2.3: Statistical models

Variables	**NB**	**FMNB**		**GFMNB**	
		Comp. 1	Comp. 2	Comp. 1	Comp. 2
Intercept	-7.94^{***}	-8.68^{***}	-7.83^{***}	-9.09^{***}	-4.38^{***}
	(0.31)	(0.65)	(0.71)	(0.59)	(0.87)
Log (Segment length)	0.87^{***}	0.90^{***}	0.88^{***}	0.91^{***}	0.79^{***}
	(0.04)	(0.09)	(0.09)	(0.07)	(0.09)
Log (AADT)	0.07^{***}	0.09^{***}	$0.06^{\cdot}$	0.07^{***}	-0.01
	(0.01)	(0.02)	(0.03)	(0.01)	(0.04)
Log (Peak-hour bicycle volume)	0.09^{***}	-0.04^{*}	0.37^{***}	-0.00	$0.08^{\cdot}$
	(0.01)	(0.02)	(0.04)	(0.01)	(0.04)
Presence of median	-0.26^{***}	-0.37^{*}	$-0.31^{\cdot}$	-0.41^{*}	-0.28^{*}
	(0.07)	(0.18)	(0.17)	(0.18)	(0.11)
Number of lanes	0.16^{***}	0.14^{**}	0.17^{**}	0.33^{***}	-0.01
	(0.03)	(0.05)	(0.06)	(0.05)	(0.05)
Local street (YES/NO)	-0.38^{***}	-0.39^{**}	-0.32^{*}	-0.31^{**}	-0.30^{**}
	(0.06)	(0.12)	(0.13)	(0.10)	(0.10)
Within bicycle network	0.24^{***}	0.06	0.14	0.09	0.01
	(0.05)	(0.11)	(0.11)	(0.11)	(0.09)
Number of bus stops	0.09^{**}	0.00	0.15^{*}	-0.02	0.17^{***}

Variables	NB	FMNB		GFMNB	
		Comp. 1	Comp. 2	Comp. 1	Comp. 2
	(0.03)	(0.06)	(0.06)	(0.05)	(0.05)
Presence of bus lines	0.17^{**} (0.05)	0.28^{*} (0.12)	0.07 (0.13)	0.40^{***} (0.11)	−0.02 (0.10)
Presence of regional transit station	0.37^{***} (0.05)	0.45^{***} (0.10)	0.24^{*} (0.12)	0.45^{***} (0.10)	0.27^{**} (0.09)
Number of on-street parking meters	0.03^{***} (0.00)	0.02^{*} (0.01)	0.08^{***} (0.01)	0.03^{***} (0.01)	0.02^{**} (0.01)
Presence of off-street parking	0.18^{***} (0.06)	0.22^{*} (0.10)	0.12 (0.13)	0.33^{***} (0.09)	0.06 (0.09)
Mean duration of parking meter transactions (multi-space)	-0.08^{***} (0.02)	−0.05 (0.03)	-0.28^{***} (0.05)	-0.31^{**} (0.11)	-0.06^{*} (0.03)
Mean duration of parking meter transactions (single-space)	-0.23^{***} (0.05)	−0.02 (0.05)	-1.34^{***} (0.27)	-0.26^{*} (0.11)	-0.24^{*} (0.11)
Speed limit ($\geq$ 35 mph)	0.42^{***} (0.10)	1.08^{***} (0.21)	−0.36 (0.29)	0.95^{***} (0.19)	−0.10 (0.16)
One-way street	-0.20^{***} (0.06)	-0.30^{*} (0.12)	−0.03 (0.16)	−0.05 (0.11)	-0.25^{*} (0.10)
Adjoining intersections, signalized	0.50^{***} (0.06)	0.60^{***} (0.15)	0.34^{*} (0.14)	0.51^{***} (0.09)	0.23^{*} (0.10)
Adjoining intersections, 4-legged	-0.10^{*} (0.04)	0.21^{*} (0.10)	-0.39^{***} (0.11)	0.11 (0.08)	-0.31^{***} (0.08)
Log (Daily average TNC pick-ups/drop-offs)	0.21^{***} (0.03)	0.29^{***} (0.08)	0.10 (0.08)	0.27^{***} (0.06)	0.03 (0.08)
Percentage of zero vehicle households	0.67^{***} (0.11)	1.23^{***} (0.24)	−0.16 (0.33)	0.61^{**} (0.21)	0.82^{***} (0.21)
ϕ	2.46	9.40	4.72	10.88	3.62
Class Membership	–	52.0%	48.0%	52.0%	48.0%
AIC	10988.92	10898.42		**10876.93**	
BIC	**11143.22**	11214.03		11234.62	
Log Likelihood	−5472.46	-5404.21		**-5387.47**	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table 2.4: Class membership model for GFMNB component 2

