Prediction of Frequencies of Truck-involved and Non-truck-involved Crashes on Roadway Segments in Ontario

Ran Ran
University of Windsor

Follow this and additional works at: https://scholar.uwindsor.ca/etd

Recommended Citation
Ran, Ran, "Prediction of Frequencies of Truck-involved and Non-truck-involved Crashes on Roadway Segments in Ontario" (2015). Electronic Theses and Dissertations. 5475.
https://scholar.uwindsor.ca/etd/5475

This online database contains the full-text of PhD dissertations and Masters' theses of University of Windsor students from 1954 forward. These documents are made available for personal study and research purposes only, in accordance with the Canadian Copyright Act and the Creative Commons license—CC BY-NC-ND (Attribution, Non-Commercial, No Derivative Works). Under this license, works must always be attributed to the copyright holder (original author), cannot be used for any commercial purposes, and may not be altered. Any other use would require the permission of the copyright holder. Students may inquire about withdrawing their dissertation and/or thesis from this database. For additional inquiries, please contact the repository administrator via email (scholarship@uwindsor.ca) or by telephone at 519-253-3000ext. 3208.
Prediction of Frequencies of Truck-involved and Non-truck-involved Crashes on Roadway Segments in Ontario

By

Ran Ran

A Thesis
Submitted to the Faculty of Graduate Studies through the Department of Civil and Environmental Engineering in Partial Fulfillment of the Requirements for the Degree of Master of Applied Science at the University of Windsor

Windsor, Ontario, Canada

2015

© 2015 Ran Ran
Prediction of Frequencies of Truck-involved and Non-truck-involved Crashes on Roadway Segments in Ontario

By
Ran Ran

APPROVED BY:

______________________________
Dr. H. Maoh, Department Reader
Department of Civil and Environmental Engineering

______________________________
Dr. X. Guo, Outside Program Reader
Odette School of Business

______________________________
Prof. J. Tofflemire, Special Advisor
Department of Civil and Environmental Engineering

______________________________
Dr. C. Lee, Advisor
Department of Civil and Environmental Engineering

September 18, 2015
Declaration of Previous Publication

This thesis includes material from one original paper that has been submitted for publication in a peer reviewed journal, as follows:

<table>
<thead>
<tr>
<th>Thesis Chapter</th>
<th>Publication title/full citation</th>
<th>Publication status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 5</td>
<td>Lee, C., Ran, R., 2015. Modeling Frequency of Truck-involved and Non-truck-involved Crashes on Roadway Segments in Ontario. Submitted for presentation at 95th Transportation Research Board Annual Meeting and publication in Transportation Research Record: Journal of the Transportation Research Board.</td>
<td>under review</td>
</tr>
</tbody>
</table>

I certify that I have obtained a written permission from the copyright owner(s) to include the above published material(s) in my thesis. I certify that the above material describes work completed during my registration as a graduate student at the University of Windsor.

I declare that, to the best of my knowledge, my thesis does not infringe upon anyone’s copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act,
I certify that I have obtained written permission from the copyright owner(s) to include such material(s) in my thesis.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.
Abstract

This study develops the methods of predicting frequency of truck-involved and non-truck-involved crashes on roadway segments and identifies unique characteristics of truck-involved crashes. To capture these nonlinear effects of the variables and temporal correlations among annual crash frequencies, Generalized Estimating Equation (GEE) models with nonlinearizing link functions were developed. Separate GEE models for total, truck-involved and non-truck-involved crashes were developed and compared. The result of the models shows that annual total and non-truck-involved crash frequencies in two successive years at a given location are correlated but the correlation does not exist for truck-involved crashes. The result also shows that nonlinearizing link functions of lane width, truck percentage and speed limit were statistically significant in the truck-involved crash. Thus, the proposed method can capture important nonlinear effects of variables on crash frequencies with temporal correlations, and identify the differences in the factors contributing to crash frequency between truck-involved and non-truck-involved crashes.
Acknowledgments

I am deeply indebted to my supervisor Dr. Chris Lee whose stimulating motivation and valuable ideas helped me to complete this research. His enthusiasm, encouragement and faith in me is extremely inspiring. He was always available for my question and he was positive and gave generously of his time and vast knowledge.

