

A Driving Violation Index Modelling Study at Intersections in Urban Precincts Using Big Data

Yan JIA^{1,2,a}, Da-lin QIAN¹ and Jin PU²

¹School of Traffic and Transportation, Beijing Jiaotong University, Beijing, 100044, China

²Traffic Engineering and Technology Research Institute, Tianjin Public Security Traffic Management Bureau, Tianjin, 300040, China

Abstract. Driving violations at intersections are profoundly prevalent in urban precinct in China. These violations, including failing to stop for a red traffic light or arrow, driving across single continuous dividing line, disobeying traffic lane markings, and illegal parking, have become a major issue not only impacting the overall safety of the cities, but also causing traffic delays and accidents. This article focuses on the behaviors of the driving violations at intersections, and investigates what the influencing factors and characteristics are. A model of Driving Violation Index (*DVI*) for these driving violations at intersections has been developed by utilizing the Big Data approach. Based on the actual traffic data from the intersections in the city of Tianjin, a relationship has been fitted and established between the driving violations and the influencing factors, and further yielded the model of calculating the related *DVI*. This research attempts to provide a mechanism to forecast the driving violations at intersections, and explore new strategies for managing the security and safety of the intersections.

1 Introduction

Road traffic crashes are a major public health concern particularly in developing countries. With the recent economic boom in China, vehicle volume and the number of traffic accident fatalities have become the highest in the world. Traffic accidents have become the leading cause of death in China. In 2015, more than 200,000 crashes resulting in damage or injury were reported in China. These crashes resulted in 67,759 deaths. Crashes at intersections account for about 30% of the total crashes in China [1]. Crashes often occur at intersections when drivers “scramble” to gain the right of way in violation of traffic regulations [2]. In the United States in 2011, more than 2.2 million police-reported motor vehicle crashes occurred at intersections or were intersection related.

In previous studies, it is well established that driving violation is one of the major risks threatening road safety in Western societies and in China [3, 4]. Violations among Western drivers mainly include drink-driving and speeding [5-7], and in addition to the violations mentioned above [8], there are distinctive violations in China, such as running the red light, driving across continuous dividing line, disobeying traffic lane markings, using a non-motor lane during traffic congestion, stopping on the road in prohibited areas. Despite most drivers believing red light running was problematic and

^a Corresponding author: jiayan2010@sina.com

dangerous, approximately one in five respondents reported running one or more red lights when entering the last ten signalized intersections [9].

Depending on the environment and the personality of the driver, a particular situation may place the driver in an aggressive disposition. Studies, mainly conducted in the West, have identified some attitudinal and personality factors that are related to driving violations. Medina suggests that driver violation or error contributes to as much as 75% of all roadway crashes, which attempts to construct a generic taxonomy of driver errors and driver error causal factors, based on a synthesis of the available literature on human error and its causal factors [10]. Brosseau study the impact of waiting time and other factors on violations at signalized intersections [11]. Akaateba investigates the influence of three distinct variables: driver educational attainment, driving experience, and form of driver training on drivers' self-reported attitudes towards the frequency of commission of traffic safety violations [12].

In recent years, the comprehensive research on the influence factors of traffic safety is a lot, for example, Abdel-Aty developed a mathematical model that explains the relationship between the frequency of accidents and highway geometric and traffic characteristics, and they confirmed that shoulder and lane, widths, and sharp horizontal curve are found to affected the safety of a roadway, while AADT is significant factor [13]. Mehmood aimed to introduce the System Dynamics approach to simulate the driver behavior in relation of law enforcement, traffic monitoring, and education [14]. McCart and Hu studied the effects of camera enforcement on red light violations, and got the result that red light violations at camera-enforced intersections declined significantly [15]. Al-Ataw assessed the characteristics of red light violations and analyzed the effect of intersection characteristics, such as geometric design, control system and location, on number of violations [16]. Yau study found that male drivers and young drivers are more likely to illegal driving [17].

2 Method

2.1 Big data sources

2.1.1 Traffic violations processing system

At present, more than 300 signal-controlled intersections in Tianjin have been installed video detection devices, which can get the data of drivers' traffic violations at area of the intersections and related data of traffic flow. The data of driving violation behaviors at intersections are obtained from the database of the Traffic Violations Processing System, maintained by the Tianjin Traffic Management Bureau. This database is used to process, store, and report information of traffic violations at intersections in Tianjin, China. The research violation data comes from the TVPS, which process the data covered all the signal-controlled intersections in Tianjin.