Variables	GFMNB Comp. 2
(Intercept)	-15.13^{***}
	(4.10)
Log (Segment length)	0.84^{*}
	(0.39)
Log (AADT)	0.45^{**}
	(0.14)
Log (Peak-hour bicycle volume)	1.19^{***}
	(0.21)
Presence of bus lines	-0.20
	(0.44)
Log (Daily average TNCpick-ups/drop-offs)	0.27
	(0.35)
Percentage of zero vehi-cle households	-1.11
	(1.00)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

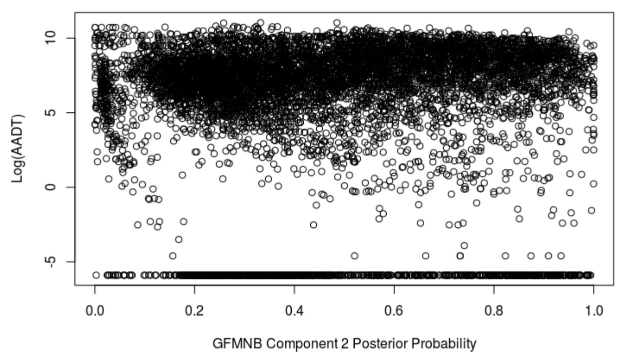
2.4 Interpretation of coefficients

2.4.1 Segment-level variables

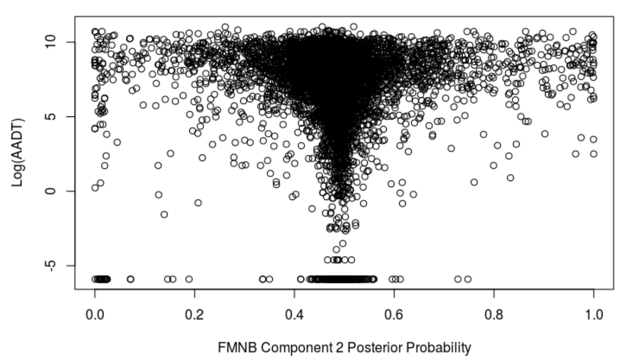
Within each of the NB models, segment length has been modeled as a variable instead of as an offset term. Considering that the coefficient of logarithm of segment length is significantly different from 1 (given the associated standard error) indicates that the assumption of a constant length offset may not have been appropriate.

As expected, log(AADT) was significant and positively associated with crashes in the NB model. In comparison, component 2 of GFMNB model was not found AADT to be statistically significant. A potential reason for the lack of signficance of AADT for component 2 could be that, since the component membership model indicates a significant, positive coefficient for AADT, the posterior probabilities are more likely to yield a larger weight for higher traffic volumes and a lower weight for the smaller values of AADT (see figure 2.2(a)). As a result, the component-specific conditional NB model estimation may not find the log-transformation of traffic volumes as a significant variable. In comparison, in the absence of varying prior weights for FMNB model, the weights associated with component 2 are closer to 0.5 for a large subset of the population (figure 2.2(b)).

A similar effect is also observed in the case of bicycle volumes, wherein the coefficients for component 2 bicycle volumes are deemed to be significant only at the 90% confidence interval, even though the component 2 observes higher peak



((a)) GFMNB Component 2 posterior probability vs Log(AADT)



((b)) FMNB Component 2 posterior probability vs Log(AADT)

Figure 2.2: Comparison of AADT distribution as a function of component 2's posterior probability

hour bicycle traffic on average than in component 1 (see Table 2.6). However, in the case of NB model, the bicycle volumes yields a significant, positive coefficient.

The presence of median is negatively associated with crashes in the NB model as well as 3 out of 4 component models. This finding is consistent with prior studies which have analyzed the improvements in safety associated with constructing medians along undivided arterial segments [28, 20].

Number of lanes are positively associated with number of crashes. An increase in the number of lanes can imply increase in sideswipe collisions.

Since a significant proportion of the segments being analyzed correspond to local streets that cater to residential population and attract less traffic, it is expected to observe this variable to negatively associated with number of crashes across all models.

In the case of the presence of bicycle infrastructure, the NB model yields a positive coefficient, which can imply the presence of heavier bicycle traffic but also greater conflicts/interactions with parked and overtaking vehicles, depending on the type of bicycle facilities installed. It is also possible that within the NB model, the presence of bicycle facilities might be an endogenous variable. In comparison, given the type of the segmentation of the population, the impact of the presence of bicycle infrastructure on the total number number of crashes is not observed to be statistically significant.

