Exploring the Impact of Truck Traffic on Road Segment-based Severe Crash Proportion using Extensive Statewide Weigh-In-Motion Data

Authors : Chuan Xu, Kaan Ozbay, Hongling Liu, Kun Xie, Di Yang

Presenter : Kun Xie

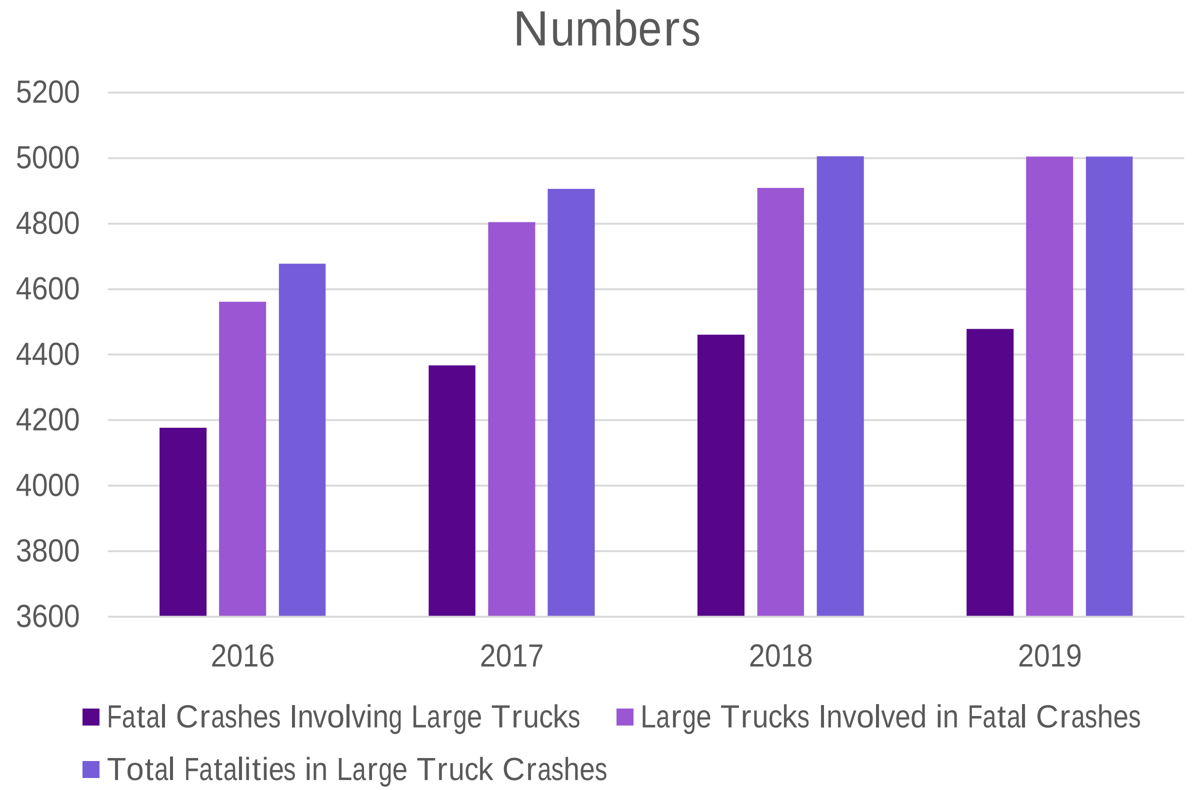
June 8, 2022

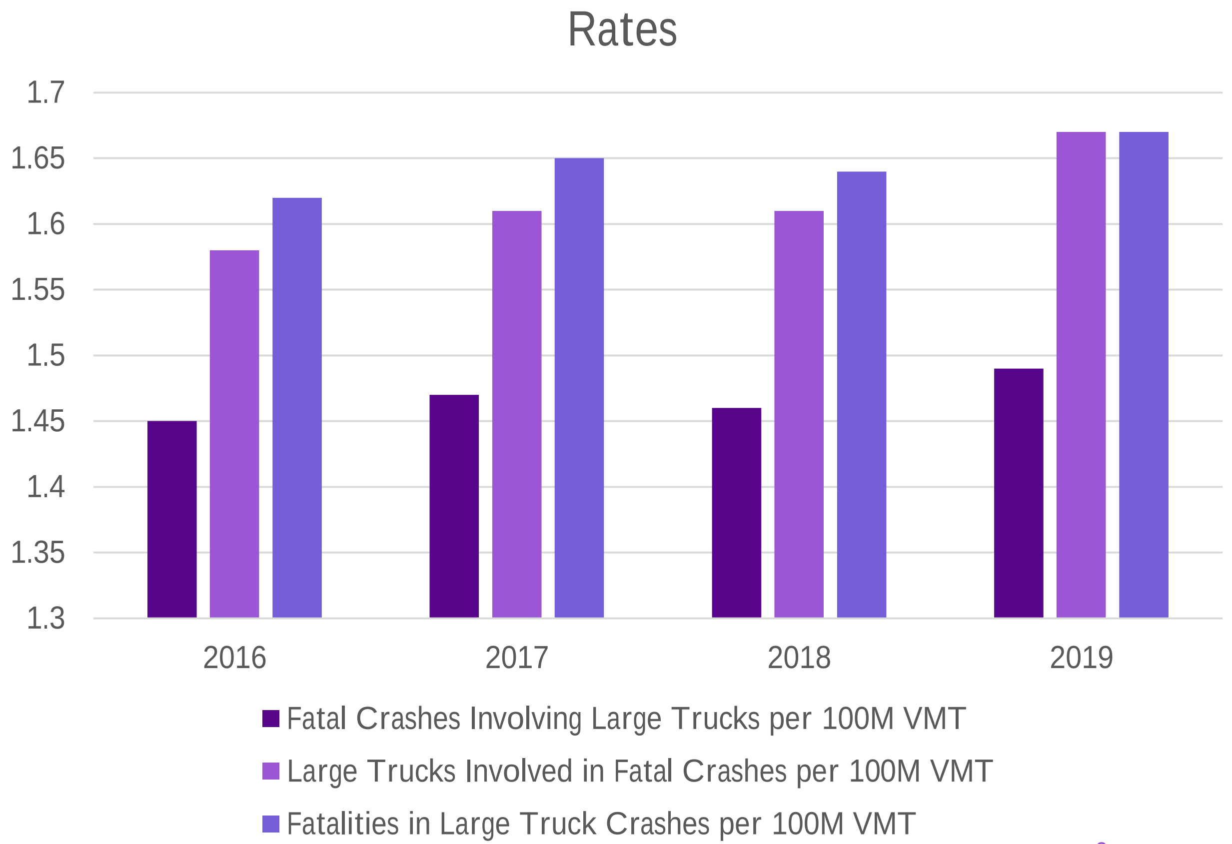
Content

- Background
  - The Trend of Large Truck related Fatal Crashes Indexes
  - Large Truck vs. Passenger Vehicles
  - Crash Proportions by Severity Levels in HSM
- Literature
- Motivation & Research Question
- Data Preparation
  - WIM Data
  - Road Feature and Crash Data
  - Data Processing and Fusion Steps
  - Descriptive Statistics of the Variables
- Methodology
  - Fractional Regression Model
  - Model Selection
- Results & Discussion
  - Specification Tests for One-part and Two-part Models
  - Average Partial Effects of The Fractional Regression Model
- Conclusions

Background: The Trend of Large Truck related Fatal Crashes Indexes

- Truck-related crashes bring serious consequences to society, especially for large truck.
- From 2016 to 2019, the trends large-truck-related fatal crash indexing went up.



