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

Extended Summary

Chuan Xu[1], Kaan Ozbay[2], Hongling Liu[3*], Kun Xie[4], Di Yang[5]

1. *School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, China Department of Civil and Urban Engineering, New York University, 6 MetroTech Center 4th Floor, Brooklyn, NY 11201, USA, cx514@nyu.edu*
2. *Department of Civil and Urban Engineering, New York University, 6 MetroTech Center 4th Floor, Brooklyn, NY 11201, USA, kaan.ozbay@nyu.edu*
3. *School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, China, lhl_lynne@my.swjtu.edu.cn*
4. *Department of Civil & Environmental Engineering, Old Dominion University (ODU), 129C Kaufman Hall, Norfolk, VA 23529, USA, kxie@odu.edu*
5. *Department of Civil and Urban Engineering, New York University, 6 MetroTech Center 4th Floor, Brooklyn, NY 11201, USA, dy855@nyu.edu*

Background

The impact of truck traffic on crash severity should not be neglected but it hasn't been applied appropriately in the practice. Highway Safety Manual (HSM), published by the American Association of State Highway Transportation Officials (AASHTO) provides information and methods for quantitatively evaluating traffic safety performance on roadways (1). It proposed a method to calculate the expected average crash frequency by crash severity for a certain roadway facility type. This method applies safety performance function, crash modification factors, and calibration factor sequentially to calculate the total expected average crash frequency, then 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) (2). The default proportions by crash severity levels are static and there's no difference between two road segments with totally different truck traffic. To study the impact of truck traffic on proportions by severity levels is important to promote the predictive methods introduced by HSM.

Objective

This study seeks to explore the impact of truck traffic on severe crash proportion under the predictive method framework proposed by HSM. Five-year Weigh-in-Motion (WIM) data from 88 WIM stations in New Jersey were used to acquire specific truck traffic information such as truck volume by classification, truck weight. Homogeneous road segments with WIM stations are filtered and used as samples. Then, the road feature data, traffic volume data, crash data are extracted and aggregated for each road segment. Fatal and Injury Proportion (FIP), defined as the fatal and injury crash percentage in the total crash, is computed based on the crash data and used as the response variable. Finally, the relationships between FIP and independent variables are explored by establishing fractional regression models.

Data

WIM data was collected from 88 WIM stations deployed by NJDOT. The WIM data quality is estimated to be satisfactory for our study because those data are used for enforcement purposes which requires higher data quality. However, the individual vehicle records are available for only 61 WIM stations. Each WIM station measured the traffic for both directions, and the period of the WIM data is from 2011 to 2015.

To identify the homogeneous road segment that the WIM station is located on, the 2017 version Straight Line Diagrams (SLD) for New Jersey is used. For the road segment attributes, segment length, rural or urban, divided or undivided, number of lanes, hard shoulder width, posted speed limit is extracted for each homogeneous road segment with a WIM station. The SLD in New Jersey has been updated for several times and some practical projects are based on it, therefore, the overall reliability should be enough for our research.

Five years of crash record data (2011-2015) was obtained from the NJDOT post-processes crash database named Voyager Safety Database (3). Crash location, crash occurrence time, crash severity was extracted. In this study,

the data were aggregated by roadway segment. Since we only considered the road segment features suggested by HSM, intersection-related crashes that are defined as crashes that occur at the intersection or crashes that occur on an intersection approach within 250 ft (2) were removed. The crash severity is coded as fatality, injury, and property damage only.
The data processing and fusion steps are presented in Figure 1. Firstly, as mentioned above, we aggregated individual vehicle records WIM data into yearly station level data. Then the verified road feature data were filtered and fused into homogenous road segments with WIM station dataset. After that, crash records were filtered for the target homogenous road segments and intersection-related crashes are removed. Then, crashes were aggregated by road segment, and the FIP of each segment was extracted. FIP is the overall FIP calculated by all crashes instead of only truck related crashes because trucks may have negative impact on a crash even if they are not involved in that crash directly. Finally, road features, traffic characteristics, truck traffic characteristics, and FIP for homogeneous road segments with WIM stations are fused.