The presence of bus lines and stops are indicative of greater mixed traffic conditions which may result in increased complexity of driving/biking, thus contributing towards more collisions [20]. The differences across the components may again be due to the differences in the types of segments within each component. For instance, component 2 of GFMNB indicates a high proportion of segments with bus lines. As a result, presence of bus stops may indicate more stop-and-go conditions and interactions of buses with turning vehicles/cyclists. In comparison, component 1 may be associated with segments where bus volumes may not be consistently high. As a result, the presence of bus lines along a segment may contribute towards the increase in complexity of driving/riding conditions.

Regional transit stations such as Bay Area Rapid Transit (BART) stations, represent centers of high activity which attracts pedestrian traffic [30], taxi pick-ups/drop-offs [34], etc. As a result, their presence is expected to induce increased amounts of chaos within the traffic stream, which may result in the positive association with increased number of crashes.

The models include two types of parking variables. The first set of variables capture the numbers of metered on-street parking and off-street parking spaces along the segment. Parking spot exhibits a positive associated with the number of crashes. Other research has also found on-street parking to be significant when analyzing segment-level crashes and have argued the presence of on-street parking to induce mid-block crossing for pedestrians [23], dooring-related crashes for cyclists[22], and limiting lines of sight near intersections and along horizontal curves[7]. Off-street parking interacts with traffic streams differently, as the behaviour of vehicles emerging from and to such spots is likely to be similar to presence of driveways.

In addition to static variables such as the number of parking spaces, our models also found the dynamic parking variables such as the mean duration of the parked vehicles to be significant and negatively associated with number of crashes. A smaller mean duration implies that the parking space is also associated with a higher number of transactions. As a result, one expects parked vehicles to be moving in and out of the traffic stream, which can lead to an increase in traffic conflicts with other motor vehicles and bicyclists. The difference between single-space and multi-space parking spots is to differentiate between the quantity of transactions as multi-space parking spots handle transactions for more than one parking spot at a time.

The presence of higher speed limits are associated with an increased number of crashes. Within the component specific NB models, component 1 models for FMNB and GFMNB observe a similar trend, whereas component 2 models do not show an significant impact of speed limits of 35 mph or greater. The differential impact of higher speed limits can be attributed to the nature of segments associated with each component. Table 2.5 indicates the average number of collisions along the segments with speed limits greater than or equal to 35 mph. The findings indicate that even though the mean number of collisions in FMNB component 1 are fewer than in component 2, the segments with higher speed limits that belong to component 2 have more collisions than those that belong to component 1. In comparison, while the average number of collisions for the higher speed segments in GFMNB components are similar, segments in component 2 of GFMNB observe twice as many crashes as those in component 1. As a result, the presence of higher speed limits among higher crash segments in general does not lead to an increased impact. Table 2.5 also shows that the AADT for the high speed segments across the various components are similar.

Table 2.5: Differences in segments with speed limits $\geq$ 35 mph across components

Variables	**FMNB**		**GFMNB**	
	Comp. 1	Comp. 2	Comp. 1	Comp. 2
Collisions along segments with speed limits $\geq$ 35 mph (mean)	1.4	0.6	1.0	1.0
AADT along segments with speed limits $\geq$ 35 mph (mean)	17399.0	17926.0	18701.0	16432.0
Mean number of collisions along all segments with posterior probability ¿0.5	0.3	0.7	0.3	0.6

Presence of one-way street segments were negatively associated with crashes. This finding is also consistent with some studies evaluating all collisions in the traffic safety literature. For instance, Eisele et al. [12] found conversion of frontage road segments from two-way to one-way streets in Texas to reduce rear-end and angle collisions. Greibe [14] analyzed collisions urban roads from Danish municipalities and observed one-way streets to have smaller crash rates than two-way streets, although the sample size of one-way streets in the dataset was around only 2%. For specific crash types, such as bicycle and pedestrian crashes, some crash prediction models have observed one-way streets to be associated with an increase in collisions possibly due to the absence of median refuge islands for crossing pedestrians[31] and an increase in parking spaces along one-way streets [22], which have been controlled for in our study.

2.4.2 Intersection-related variables

The presence of signalized intersections along either end of the segments is found to be associated with an increase in the number of crashes. In comparison, the presence of 4-legged intersections is negatively associated with a decrease in collisions. Given that we excluded only crashes associated with the main intersection region and smaller influence around it, the presence of the signalized intersection may contribute towards an increase in rear-end collisions near the edges of the segment [17]. In comparison, while crashes at three-legged intersections are associated with fewer collisions than four-legged intersections because of the presence of fewer conflict points [1], it is possible that the presence of fewer conflict points may lead to increased entry/exit speeds for the adjoining segments, which may adversely affect safety.

2.4.3 Macroscopic variables

The presence of high TNC-related pick-ups/drop-offs in the vicinity of a segment is positively associated with crashes. This finding may be attributed to the associated increase in driving complexity (lane changes, stopping/slow-down) for pick-ups and drop-offs. Alternatively, pick-up and drop-offs also contribute to an increased in traffic in the region which may not be captured within the AADT variable.