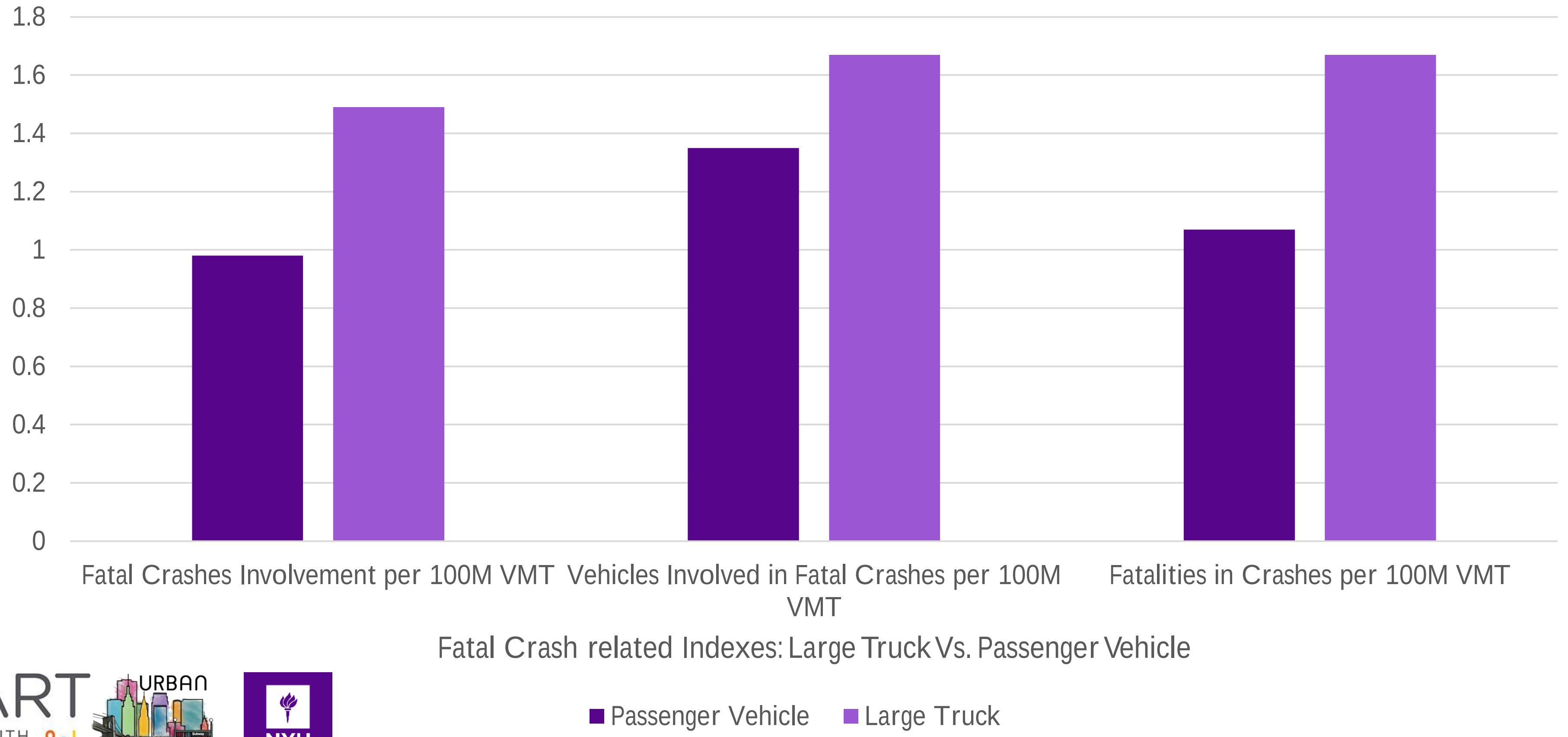


C2SMART CONNECTED CITIES WITH SMART TRANSPORTATION
URBAN MITS Lab@NYU
NYU

Data Source: FMCSA, Large Truck and Bus Crash Facts 2019

Background: Large Truck vs. Passenger Vehicles

- In 2019, compared to a passenger vehicle, large trucks fatal crashes involvement per 100 million VMT is 52% higher, the vehicles involved in fatal crashes per 100 million VMT for large truck is 24% higher, the number of fatalities in large truck crashes per 100 million VMT is 56% higher.



Data Source: FMCSA, Large Truck and Bus Crash Facts 2019

Background: Crash Proportions by Severity Levels in HSM

- HSM proposed a method to calculate the expected average crash frequency by crash severity for a certain roadway facility type.
- Safety performance function, crash modification factors, calibration factor total expected average crash frequency
- HSM also provides default proportions by crash severity levels for different roadway facility types based on the Highway Safety Information System (HSIS) data from some states (such as Washington 2002-2006, California 2002-2006).

Table 10-3. Default Distribution for Crash Severity Level on Rural Two-Lane, Two-Way Roadway Segments

Crash Severity Level	Percentage of Total Roadway Segment Crashes[a]
Fatal	1.3
Incapacitating Injury	5.4
Nonincapacitating injury	10.9
Possible injury	14.5
Total fatal plus injury	32.1
Property damage only	67.9
Total	100.0

[a] Based on HSIS data for Washington (2002–2006)

Literature

- Most crash severity studies are based on disaggregated crash data
  - These studies can help to identify hazard factors to truck-involved crash severity and provide insights mitigating its effect
  - But can't be used in the HSM crash prediction framework
- Some studies are based on data aggregated by entities
  - Crash proportion by severity levels for multilane highway segments - Yasmin et al. (2016)
  - Rural two-lane, two-way intersections - Qin et al. (2019)
  - Census tract - Xie et al. (2019)
  - Urban and suburban intersections - Wang et al. (2021)
  - But few studies examined the truck traffic impact for a roadway segment.