2.1.2 Traffic flow collection system

The data of traffic flow comes from the statistical database of Traffic Flow Collection System, which also maintained by the TTMB. Real-time traffic flow data are obtained through video analysis and geomagnetic detection. The traffic flow data in this study are collected by the hour.

2.1.3 Area coordinated system of signal control

The data of signal control is derived from the statistical data of Tianjin Area Coordinated System of Signal Control, and the system adopts the technology of Sydney Coordinated Adaptive Traffic System (SCATS). At present, Tianjin urban intersections are all using SCATS. This study collects the system control mode, the signal circle and the red light time data for analysis and research.

2.2 Measures

2.2.1 Design

The data obtained from the above systems show that the number of traffic violations and the objective conditions are related, because some of the intersection traffic violation data significantly more than others, while the number of violations will change over time and the occurrence of a certain pattern of changes. Therefore, this study aims to find the relationship between traffic violations and various external factors.

In the previous research, there are many methods of data source. Some scholars use the questionnaire [12,18-20], some are observational studies [21], and some use historical statistics [22].

Referring to the previous research methods, this study combines the quantitative data and qualitative data, including state data and dynamic data. Intersections were selected with characteristic differences all over the Tianjin city, in China. Tianjin, as one of China's four municipalities, is located in the south of Beijing. The city's population is more than 10 million, and the motor vehicles are more than 3 million. The traffic violations in Tianjin are prominent. According to the Tianjin Public Security Traffic Management Bureau statistics, in 2015 the Traffic Violations Processing System found and punished traffic violations in the intersections 1000 times. Among them, running a red light 200 times, driving across continuous dividing line 250 times, disobeying traffic lane markings 550 times.

2.2.2 Data processing

In consideration of detection errors of video detection devices, all the data used in this study has been dealt with through error analysis, data reduction and additional treatment of data incompleteness. Based on the statistical data of one intersection and one entrance lane of one day, this study selected the sample data of 300 days in 2015 in the selected 10 intersections in Tianjin, China.

In order to derive the Driving Violation Index (DVI), the general idea is to analysis of the relationship between traffic violations and the influence factors. This DVI is intended as a measure of the propensity for driving violation to be experienced at a signalized intersection. Hamdar got an aggressiveness propensity index for driving behavior at signalized intersections using the SEM method [23]. On the basis of previous studies, this study has established DVI at intersections through the analysis and the research of objective factors of urban intersections.

2.3. Analytical model

The Structural Equation Model (SEM) approach can handle a large number of endogenous and exogenous variables simultaneously, as well as latent variables specified as linear combinations of the observed variables. SEM is the combination of confirmatory factor model and causal model. The factor SEM includes is called measurement model, and its equation is called measurement equation which describes the relationship between latent variables and indicators. Besides, the causal model SEM contains is called latent variable model, also known as structural model, and its equation is called structural equation which describes the relationship among latent variables [24]. With the popularization of vehicles and the development of urban traffic, the influence factors and the complexity of traffic behaviors have increased significantly. Thus, SEM has been applied to traffic field by foreign scholars. Early in the 1980s, Golob applied SEM to the research of traffic demand model [25]. At the beginning of 20th century, foreign scholars were the first to apply SEM to the research of drivers' behavioral characteristics, and got a very good application. Golob using SEM to the study of travel behaviour, and de Abreu e Silva using structural equations modeling to unravel the influence of land use patterns on travel behavior of workers in Montreal [26-27]. Eboli explored land use and transport interaction through Structural Equation Modelling [28].

3. Results

3.1 Analysis of impact factor

Owing to the complexity of objective conditions of urban road intersections, the objective factors that influence driving behaviors are also a comprehensive and complex collection. Table 1-2 provides the factor variable classification, and also analyzes the impact factors through SPSS software. First, the data were tested through the test of KMO and Bartlett, and the KMO value is greater than 0.7, which illustrates that the data is suitable for factor analysis. Besides, the Sig value is less than 0.05, which shows the test is significant. Through the factor analysis of the data, this thesis got a coefficient matrix of factor scores.