The presence of more zero vehicle households may imply the reliance of alternate forms of transportation such as transit, ride-sharing, cycling. As a result, an increased heterogeneity in the traffic mix may explain the positive association with number of crashes.

2.5 Model selection criteria

While the GFMNB model has the best log-likelihood and AIC, it underperforms on BIC which imposes a higher penalty on increased number of parameters. Since the finite mixture modeling approach used in the paper assumed an identical NB model specification within each model, the use of overlapping explanatory variables in the prior class membership model led to some explanatory variables such as AADT, peak hour bicycle volumes not being significant in either one or both model components. Thus, re-estimating the finite mixture models with asymmetric explanatory variables may lead to an improvement in BIC and AIC of the finite mixture models.

2.6 Differences in component characteristics

Based on the differences in the dominant characteristics of each component within FMNB and GFMNB (Table 2.6), we can summarize the model component 2 of FMNB and GFMNB to model relatively high crash presence sites. Sites with higher posterior component 2 probability (> 0.5) are also associated with

Table 2.6: Summary statistics of component characteristics

Variables	FMNB		GFMNB	
	Comp. 1	Comp. 2	Comp. 1	Comp. 2
Total collisions (mean)	0.3	0.7	0.2	0.6
Total collisions (variance)	0.6	1.8	0.6	1.4
Bicycle collisions (mean)	0.06	0.23	0.03	0.19
% freeways	1.3%	1.2%	0.7%	0.2%
% major streets	8.3%	13.4%	7.1%	12.8%
% secondary streets	16.1%	22.0%	10.9%	25.9%
% local streets	74.4%	63.4%	81.3%	59.3%
Peak hour bicycle volumes	167.8	246.8	65.5	340.3
% cycle network	0.2	0.2	0.1	0.3
Number of bus stops (mean)	0.28	0.35	0.17	0.46
Presence of bus line	33.0%	40.0%	13.0%	61.0%
% speed limit ($\geq$ 35 mph)	1.9%	4.4%	2.6%	2.6%
% one-way	16.3%	22.0%	11.7%	25.3%
TNC pick-ups/drop-offs (mean)	513.7	647.7	450.0	671.8
AADT (mean)	4205.4	5053.2	3695.0	5337.7
Population (mean)	1451.7	1512.2	1431.5	1512.6
Employment density (mean)	2777.8	3701.8	1781.4	4551.0
% of zero vehicle households	27.2%	33.9%	20.8%	39.0%
Mean duration of parking meter transactions (multi-space)	65.6	141.1	43.8	137.5
Mean duration of parking meter transactions (single-space)	143.3	286.5	100.0	282.0
Mean segment length	92.5	118.9	112.7	83.5

higher AADT, bicycle volumes, parking activity and TNC pick-ups and drop-offs. These differences are amplified in the case of the GFMNB model since the class membership model skews the posterior probability distributions even further.