C2SMART CONNECTED CITIES WITH SMART TRANSPORTATION URBAN MITS Lab@NYU NYU

Literature

- Vehicle Weight and Traffic Safety
  - Post-crash observed vehicle weight
    - The weight of vehicles involved in crashes
    - Curb weight - Wang and Kockelman (2005)
    - GVWR - Lemp et al. (2011)
    - Trucks hauling a trailer with heavy cargo (>20,000 kg) - Zhu and Srinivasan (2011)
    - Weight difference and weight ratio between vehicles involving rear-end crashes - Yuan et al. (2017)
    - Mean truck weight, maximum and minimum weight (in multi-vehicle crashes), truck difference between the maximum truck weight and the minimum truck weight, registered-weight - Zou et al. (2017)
    - <u>Can’t be used in HSM crash prediction framework</u>

Literature

- Vehicle Weight and Traffic Safety
  - Traffic Vehicle Weight
    - is firstly observed for protecting infrastructure purpose
    - WIM technology is an uninterrupted way to observe traffic vehicle weight in the traffic.
      - used WIM data to investigate the safety performance of driver behavior - Karim (2014)
      - estimate heavy vehicle-involved rear-end crash potential - Jo et al. (2019)
      - created novel risk indicators considering vehicle weight for the diagnosis of traffic conflict risk map - Wang et al. (2021)
      - <u>But we haven’t found any literature studying the impact of vehicle weight to severe crash proportion</u>

Motivation & Research Question

- Motivation
  - No Truck Information Involved: the default proportions by crash severity levels in HSM are static and there's no difference between two road segments with totally different truck traffic
  - Data Aggregation: most safety studies related to truck and crash severity are crash-based and only a few studies are facilities-based (such as road segment, intersection)
  - Truck Weight Information from Extensive WIM Data: we have 5-year 88 WIM station data that the truck weight was included. (As far as we know, no existing studies used this amount of WIM data)
- Research Question: Do truck traffic variables (including truck weight) have impacts on road segment-based severe crash proportion?