Table 1. Observed exogenous variables' description in model

Variable	Abbreviation	Variable Description (Value Unit)
X ₁	Legs	The number of entrances, e.g.: crossroad values four
X ₂	Location	The location of intersections (Downtown area/non-central area/suburb, values 1/2/3)
X ₃	Mlanes	The number of lanes of major road
X ₄	Slanes	The number of lanes of minor road
X ₅	Angle	The included angle of intersected roads (Degree)
X ₆	Alanes	The number of accommodation lanes of all the entrances, and the accommodation lanes include left-lane and right-lane
X ₇	Widthr	Width of main entrance road(m)
X ₈	Widthl	Width of main entrance lane(m)
X ₉	Divided	The number of parts road are divided
X ₁₀	Dsd	Design speed of road (10 km/h)
X ₁₁	Direct	One-way road values 1, and two-way road values 0
X ₁₂	Control	Mode of control signal at intersection (Fixed period/induction/coordinated control value 1/2/3)
X ₁₃	Mred	Average red time of main road (min)
X ₁₄	Camara	The number of Video surveillance facilities at intersection
X ₁₅	Vol	Daily traffic volume at intersection (thousand vehicles)
X ₁₆	Hvol	Daily traffic volume of large scale vehicles at intersections (thousand vehicles)
X ₁₇	Bvol	Daily traffic volume of non-motor vehicles at intersection (thousand vehicles)
X ₁₈	Mque	Average queue length of main road (m)
X ₁₉	Bsp	Is there a bus stop near the intersections? Yes values 1, and no values 0
X ₂₀	Wea	Sunny/ Rainy/Snowy values 1/2/3
X ₂₁	Time	Weekday/weekend/holiday (values 1/2/3)

Table 2. Observed endogenous variables' description

Variable	Abbreviation	Variable Description (Value Unit)
Y ₁	RED	The number of running a red light
Y ₂	LINE	The number of driving across continuous dividing line
Y ₃	CROSS	The number of disobeying traffic lane markings
Y ₄	PARK	The number of illegal parking

According to the analysis result, this thesis reduced the original five impact factors to four impact factors, excluding the environmental factor. Meanwhile, the factor scores of the angle of intersection, the width of lane, the design speed and the one-way road are all less than 0.1, so they are all excluded. Based on the actual experience, because the differences of lane width and design speed in this thesis are small, the two variables of lane width and design speed cannot be treated and analyzed as significant impact factors, which coincides with the analysis result of data software.

3.2 Selected model variables

According to the analysis result of impact factors, this thesis selected among the effective factors, and totally selected four exogenous variables and 1 endogenous variable as variables of structural equation model. The detailed descriptions of selected variables in the model are shown in Table 4.

Table 4. Selected model variables

	Factor	Variable	Abbreviation
Exogenous Variables	F ₁	X ₁	Legs
		X ₂	Location
		X ₃	Mlanes
	F ₂	X ₄	Slanes
		X ₆	Alanes
		X ₇	WidthR
	F ₃	X ₉	Divided
		X ₁₂	Control
		X ₁₃	Mred
		X ₁₄	Camara
	F ₄	X ₁₅	Vol
		X ₁₆	Hvol
		X ₁₇	Bvol
Endogenous Variable	Index	Y ₁	RED
		Y ₂	LINE
		Y ₃	CROSS
		Y ₄	PARK

3.3 Model construction

This study uses the software of LISREL 8.8 to program and establish SEM. Applied the data samples of Tianjin intersections to model training, this thesis has obtained the measurement equation of exogenous latent variables and structural equation model. Establishing models and modifying them through software programming, this thesis has got a path diagram of the structural equation model to calculate driving violation index which includes factor loading and path coefficient, as shown in Figure 1.

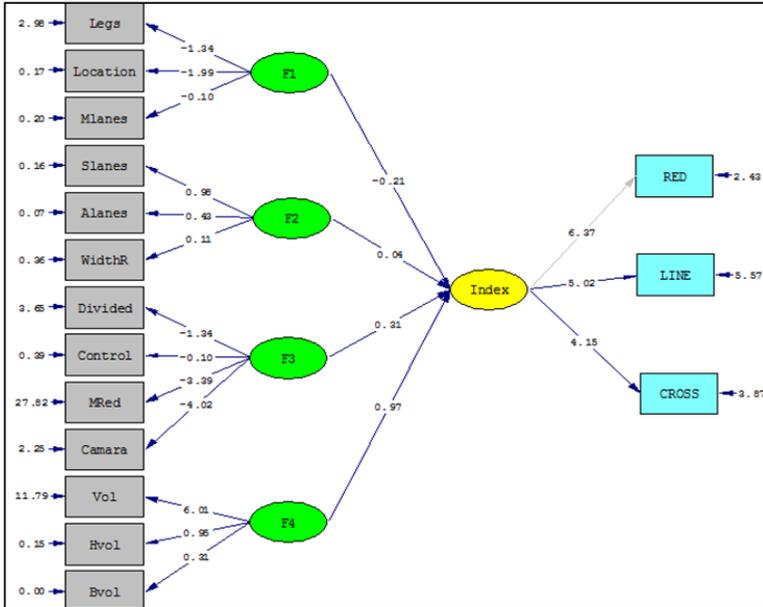


Figure.1 Final Driving violation index structural model by LISREL 8.8.