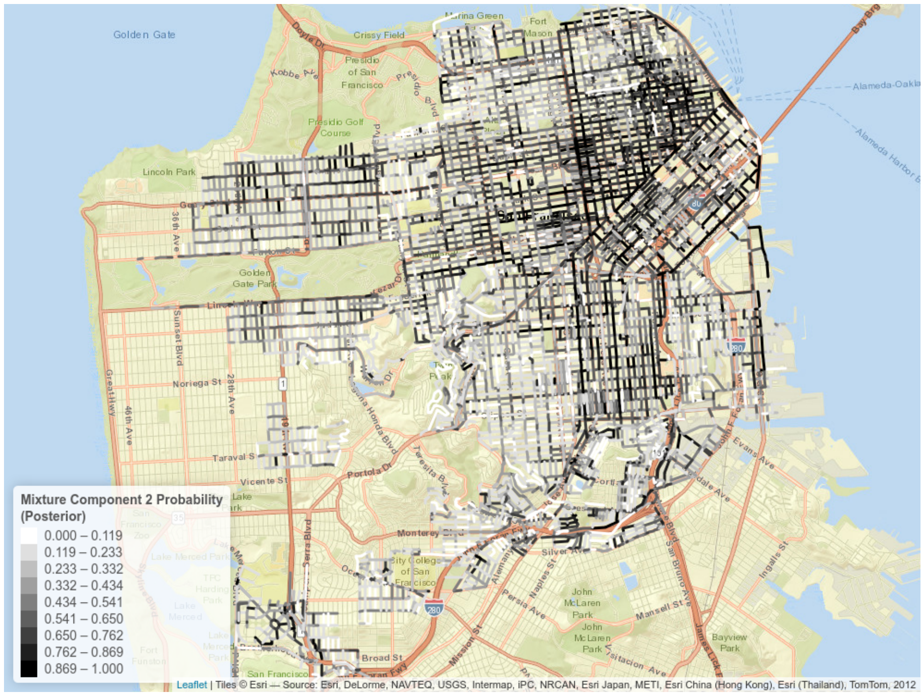


Figure 2.3: GFMNB Component 2 posterior probability

A key point of differentiation for the FMNB and GFMNB components are the distribution of segment lengths. GFMNB component 2 favours shorter blocks due to the inclusion of segment length as variable in the class membership model. As a result, there are significant spatial differences in the sites with higher posterior probabilities for component 2. Figures 2.4 and 2.3 show spatial distribution of the class membership probability. We can see that GFMNB component 2 sites are predominantly located in downtown San Francisco where employment density is the highest, as also revealed in Table 2.6. In comparison, the high probability locations for FMNB component 2 appear to be more spatially dispersed.

2.7 Impact on Network Screening

Figure 2.5 shows the correlation between the EB estimates of the NB, FMNB and GFMNB models. The EB estimates show high degree of correlation, implying that the different models may produce high overlaps when prioritizing sites for investigation when using EB as the ranking criteria. The high correlation

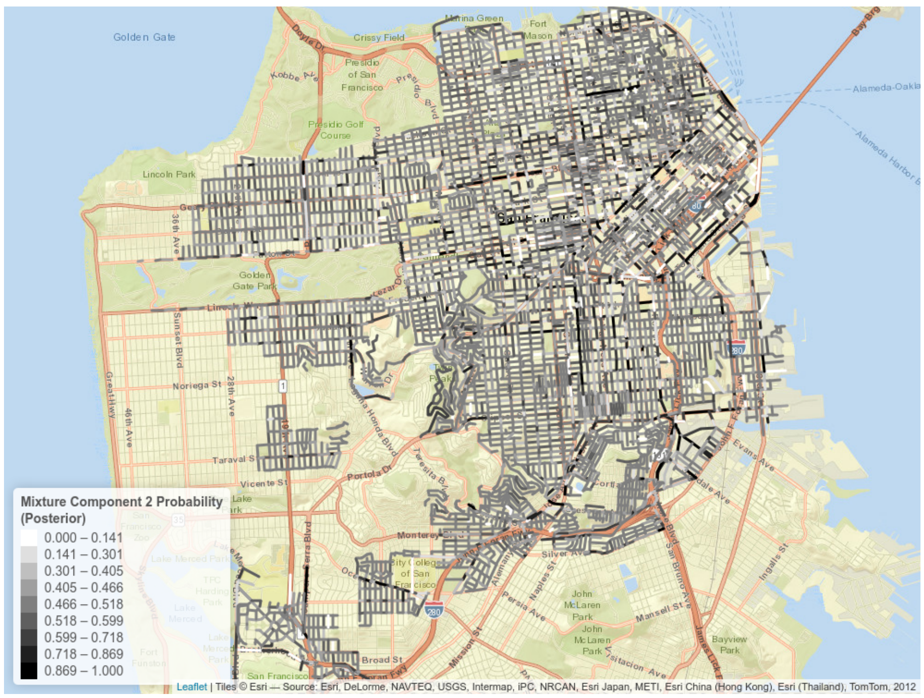


Figure 2.4: FMNB Component 2 posterior probability

can also be attributed to using 5 years of crash data, which reduces the issues associated with regression-to-the-mean.