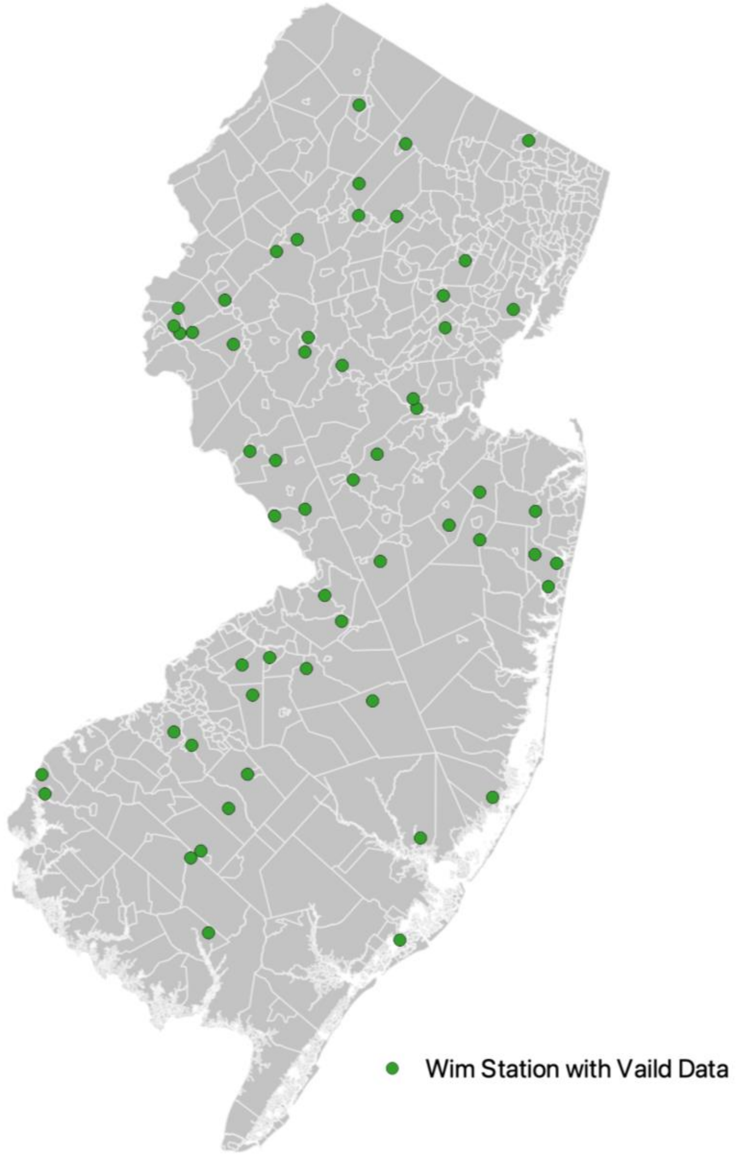
- This study seeks to explore the impact of truck traffic on road segment-based severe crash proportion under the predictive method framework proposed by HSM

C2SMART CONNECTED CITIES WITH SMART TRANSPORTATION URBAN MITS Lab@NYU NYU

Data Source: FMCSA, Large Truck and Bus Crash Facts 2019

Data Preparation: WIM Data

- WIM devices: capture and record axle weights and total vehicle weights as vehicles pass a measurement site
- Available Features:
  - Vehicle Classification
  - Vehicle Weight
  - Traffic Volume
- Data Period: 2011-2015, 5-year data
- 88 WIM stations deployed by NJDOT, 61 WIM stations have individual vehicle data for both directions
- Remove missing data record, and 165 records with zero crash frequency
- Finally, 339 observations