3.4 Calculation model of driving violation index

The results of calculation model can be acquired in accordance with SEM, among which *Err* is the error term. Models include the measurement model of endogenous variables, measurement model of exogenous variables and calculation model of violation propensity index. The result summarizing the model are presented in Table 5.

Table 5. Results of SEM

Exogenous measurement model		Endogenous measurement model	
1	$X_1 = -1.34 * F_1$; err=2.98	1	$Y_1 = 6.37 * DVI$; err=2.43
2	$X_2 = -1.99 * F_1$; err=0.17	2	$Y_2 = 5.02 * DVI$; err=5.57
3	$X_3 = -0.10 * F_1$; err=0.20	3	$Y_3 = 4.15 * DVI$; err=3.87
4	$X_4 = 0.98 * F_2$; err=0.16		
5	$X_6 = 0.43 * F_2$; err=0.07		
6	$X_7 = 0.11 * F_2$; err=0.36		
7	$X_9 = -1.34 * F_3$; err=3.65		
8	$X_{12} = -0.10 * F_3$; err=0.39		
9	$X_{13} = -3.39 * F_3$; err=27.82		
10	$X_{14} = -4.02 * F_3$; err=2.25		
11	$X_{15} = -6.01 * F_4$; err=11.79		
12	$X_{16} = -0.95 * F_4$; err=0.15		
13	$X_{17} = -0.31 * F_4$; err=0.00		
		$DVI = -0.21 * F_1 + 0.04 * F_2 + 0.31 * F_3 + 0.97 * F_4$	

4 Discussion

4.1 Analysis of influencing factors

Through the model fitting of 300 data samples, this thesis uses structural equation model to obtain the calculation model of *DVI*. In the final fitting model, the Chi-Square value is 171.9, *RMSEA* value is 0.089, *GFI* value is 0.68 and *AGFI* value is 0.76. The fitting results are in an acceptable range when the sample size is much larger than observed variables. Because the model fitting is conducted with real data, it is inevitable that the data maybe uncertain and mutant. Therefore, it can be assumed that the result of model fitting is an ideal result.

The final structural equation model classifies the exogenous variables X_1 , X_2 and X_3 as impact factor F_1 which represents the characteristic index of intersections; classifies the exogenous variables X_4 , X_6 and X_7 as impact factor F_2 represents the characteristic index of entrance lanes at intersections; classifies the exogenous variables X_9 , X_{12} , X_{13} and X_{14} as impact factor F_3 which represents the characteristic index of traffic control and traffic management; and classifies the exogenous variables X_{15} , X_{16} and X_{17} as impact factor F_4 which represents the characteristic index of traffic flow. The model fitting has excluded the exogenous X_{18} for its relevancy to the exogenous variable X_{15} is larger than 0.9. Through the comparison of path coefficient of impact factors, the sort of effects of impact factors on *DVI* is as follows: $F_4 > F_3 > F_1 > F_2$. It shows that the traffic flow characteristic has the greatest influence on the traffic violation behaviors, and the impact factor of the entrance lane characteristics at intersections is the least.

Meanwhile, through the comparison of the path coefficients Y_1 , Y_2 and Y_3 , running the red light has the greatest influence on the *DVI* among traffic violation behaviours at intersections, followed by pressing the solid lines, and the third is driving not according to the guiding lanes. The illegal parking is eliminated by the model, which shows the low correlation between illegal parking at intersections and driving violation index.

4.2. Driving Violation Index Calculation

Making use of the model results obtained in this study, the *DVI* value of a certain intersection can be obtained. The violation propensity index of each intersection is related to traffic flow, traffic condition and other dynamic variables, so the average traffic data of intersections should be taken into account when calculating the violation propensity index of intersections. Meanwhile, the data of different entrance lanes should be averaged to calculate the result of *DVI*. According to the calculation model, the corresponding F_i value can be obtained by taking one factor in each factor, and then the Index value is calculated. At the same time, we can estimate the number of traffic violations according to the Index value. Research shows that with the increase of Index value, the tendency of traffic violations also increases. Table 6. presents the calculation results of the classical intersections in this study in Tianjin, China. Similarly, we can extend the results to other intersections or other Chinese cities, because of the similar characteristics of drivers and traffic. When the standard index is greater than 0.8, the intersection is considered to be more dangerous, it is recommended to take countermeasures, such as the number 7, number 8 intersections in this study.