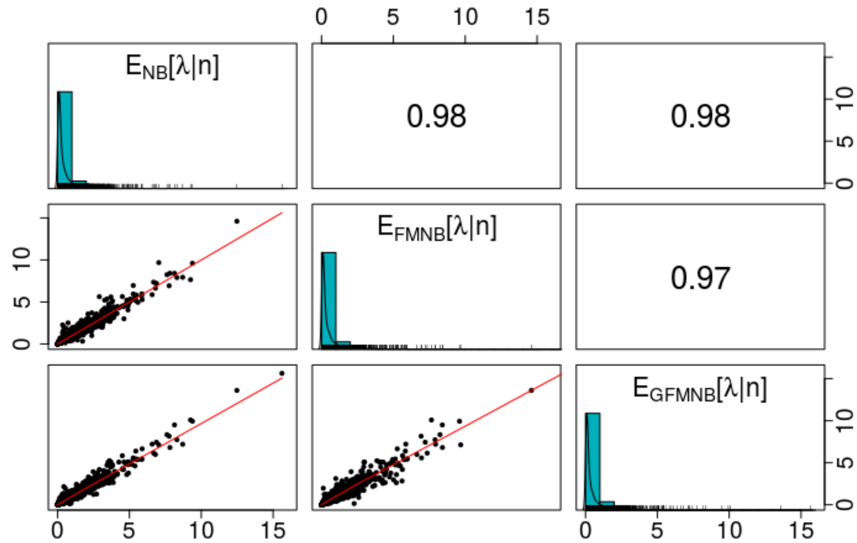


Figure 2.5: Correlation between the EB estimates across models

Figure 2.6 shows the correlation between the PSI estimates of the different models. In comparison to the EB estimates, the PSI estimates demonstrate much lower correlations, especially between GFMNB and other models. In the absence of a ground truth, it is beyond the scope of this paper to ascertain the causes of the underlying differences. However, a simulation study with known ground truths may be useful in providing insights about the implication of GFMNB/FMNB/NB models on network screening, similar to the study conducted by Park, Lord, and Hart [25].

2.8 Sub-segment modeling

2.8.1 Definition of sub-segments

The definition of the spatial unit of observation in our approach is based on intersection influence areas as well as vehicle stopping times. Either due to red phasing of a traffic signal or traffic congestion at the intersection ahead, drivers are forced to decelerate from the desired speed, sometimes come to a stop position and accelerate afterwards. The location at which the deceleration of vehicle starts till the location of the vehicle at the acceleration ends are considered as stretch of intersection influence zone. According to these landmarks, segments

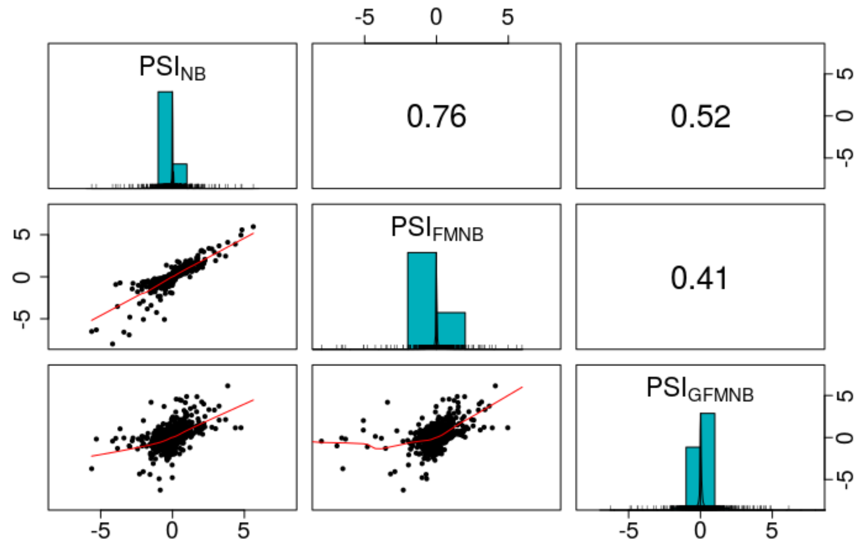


Figure 2.6: Correlation between the PSI estimates across models

are split to in and out parts, as a function of the segment's length and the posted speed limit.

2.8.2 Discussion and Future Work

FMNB and GFMNB models provide a distribution-free method to incorporate unobserved heterogeneity in crash prediction models, by assuming that the underlying crash generation mechanism comprises of different sub-populations. To this end, the differences observed in the component-specific models illustrates the possibility that finite mixture models may capture different safety regimes which can collectively explain the overall crash data. Given the mixed evidence from the model selection criteria, further refinements to the proposed FMNB and GFMNB models are desirable so as to improve the performance of the finite mixture model relative to the NB model. To further extend the finite mixture modeling approach, the use of random parameter models and spatial correlation structures should be explored in future research.

With regards to the variables explored, this study demonstrates the the improved model fit from incorporating dynamic parking data and TNC pick-up/drop-off information in segment-level traffic safety assessment. The results also reveal that when designing for curb usage in urban areas, safety considerations must be considered in addition to competing mobility needs of different road users. Since the results of the estimated models do not imply any causality in the variable considered, future studies must be consider more detailed safety assessments of curbside infrastructure such as temporary pick-up/drop-off loading zones, no-parking zones when compared to segments with on-street parking. Similarly, as cities explore flexible pricing mechanisms to maximize parking available or reduce cruising [8], the impact of the frequent vehicle turnovers on safety needs to be investigated further. The study also does not provide any definitive assessment of the impact of TNC pick-ups/drop-offs on safety. Depending on the substitution patterns of modes used, an increase in TNC activity can lead to an increase in traffic volumes, which in turn can increase the expected number of crashes. However, the dynamics of passenger loading and pick-up may be different from those of curbside parking, and needs to be explored further.