C2SMART CONNECTED CITIES WITH SMART TRANSPORTATION

URBAN MITS Lab@NYU

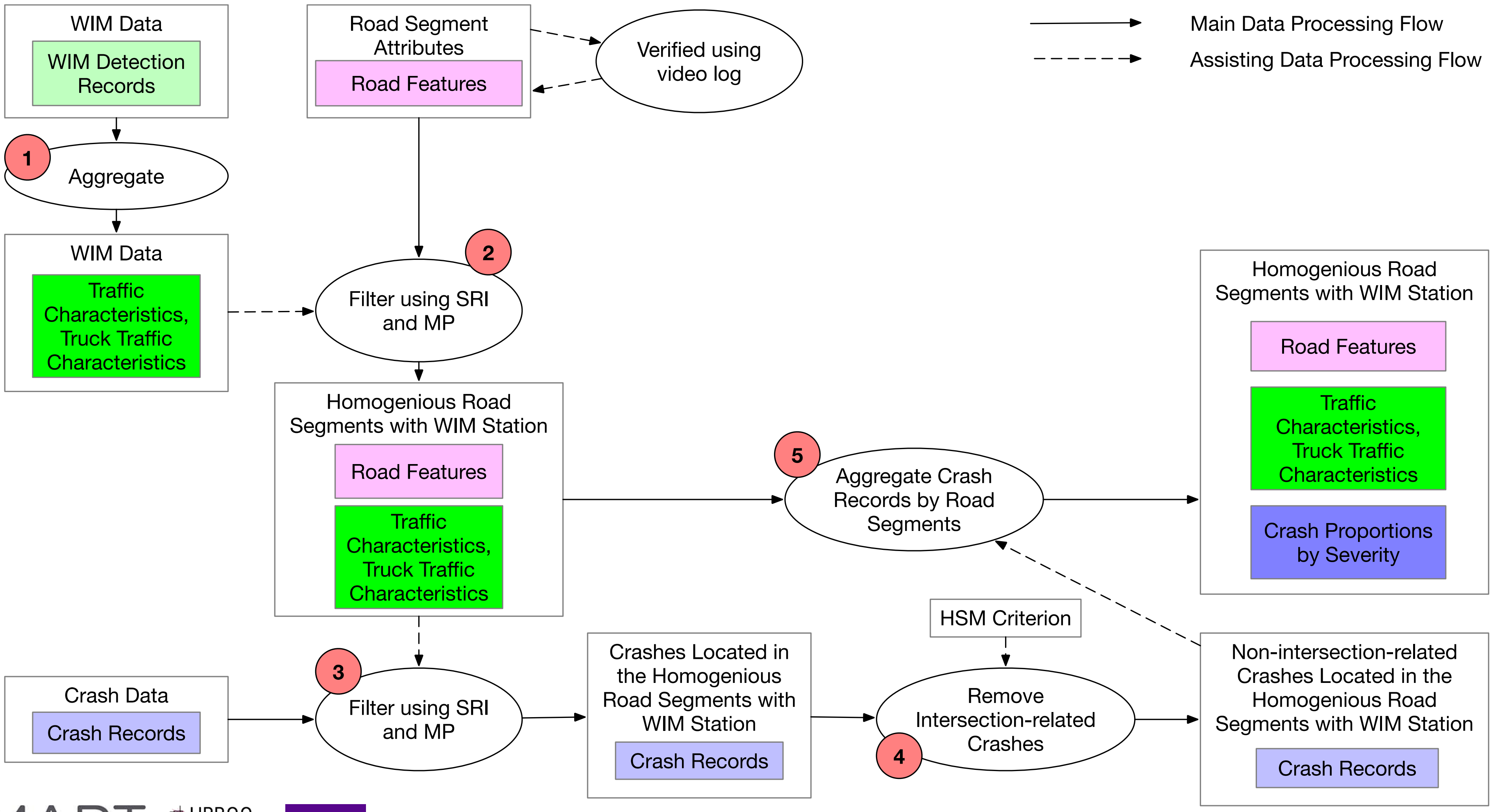
NYU

Data Preparation: Road Feature and Crash Data

- Road Feature Data
  - Data Source: Straight Line Diagrams (SLD) for New Jersey (*from NJ SPF Project)
  - Used Features: segment length, rural or urban, divided or undivided, number of lanes, shoulder width, posted speed limit
  - Aggregation Unit: Homogeneous road segment with a WIM station
- Crash Data
  - Data Source: Voyager Safety Database (*from NJ SPF Project)
  - Period: 2011-2015
  - Crash Attributes: Crash location, crash occurrence time, crash severity were extracted
  - Crash Severity: coded as fatality, injury, and property damage only
  - Remove Intersection-related Crashes: defined as crashes that occur at the intersection itself or crashes that occur on an intersection approach within 250 ft (HSM)

*Data Source: Ozbay, K., H. Nassif, B. Bartin, C. Xu, and A. Bhattacharyya. Calibration/development of safety performance functions for new jersey. 2019. https://rosap.ntl.bts.gov/view/dot/56146