Table 6 *DVI* of the intersections in this study

Num	Intersection name	DVI Standard	Num	Intersection name	DVI Standard
1	Jiefang Road-Taian Street	0.76	3	Youyi Road-Shaoxing Street	0.65
2	Youyi Road-Waihuan Road	0.89	4	Waihuan Road-Xianfeng Road	0.98

Conclusions

In this paper, we have creatively proposed the concept of violation propensity index. Through collecting and organizing the data of illegal driving at intersections and other relevant data of road traffic, it established a SEM-based calculation model of violation propensity index. With the application of real data of Tianjin, it obtained the path coefficient and factor loading chart of SEM by programming with the software LISREL 8.8. Through analysing the impact factors of traffic violation behaviours at urban intersections in China, this thesis explored the relationship between traffic violation behaviours at intersections and intersection characteristics, road characteristics, traffic control characteristics and traffic flow characteristics. Besides, it also innovatively acquires the SEM-based calculation method of driving violation index at intersections through the combination of qualitative analysis and quantitative calculation, which provides new ideas and methods for safety evaluation of intersections. This study is not perfect, and further research in the following areas is suggested.

References

1. Pei, Y.-L., Ma, Y.-L. *Journal of Harbin Institute of Technology* **37**, 3 (2005)
2. Shi, Jing, et al. *Accid Anal Prev* **43**, 1540-1546 (2011)
3. Xie, Cheng-qiu, and Dianne Parker. *Transport. Res. Part F* **5**, 293-308 (2002)
4. Zhang, Guangnan, Kelvin K. W. Yau, and Guanghan Chen. *Accid Anal Prev* **59**, 18-25 (2013)
5. Bhalla, Kavi, et al. *Injury* **44**, Supplement **4**, S49-S56 (2013).
6. Gheorghiu, Alexandra, Patricia Delhomme, and Marie Line Felonneau. *Transport. Res. Part F* **35**, 101-111 (2015)
7. MacLeod, Kara E., et al. *Accid Anal Prev* **81**, 134-142 (2015)
8. Zhang, Guangnan, Kelvin K. W. Yau, and Xiangpu Gong. *Accid Anal Prev* **64**, 30-40 (2014)
9. Porter, Bryan E., and Thomas D. Berry. *Accid Anal Prev* **33**, 735-741 (2001)
10. Medina, A.L., Lee, S.E., Wierwille, W.W., Hanowski, R.J. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting, 2075–2080* (2004)
11. Brosseau, Marilyne, et al. *Transport. Res. Part F* **21**, 159-172 (2013)
12. Akaateba, Millicent Awialie, Richard Amoh-Gyimah, and Owusu Amponsah. *Safety Science* **75**, 156-162 (2015)
13. Abdel-Aty, Mohamed A., and A. Essam Radwan. *Accid Anal Prev* **32**, 633-642 (2000)
14. Mehmood, Arif. *Accid Anal Prev* **42**, 427-436 (2010)
15. McCart, Anne T., and Wen Hu. *Journal of Safety Research* **48**, 57-62 (2014)
16. Al-Atawi, Attiyah M. *IATSS Research* **37**, 119-123 (2014)
17. Yau, Kelvin K. W. *Accid Anal Prev* **36**, 333-340 (2004)
18. Forward, Sonja E. *Transport. Res. Part F* **9**, 412-426 (2006)
19. Yagil, Dana. *Transport. Res. Part F* **1**, 123-135 (1998)
20. Yeh, Ming-Shan, et al. *Transport. Res. Part F* **32**, 35-45 (1998)
21. Yan, Fangfang, et al. *Accid Anal Prev* (2015).
22. Chen, Chen, and Yuanchang Xie. *Accid Anal Prev* **91**, 72-83 (2016)
23. Hamdar, Samer H., Hani S. Mahmassani, and Roger B. Chen. *Accid Anal Prev* **40**, 315-326 (2008)
24. Muthen, B. *Psychometrika* **49**, 115-132 (1984)
25. Golob, T.F., Meurs, T. *Transport. Res. Part A* **21**, 391-400 (1987)
26. Golob, Thomas F. *Transport. Res. Part B* **37**, 1-25 (2003)
27. de Abreu e Silva, João, Catherine Morency, and Konstadinos G. Goulias. *Transport. Res. Part A* **46**, 1252-1264 (2012)
28. Eboli, Laura, Carmen Forciniti, and Gabriella Mazzulla. *Procedia Social and Behavioral Sciences* **54**, 107-116 (2012)