Lastly, in order to consider the diverse set of variables considered in this study, some assumptions were made which may not be realistic. The temporal aggregation of the crash data was driven by the absence of annual estimates of traffic counts, infrastructure installation, and TNC data. The temporal variation of TNC activity would require more detailed data collection efforts and likely some collaboration with TNC companies. Similarly, in order to analyze the influence of the vehicle turnover at metered parking spots, parking transaction data prior to 2018 was not available. Moving forward, it would be beneficial to collect this data for a larger time period so as to analyze the influence of parking occupancy and turnover at a finer, more consistent temporal resolution.

Bibliography

[1] Mohamed Abdel-Aty and Xuesong Wang. "Crash estimation at signalized intersections along corridors: analyzing spatial effect and identifying significant factors". In: *Transportation Research Record* 1953.1 (2006), pp. 98–111.

[2] Saif A. Alarifi et al. "Crash modeling for intersections and segments along corridors: A Bayesian multilevel joint model with random parameters". In: *Analytic Methods in Accident Research* 16 (2017), pp. 48–59. ISSN: 22136657. DOI: 10.1016/j.amar.2017.08.002. URL: https://doi.org/10.1016/j.amar.2017.08.002.

[3] Panagiotis Ch Anastasopoulos and Fred L. Mannering. "A note on modeling vehicle accident frequencies with random-parameters count models". In: *Accident Analysis and Prevention* 41.1 (2009), pp. 153–159. ISSN: 00014575. DOI: 10.1016/j.aap.2008.10.005.

[4] Panagiotis Ch Anastasopoulos et al. "A study of factors affecting highway accident rates using the random-parameters tobit model". In: *Accident Analysis and Prevention* 45 (2012), pp. 628–633. ISSN: 00014575. DOI: 10.1016/j.aap.2011.09.015. URL: http://dx.doi.org/10.1016/j.aap.2011.09.015.

[5] San Francisco County Transportation Authority. *TNCs Today: A Profile of San Francisco Transportation, Transportation Network Company Activity*. Tech. rep. URL: https://www.sfcta.org/sites/default/files/2019-02/TNCs_Today_112917_0.pdf.

[6] Sumit Bindra, John N. Ivan, and Thomas Jonsson. "Predicting Segment-Intersection Crashes with Land Development Data". In: *Transportation Research Record: Journal of the Transportation Research Board* 2102.1 (2009), pp. 9–17. ISSN: 0361-1981. DOI: 10.3141/2102-02.

[7] Kiante Bush and Celeste Chavis. *Safety analysis of on-street parking on an urban principal arterial*. Tech. rep. 2017.

[8] Daniel G Chatman and Michael Manville. "Theory versus implementation in congestion-priced parking: An evaluation of SFpark, 2011–2012". In: *Research in Transportation Economics* 44 (2014), pp. 52–60.

[9] Wayne D Cottrell and Sichun Mu. "Recurrent Crash Sites : Identification , Issues and Factors". In: *University of Utah* December (2005).

[10] Abhishek Das and Mohamed A. Abdel-Aty. "A combined frequency-severity approach for the analysis of rear-end crashes on urban arterials". In: *Safety Science* 49.8-9 (2011), pp. 1156–1163. ISSN: 09257535. DOI: 10.1016/j.ssci.2011.03.007.

[11] Partha Deb and Pravin K. Trivedi. "The structure of demand for health care: Latent class versus two-part models". In: *Journal of health economics* (2002). ISSN: 01676296. DOI: 10.1016/S0167-6296(02)00008-5.

[12] William L Eisele et al. "Safety and economic impacts of converting two-way frontage roads to one-way operation". In: *Transportation research record* 2301.1 (2012), pp. 66–75.

[13] Sylvia Frühwirth-Schnatter. *Finite Mixture and Markov Switching Models.* 2006. DOI: 10.1007/978-0-387-35768-3.

[14] Poul Greibe. "Accident prediction models for urban roads". In: *Accident Analysis & Prevention* 35.2 (2003), pp. 273–285.

[15] Bettina Grun and Friedrich Leisch. "FlexMix version 2: finite mixtures with concomitant variables and varying and constant parameters". In: (2008).

[16] Qinzhong Hou et al. "Application of a random effects negative binomial model to examine crash frequency for freeways in China". In: *Physica A: Statistical Mechanics and its Applications* 509.2018 (2018), pp. 937–944. ISSN: 03784371. DOI: 10.1016/j.physa.2018.06.088.

[17] Do-Gyeong Kim et al. "Modeling crash outcome probabilities at rural intersections: Application of hierarchical binomial logistic models". In: *Accident Analysis & Prevention* 39.1 (2007), pp. 125–134.

[18] Jaeyoung Lee, Mohamed Abdel-Aty, and Qing Cai. "Intersection crash prediction modeling with macro-level data from various geographic units". In: *Accident Analysis & Prevention* 102 (2017), pp. 213–226.