Data Preparation: Data Processing and Fusion Steps



C2SMART CONNECTED CITIES WITH SMART TRANSPORTATION

URBAN MITS Lab @ NYU

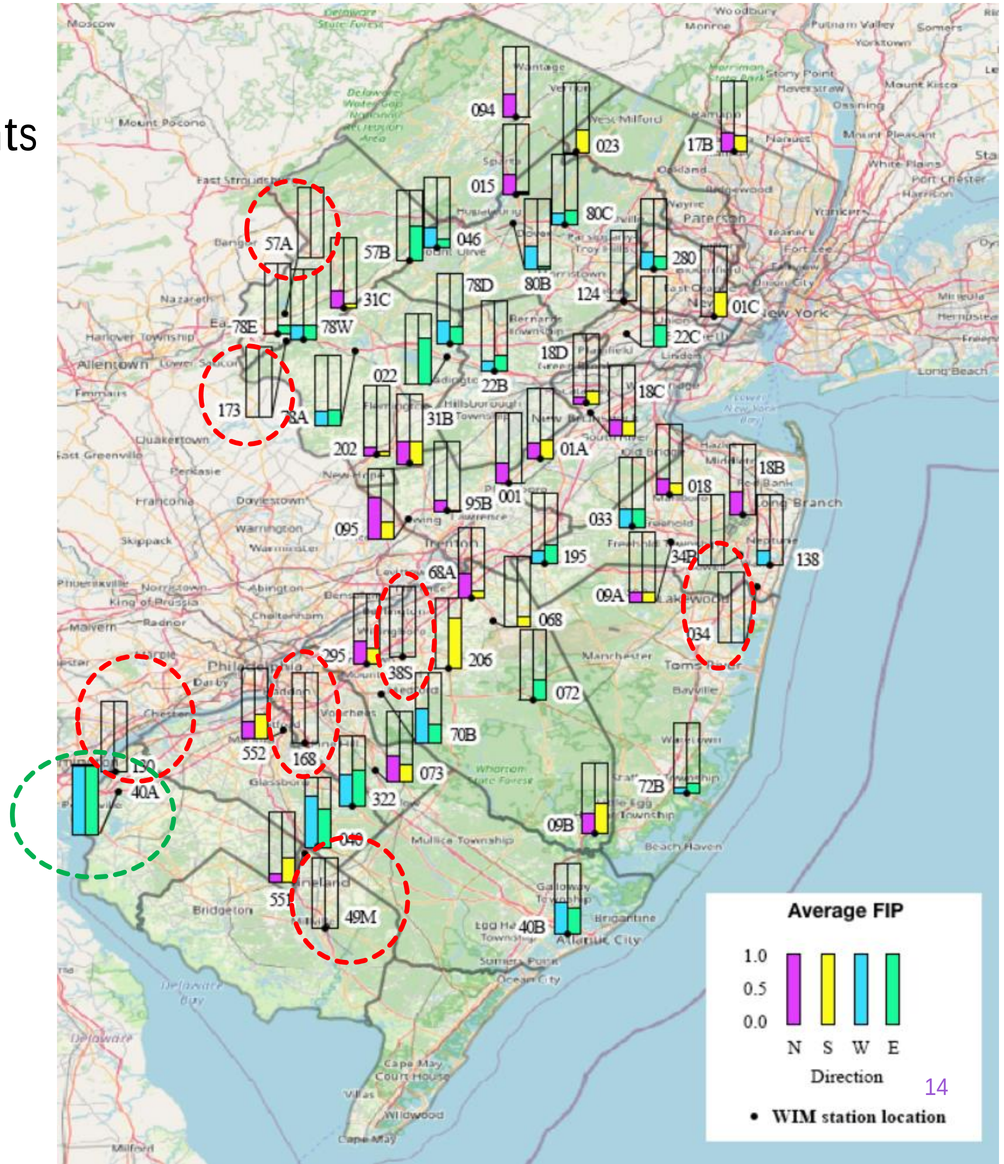
NYU

Data Preparation: Descriptive Statistics of the Variables

Type	Variable	Description	Mean	S.D.	Min	Max
Response Variable	FIP	Fatal & injury crash proportion	0.25	0.25	0.00	1.00
Road Features	M_d	Median type: 1-The road segment is divided by a physical median; otherwise, 0.	0.67	0.47	0.00	1.00
	N_{lane}	Number of lanes	2.44	0.72	2.00	5.00
	L_{ru}	Location: 1- urban, 0- rural.	0.81	0.39	0.00	1.00
	S_{limit}	Posted speed limit (mph)	55.00	7.20	40.00	65.00
	S_{Width}	The width of road segment shoulder in feet	10.33	3.08	0.00	18.00
	S_{length}	Segment length (mile)	0.99	1.01	0.10	4.75
Truck Traffic Characteristics	AADTT	Annual average daily truck traffic (vehicle/day)	1143	1672	15	8442
	P_{tt}	Truck traffic proportion	0.05	0.03	0.00	0.19
	SD_{tw}	SD of truck weight (kips)	20.16	4.66	9.49	44.94
	M_{tw}	Mean of truck weight (kips)	34.73	7.17	19.09	55.27
	N_{50Kips}	The count of truck over 50 kips (10^4)	9.89	22.59	0.00	145.41
Traffic Characteristics	AADT	Annual average daily traffic (vehicle/day)	20,023	16,809	2,202	78,696

Data Preparation

- Average Fatal and Injury Proportion
  - The average FIP varies greatly among different road segments
  - Several FIPs are zero or close to zero
  - FIPs of a road segment (both direction) are 1



C2SMART CONNECTED CITIES WITH SMART TRANSPORTATION

URBAN MITS Lab@NYU

NYU

Methodology: Fractional Regression Model

- Fractional Regression Model (FRM): Let y_{it} denote the fractional response variable, to be explained for the FIP in WIM station i, $i = 1, \dots N$ at time t, $t = 1, \dots T$, and let x_{it} denote a k-vector of explanatory variables. The standard factional regression model requires the assumption of a functional form for y that imposes the desired constraints on the conditional mean of the dependent variable:

$$E(y_{it}|\, x_{it}) = G(x_{it}^T\theta)$$

where $\boldsymbol{\theta}$ is the vector of parameters of interest and $G(\cdot)$ is a (nonlinear) function satisfying $0 < G(\cdot) < 1$.