I would like to thank my parents for their precious love and support. They are always standing by me during the hard time.

I would like to thank all my friends and staff of department for their help in the whole process leading to the success of this study.
# Table of Contents

Declaration of Previous Publication .................................................................................. iii
Abstract.......................................................................................................................... v
Acknowledgments.......................................................................................................... vi
List of Tables ................................................................................................................... ix
List of Figures ................................................................................................................ xi
1. Introduction.................................................................................................................. 1
2. Literature Review........................................................................................................ 4
   2.1. Factors Related to Crash Frequency....................................................................... 4
      2.1.1. Speed............................................................................................................. 4
      2.1.2. Geometric characteristics............................................................................. 5
      2.1.3. Traffic volume.............................................................................................. 8
      2.1.4. Other factors................................................................................................. 9
      2.1.5. Transportation polices to improve safety..................................................... 10
   2.2. Crash Frequency Models....................................................................................... 11
3. Data............................................................................................................................ 16
4. Methods....................................................................................................................... 25
   4.1. Generalized Nonlinear Models............................................................................ 25
   4.2. Generalized Estimating Equation....................................................................... 27
5. Results and Discussion ............................................................................................... 33
   5.1. Nonlinearizing Link Functions ........................................................................... 33
   5.2. Result of GNM..................................................................................................... 47
   5.3. Result of GLM...................................................................................................... 49
   5.4. Result of GEE..................................................................................................... 51
5.5. Model Fit of GEE................................................................. 55

6. Conclusions and Recommendations ............................................. 61
   6.1. Conclusions ........................................................................ 61
   6.2. Contributions ..................................................................... 63
   6.3. Limitations and Recommendations ........................................ 64

References ................................................................................. 65
Appendix A. Results of GEE Models ............................................. 74
Appendix B. Cumulative Residual Plots of GEE Models ..................... 77
Vita Auctoris ............................................................................. 97
List of Tables

Table 3-1. Descriptive statistics of continuous explanatory variables.......................... 16
Table 3-2. Descriptive statistics of crash frequency ..................................................... 17
Table 3-3. List of categorical geometric variables......................................................... 18
Table 5-1. Estimated parameters of Generalized Nonlinear Models for total crashes .... 47
Table 5-2. Estimated Parameters of Generalized Nonlinear Models for truck-involved crashes ................................................................. 48
Table 5-3. Estimated parameters of Generalized Nonlinear Models for non-truck-involved crashes ........................................................................................................... 48
Table 5-4. Estimated parameters of Generalized Linear Models for total crashes ........ 49
Table 5-5. Estimated Parameters of Generalized Linear Models for truck-involved crashes ............................................................................................................................ 49
Table 5-6. Estimated parameters of Generalized Linear Models for non-truck-involved crashes ............................................................................................................................ 50
Table 5-7. Comparison of model fit between GLM and GNM ........................................ 50
Table 5-8. Estimated parameters of Generalized Estimating Equation for total crashes (Exchangeable correlation structure) ................................................................. 52
Table 5-9. Estimated parameters of Generalized Estimating Equation for truck-involved crashes (Independent correlation structure) .......................................................... 53
Table 5-10. Estimated parameters of Generalized Estimating Equation for non-truck-involved crashes (Exchangeable correlation structure) .................................................. 54
Table 5-11. Statistically significant variables for different crash types.......................... 54
Table 5-12. QIC values of Generalized Estimating Equation for different crash types ... 56
Table A-1. Estimated parameters of Generalized Estimating Equation for total crashes (Independent correlation structure) ................................................................. 74
Table A-2. Estimated parameters of Generalized Estimating Equation for total crashes (Autoregressive correlation structure) ................................................................. 74
Table A-3. Estimated parameters of Generalized Estimating Equation for total crashes (Unstructured correlation structure) ................................................................. 74