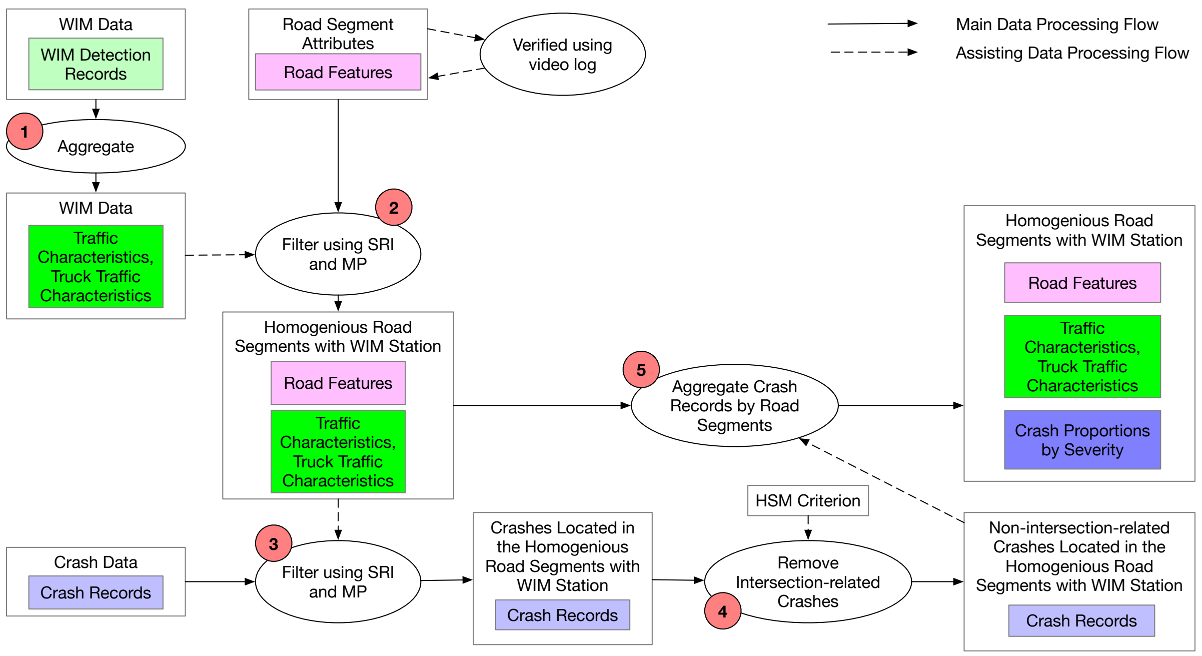

Figure 1: Data processing and fusion steps

Method

FRM (4) applies a functional from G(.) to limit the conditional mean of the dependent variable (y) in the desired constraints. So that, E(y|x) = G(xθ) is constrained in the same interval, where G(.) represents a non-linear function (or named as link function in generalized linear models) satisfying $0 \leqslant G(.) \leqslant 1$, x represents a vector with independent variables (such as road features, traffic characteristics, truck traffic characteristics) and θ is a vector of parameters to be estimated. Papke and Wooldridge (4) suggested any cumulative distribution function can be a possible specification. Some popular choices are Cauchit, Logistic, Standard normal, Extreme maximum (loglog), Extreme minimum (cloglog). Quasi-maximum likelihood (QML) method based on Bernoulli log-likelihood function $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \sum_{i=1}^{N} \boldsymbol{y_i} \log[\boldsymbol{G}(\boldsymbol{x_i\theta})] + (\mathbf{1} - \boldsymbol{y_i})\log[\mathbf{1} - \boldsymbol{G}(\boldsymbol{x_i\theta})]$ is proposed by Papke and Wooldridge (4) to estimate the model. The basic model can be augmented by changing the relationship form between the conditional mean of the y and G(.). As mentioned above, in the basic model, E(y|x) = G(xθ). Ramalho, Ramalho and Murteira (5) proposed three ways to augment the basic model, designated as GOFF1, GOFF2, Generalized GOFF (GGOFF). In GOFF1, E(y|x,)= $\mathbf{G}(\mathbf{x}\boldsymbol{\theta})^{\alpha}$ $\alpha > 0$. The H_0 of GOFF1 is $\alpha = 1$. For GOFF2, E(y|x) = $\mathbf{1} - (\mathbf{1} - \mathbf{G}(\mathbf{x}\boldsymbol{\theta}))^{\alpha}$, $\alpha > 0$. The H_0 of GOFF2 is $\alpha = 1$. In GGOFF, we have E(y|x) = $\boldsymbol{\lambda}\mathbf{G}(\mathbf{x}\boldsymbol{\theta})^{\alpha_1} + (\mathbf{1} - \boldsymbol{\lambda})\{\mathbf{1} - (\mathbf{1} - \mathbf{G}(\mathbf{x}\boldsymbol{\theta}))^{\alpha_2}\}$, $\alpha_1 > 0$, $\alpha_2 > 0$. The H_0 of GGOFF is $\alpha_1 = \alpha_2 = 1$. Moreover, the Goodness-of-Functional From (GOFF) test, including GOFF-I, GOFF-II and GGOFF tests (6) can help to decide whether we can accept the basic G(.) model. In addition, two kinds of FRM structures are tested namely, the one-part model and the two-part model.