- Some popular link function choices are Cauchit, Logistic(logit), Standard normal(probit), Extreme maximum (loglog), Extreme minimum (cloglog).
- The link functions are used to map $x_{it}^T\theta$ from $(-\infty,+\infty)$ to (0,1)
- FRM can be structured as one-part model or two-part model.
  - One-part model: all data in one model
  - Two-part model: 0 and non-zero y_{it} in the first part, and non-zero y_{it} only in the second part

C2SMART CONNECTED CITIES WITH SMART TRANSPORTATION URBAN MITS Lab@NYU NYU

Methodology: Model Selection

- Link function Selection
  - Goodness-of-Functional From(GOFF): GOFF1, GOFF2, Generalized GOFF
  - Regression Specification Error Test (RESET)
- One-part or Two-part models?
  - P test: test one-part models against two-part models

	Model designation	Distribution function	$G(x\theta)$
Symmetrical	Cauchit	Cauchy	$\frac{1}{2} + \frac{1}{\pi}\arctan(x\theta)$
	Logit	Logistic	$\frac{e^{x\theta}}{1+e^{x\theta}}$
	Probit	Standard normal	$\Phi(x\theta)$
Asymmetrical	Loglog	Extreme maximum	$e^{-e^{-x\theta}}$
	Complementary loglog	Extreme minimum	$1 - e^{-e^{x\theta}}$

Increases sharply at small values of G(·) and slowly when G(·) is near 1

Increases slowly at small values of G(·) and sharply when G(·) is near 1

C2SMART CONNECTED CITIES WITH SMART TRANSPORTATION URBAN MITS Lab@NYU NYU

Results: Specification Tests for One-part and Two-part Models

Table 2: Specification tests for one-part and two-part models

	Logit	Probit	Cauchit	Loglog	Cloglog
One-part Models (n=339)					
RESET test	0.820 (0.664)	1.221(0.543)	0.236(0.889)	2.090(0.352)	0.527(0.768)
GOFF-I test	0.309(0.578)	0.245(0.620)	0.120(0.729)	-	0.217(0.641)
GOFF-II test	0.336(0.562)	0.191(0.662)	0.001(0.971)	0.116(0.733)	-
GGOFF test	0.363(0.834)	0.834(0.659)	2.387(0.303)	-	-
P test					
H1: Logit	-	0.806(0.421)	0.053(0.958)	0.955(0.340)	-0.292(0.770)
H1: Probit	-0.608(0.544)	-	0.155(0.877)	0.858(0.392)	-0.332 (0.740)
H1: Cauchit	1.251(0.212)	1.331(0.184)	-	1.449(0.148)	1.281(0.201)
H1: Loglog	-0.394(0.694)	-0.492(0.623)	0.372(0.710)	-	-0.201(0.841)
H1: Cloglog	0.548(0.584)	0.756(0.450)	-0.158(0.875)	0.981(0.327)	-
Two-part Models – First Part (n=339)					
RESET test	3.988(0.136)	2.845(0.241)	**15.200(0.001)**	4.737(0.094)	2.471(0.291)
GOFF-I test	2.083(0.149)	1.658(0.198)	0.002(0.961)	-	2.161(0.142)
GOFF-II test	1.213(0.271)	2.131(0.144)	**4.565(0.033)**	0.360(0.548)	-
GGOFF test	3.711(0.156)	2.809(0.245)	**11.857(0.003)**	-	-
P test					
H1: Logit	-	1.055(0.292)	1.159(0.247)	1.694(0.091)	-0.671(0.503)
H1: Probit	0.038(0.970)	-	1.204(0.229)	1.687(0.092)	-1.148(0.252)
H1: Cauchit	**3.550(0.000)**	**3.668(0.000)**	-	**3.635(0.000)**	**3.646(0.000)**
H1: Loglog	-0.886(0.376)	-0.449(0.654)	0.726(0.468)	-	-0.245(0.806)
H1: Cloglog	1.792(0.074)	1.850(0.065)	1.677(0.094)	**2.112(0.035)**	-
Two-part Models – Second Part (n=242)					
RESET test	4.893(0.087)	5.018(0.081)	4.085(0.130)	0.568(0.062)	4.225(0.121)
GOFF-I test	**4.216(0.040)**	**4.226(0.040)**	2.284(0.131)	-	2.873(0.09)
GOFF-II test	3.751(0.053)	**4.524(0.033)**	1.707(0.191)	**5.409(0.020)**	-
GGOFF test	4.849(0.089)	5.009(0.082)	3.781(0.151)	-	-
P test					
H1: Logit	-	**2.167(0.031)**	-1.600(0.111)	**2.355(0.019)**	-1.426(0.155)
H1: Probit	**-2.090(0.038)**	-	-1.569(0.118)	**2.296(0.023)**	-1.561(0.120)
H1: Cauchit	**2.230(0.027)**	**2.282(0.023)**	-	**2.454(0.015)**	1.493(0.137)
H1: Loglog	-1.954(0.052)	**-1.968(0.050)**	-1.456(0.147)	-	-1.563(0.119)
H1: Cloglog	1.732(0.085)	1.946(0.053)	-0.845(0.399)	**2.279(0.024)**	-

Note: test statistics are followed by p-value in the brackets, bolded numbers denote that test statistics are significant in 5%; GOFF-I, GOFF-II, GGOFF, and RESET tests use Lagrange Multiplier; P test uses Wald method.