[19] Dominique Lord and Fred Mannering. "The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives". In: *Transportation Research Part A: Policy and Practice* 44.5 (2010), pp. 291–305. ISSN: 0965-8564. DOI: https://doi.org/10.1016/j.tra.2010.02.001. URL: http://www.sciencedirect.com/science/article/pii/S0965856410000376.

[20] Ming Ma et al. "Safety analysis of urban arterials under mixed-traffic patterns in Beijing". In: *Transportation Research Record* 2193.1 (2010), pp. 105–115.

[21] Maren L Outwater and Billy Charlton. "The San Francisco Model in Practice". In: *Innovations in Travel Demand Modeling* 24 (2006).

[22] Chih-Wei Pai. "Overtaking, rear-end, and door crashes involving bicycles: An empirical investigation". In: *Accident Analysis & Prevention* 43.3 (2011), pp. 1228–1235.

[23] Anurag Pande and Mohamed Abdel-Aty. "A novel approach for analyzing severe crash patterns on multilane highways". In: *Accident Analysis & Prevention* 41.5 (2009), pp. 985–994.

[24] Byung Jung Park, Dominique Lord, and Chungwon Lee. "Finite mixture modeling for vehicle crash data with application to hotspot identification". In: *Accident Analysis and Prevention* 71 (2014), pp. 319–326. ISSN: 00014575. DOI: 10.1016/j.aap.2014.05.030. URL: http://dx.doi.org/10.1016/j.aap.2014.05.030.

[25] Byung-Jung Park, Dominique Lord, and Jeffrey D Hart. "Bias properties of Bayesian statistics in finite mixture of negative binomial regression models in crash data analysis". In: *Accident Analysis & Prevention* 42.2 (2010), pp. 741–749.

[26] Juneyoung Park et al. "Developing crash modification functions to assess safety effects of adding bike lanes for urban arterials with different roadway and socio-economic characteristics". In: *Accident Analysis and Prevention* 74 (2015), pp. 179–191. ISSN: 00014575. DOI: 10.1016/j.aap.2014.10.024. URL: http://dx.doi.org/10.1016/j.aap.2014.10.024.

[27] Kevin Ramsey and Alexander Bell. "Smart Location Database". In: (2014).

[28] Ziad Sawalha and Tarek Sayed. "Evaluating Safety of Urban Arterial Roadways". In: *Journal of Transportation Engineering-asce - J TRANSP ENG-ASCE* 127 (2001). DOI: 10.1061/(ASCE)0733-947X(2001)127:2(151).

[29] B. Schaller. *UNSUSTAINABLE? The Growth of App-Based Ride Services and Traffic, Travel and the Future of New York City.* Tech. rep. Schaller Consulting, 2017.

[30] Robert J Schneider, Lindsay S Arnold, and David R Ragland. "Pilot model for estimating pedestrian intersection crossing volumes". In: *Transportation research record* 2140.1 (2009), pp. 13–26.

[31] Satish Ukkusuri, Samiul Hasan, and HM Abdul Aziz. "Random parameter model used to explain effects of built-environment characteristics on pedestrian crash frequency". In: *Transportation research record* 2237.1 (2011), pp. 98–106.

[32] Xuesong Wang et al. "Investigation of Safety Influence Area for Four-Legged Signalized Intersections Nationwide Survey and Empirical Inquiry CRASH IDENTIFICATION AND CODING". In: (), pp. 86–95. DOI: 10.3141/2083-10.

[33] Yingge Xiong and Fred L. Mannering. "The heterogeneous effects of guardian supervision on adolescent driver-injury severities: A finite-mixture random-parameters approach". In: *Transportation Research Part B: Methodological* (2013). ISSN: 01912615. DOI: 10.1016/j.trb.2013.01.002.

[34] Zhuo Yang et al. “Analysis of Washington, DC taxi demand using GPS and land-use data”. In: *Journal of Transport Geography* 66 (2018), pp. 35–44.

[35] Andrew James Zalewski, Stephen M Buckley, and Rachel R Weinberger. *Regulating curb space: developing a framework to understand and improve curbside management*. Tech. rep. 2012.

[36] Yajie Zou, Yunlong Zhang, and Dominique Lord. “Application of finite mixture of negative binomial regression models with varying weight parameters for vehicle crash data analysis”. In: *Accident Analysis and Prevention* 50 (2013), pp. 1042–1051. ISSN: 00014575. DOI: `10.1016/j.aap.2012.08.004`. URL: `http://dx.doi.org/10.1016/j.aap.2012.08.004`.

[37] Yajie Zou et al. “Empirical Bayes estimates of finite mixture of negative binomial regression models and its application to highway safety”. In: *Journal of Applied Statistics* 45.9 (2018), pp. 1652–1669.