Results

After comparing different link functions, for the one-part FRM, loglog link function is selected, and 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, while P tests suggest insignificant performance difference between these two models.

The estimation results of FRM models are shown in Table 1. The results show that two road feature variables, divided by median barrier and hard shoulder width are significant in both models. For M_d in the one-part model, compared to the undivided road, FIP of road segment divided by median barrier can be reduced by 0.123 on average while controlling other contributing factors and in the second part of the two-part model, the partial effect of M_d is -0.132, which is close to that in the one-part model. The coefficients of S_{width} are significantly positive in both the one-part model and the first part of the two-part model which means wider hard shoulder width is related to higher FIP. In the one-part model, the partial effect of S_{width} indicates that if hard shoulder width increases by 1 ft., the FIP will also increase by 0.016 on average. As the estimated result in the first part of the two-part model, the road segments with wider hard shoulder are more likely to have a positive FIP than zero FIP. N_{50kips} are significantly negatively related to FIP, and the partial effects are -0.002 and -0.003 in the one-part model and the second part of the two-part model respectively. This is possible 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. The mean of truck weight is significant (at 0.1% level) in the first part of the two-part model and larger mean of truck weight is related to a higher possibility to have a positive FIP than zero and the partial effect is 0.016 that means if the mean of truck weight on a road segment increases 10,000 kips, the likelihood of a positive FIP increases 0.016. P_{tt} is the proportion of truck traffic, and its value is bounded in [0,1]. It is significantly positive in the one-part model and the second part of the two-part model with partial effects 1.722 and 2.611 respectively.

Table 1: The Estimation Results of the Fractional Regression Models

	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		
RMSE	0.232			0.233		

Note: Test statistics are followed by p-value in the brackets, bolded numbers and ** denote test statistics are significant in 5%, bolded number and * denote test statistics are significant in 10%; partial effect denotes the results of average partial effect for each variable.

Conclusions

In this paper, we explored the impact of truck traffic on severe crash proportions for road segments while controlling other contributing factors. Since large sample size WIM data are not available in most studies, five-year WIM data from 88 WIM stations was utilized in this study to capture the truck weight distribution and other truck traffic characteristics, such as the count of trucks over 50 kips, the mean and standard deviation of truck weight, truck traffic proportion are generated. Road feature, traffic volume, and crash data are also collected and aggregated at the road segment level. To account for the bounded nature of FIP, one-part and two-part FRMs are developed, and the link functions are properly selected based on corresponding statistical tests. In the final models, the mean of truck weight, truck traffic proportion are found to be significant and positively related to FIP. This implies truck traffic has unneglectable effects on severe crash proportion and using static proportions by crash severity levels in HSM may lead to biased estimations.

Acknowledgment

The study is supported by C2SMART, a Tier 1 UTC at New York University funded by the USDOT. The contents of this paper only reflect the views of the authors who are responsible for the facts and do not represent any official views of any sponsoring organizations or agencies. We need to thank NJDOT for providing this data as part of our NJDOT study (7). The authors also thank Bekir Bartin, Sami Demiroluk, Chenchen Wang, Manyu Cheng for the help of obtaining and processing data, and Fan Zuo for his valuable comments.

References

[1] FHWA. Roadway safety data program. https://safety.fhwa.dot.gov/rsdp/hsm.aspx. Accessed November 13, 2021.
[2] AASHTO. Highway Safety Manual. American Association of State Highway and Transportation Officials: Washington, DC, USA, 2010.
[3] Bartin, B., A. Bhattacharyya, K. Ozbay, C. Xu, and H. Nassif. Calibration and development of safety performance functions for rural two-lane two way roadways: A New Jersey case study. 2021.
[4] Papke, L. E., and J. M. Wooldridge. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. Journal of Applied Econometrics, Vol. 11, No. 6, 1996, pp. 619-632.
[5] Ramalho, E. A., J. J. Ramalho, and J. M. Murteira. A generalized goodness-of-functional form test for binary and fractional regression models. The Manchester School, Vol. 82, No. 4, 2014, pp. 488-507.
[6] Ramalho, E. A., J. J. Ramalho, and J. M. Murteira. Alternative estimating and testing empirical strategies for fractional regression models. Journal of Economic Surveys, Vol. 25, No. 1, 2011, pp. 19-68.
[7] Ozbay, K., H. Nassif, B. Bartin, C. Xu, A. Bhattacharyya, P. Ukpah, New Jersey Department of Transportation Bureau of Research, and United States Department of Transportation Federal Highway Administration. Calibration/development of safety performance functions for New Jersey: Final report. https://hdl.handle.net/10929/60785. 2019.