All link function forms are also not rejected by GGOFF tests, RESET tests and P tests

loglog (increase sharply at small values of G(.) and slowly when G(.) is near 1)

Asymmetric distribution function
P-test show that cloglog can be better than loglog

Only Cauchit specification is insignificant at 0.1 level

Results: Average Partial Effects of The Fractional Regression Model

	One-part model (Loglog)		Two-part model (Cloglog + Cauchit)			
			First part		Second part	
Variable	Coefficient	Partial Effect	Coefficient	Partial Effect	Coefficient	Partial Effect
Intercept	-0.571(0.014)**	-	-1.710(0.002)**	-	0.282(0.473)	
M_d	-0.364(0.001)**	-0.123(0.001)**	-0.174(0.392)	-0.057(0.391)	-0.580(0.000)**	-0.132(0.000)**
S_{Width}	0.047(0.007)**	0.016(0.007) **	0.123(0.003)**	0.040(0.002) **	0.011(0.698)	0.003(0.697)
N_{50Kips}	-0.005(0.060)*	-0.002(0.059)*	0.004(0.444)	0.001(0.443)	-0.013(0.007)**	-0.003(0.007)**
SD_{tw}	-0.012(0.477)	-0.004(0.477)	-0.039(0.208)	-0.013(0.205)	-0.008(0.782)	-0.002(0.782)
M_{tw}	0.001(0.914)	0.001(0.914)	0.049(0.062)*	0.016(0.058)*	-0.021(0.317)	-0.005(0.315)
P_{tt}	5.098(0.046)**	1.722(0.045)**	-4.587(0.278)	-1.495(0.275)	11.480(0.001)**	2.611(0.001)**
MAE	0.174		0.175			

- M_d : negative
- M_{tw}: positive, in the first part of the two-part model, the average truck weight is higher, the FIP of that road segment is more likely to be non-zero.
- P_{tt}: positive, in the one-part model and the second part of the two-part model.
- N_{50Kips}: negative;
- S_{width}: positive

C2SMART CONNECTED CITIES WITH SMART TRANSPORTATION

URBAN MITS Lab@NYU

NYU

Results & Discussion

- To account for the bounded nature of FIP, one-part and two-part Fractional Regression Models (FRMs) are developed:
  - For the one-part FRM, loglog link function is favored
  - For the two-part FRM, the cloglog and Cauchit link functions are preferred for the first and second parts respectively
  - The mean absolute error indicates that the one-part FRM is slightly better in prediction accuracy than the two-part FRM, P tests suggest insignificant performance difference between these two models.

- The mean of truck weight are statistically significant and positively related to FIP
- Truck traffic proportion are statistically significant and positively related to FIP
- The FIPs of road segments divided by physical median are found to be lower than those of undivided roads.
- Road segments with wider shoulders are associated with higher FIPs
- N_{50kips} has negative and significant coefficients in both the one-part model and the second part of the two-part model.
  - This is possibly because trucks with weights heavier than 50 kips may require the drivers have additional training and experience that enable them to maneuver these large and heavy vehicles safely.
  - In addition, vehicles around these large-sized trucks may also pay more attention to crash avoidance which reduces the crash involvement of these trucks and so the FIP was observed to be reduced.

Conclusions, Limitation & Future Study

- The mean of truck weight and truck traffic proportion are statistically significant and positively related to FIP.
- No significant association was found between FIP and truck weight variance.
- AADT, Segment Length were not significant in FRM models.
- The FIPs of road segments divided by physical median are found to be lower than those of undivided roads.
- Meanwhile, road segments with wider shoulders are associated with higher FIPs.
- Limitation & Future Study
  - The FRMs are built based on mixed road segment types, unobserved heterogeneity
  - Temporal correlation in crash modeling and should be considered in future research possibly using a panel data structure

The End

Thank you!

Q&A

Appendix: Other Truck Traffic Characteristics Variables

Type	Variable	Description	Mean	S.D.	Min	Max
Truck Traffic Characteristics	AADLTT	Annual average daily large truck traffic (vehicle/day)	1142	1672.02	15	8442
	AADLHTT	Annual average daily large heavy truck traffic (vehicle/day)	847.40	1484.20	2	7232
	AADOTT	Annual average daily overweight truck traffic (vehicle/day)	73.62	159.83	0	1239
	$N_{100kips}$	The count of truck over 100 kips (10^4)	0.19	0.33	0	2.28
	$W_{skewness}$	The skewness of truck weight	1.40	0.74	-0.042	5.17
	$W_{Kurtosis}$	The kurtosis of truck weight	5.85	4.41	1.76	47.82
	W_{q85}	The 85 percentile of truck weight	56.39	14.32	25.00	85.00
	$W_{logmean}$	The log mean of truck weight	3.38	0.20	2.91	3.91
	W_{logsd}	The log SD of truck weight	0.53	0.06	0.37	0.78

Note: The above variables are tested in the models, but not significant.