Table A-4. Estimated parameters of Generalized Estimating Equation for truck-involved crashes (Exchangeable correlation structure) ........................................ 75

Table A-5. Estimated parameters of Generalized Estimating Equation for non-truck-involved crashes (Independent correlation structure) ........................................ 75

Table A-6. Estimated parameters of Generalized Estimating Equation for non-truck-involved crashes (Autoregressive correlation structure) ....................................... 75

Table A-7. Estimated parameters of Generalized Estimating Equation for non-truck-involved crashes (Unstructured correlation structure) ........................................ 76
List of Figures

Figure 3-1. Crash rates by road classification................................................................. 19
Figure 3-2. Crash rates by median type ......................................................................... 20
Figure 3-3. Crash rates by posted speed limit................................................................. 21
Figure 3-4. Crash rates by truck percentage ................................................................. 22
Figure 3-5. Crash frequency trend for different road segments ................................. 23
Figure 3-6. Average annual precipitation of Ontario..................................................... 24
Figure 4-1. Rear-end crash rate (crash frequency per mile) in 5 years (2002–2006) from 10 highways in Washington State by grade ................................................................. 26
Figure 5-1. Observed relationship between total crash rate and lane width ............ 34
Figure 5-2. Nonlinearizing link function for lane width................................................. 35
Figure 5-3. Observed relationship between total crash rate and AADT ....................... 36
Figure 5-4. Nonlinearizing link function for AADT ....................................................... 37
Figure 5-5. Observed relationship between total crash rate and speed limit .......... 38
Figure 5-6. Nonlinearizing link function for speed limit ............................................. 39
Figure 5-7. Observed relationship between crash rate and truck percentage .......... 39
Figure 5-8. Nonlinearizing link function for truck percentage .................................... 40
Figure 5-9. Observed relationship between total crash rate and right shoulder width .... 41
Figure 5-10. Observed relationship between total crash rate and number of lanes ...... 42
Figure 5-11. Observed relationships between total crash rate and different combination of speed limit and truck percentage (Total Crash) ................................................................. 43
Figure 5-12. Relationships between total crash rate and different combination of speed limit and truck percentage (Truck-involved crash) .................................................... 44
Figure 5-13. Relationships between total crash rate and different combination of speed limit and truck percentage (Non-truck-involved crash) .................................................... 44
Figure 5-14. Nonlinearizing link functions for interaction between truck percentage and speed limit ........................................................................................................................................ 46

Figure 5-15. Cumulative residuals plot for surface width in different correlation structures for non-truck-involved crashes ........................................................................................................ 58

Figure B-1. Cumulative residuals plot for shoulder width in different correlation structures for total crashes ...................................................................................................................................... 77

Figure B-2. Cumulative residuals plot for terrain in different correlation structures for total crashes ...................................................................................................................................... 79

Figure B-3. Cumulative residuals plot for terrain in different correlation structures for total crashes ...................................................................................................................................... 81

Figure B-4. Cumulative residuals plot for length in different correlation structures for total crashes ...................................................................................................................................... 83

Figure B-5. Cumulative residuals plot for shoulder width in different correlation structures for truck-involved crashes ........................................................................................................ 85

Figure B-6. Cumulative residuals plot for median shoulder width in different correlation structures for truck-involved crashes ........................................................................................................ 86

Figure B-7. Cumulative residuals plot for median width in different correlation structures for truck-involved crashes ........................................................................................................ 87

Figure B-8. Cumulative residuals plot for length in different correlation structures for truck-involved crashes ...................................................................................................................................... 88

Figure B-9. Cumulative residuals plot for length in different correlation structures for non-truck-involved crashes ........................................................................................................ 89

Figure B-10. Cumulative residuals plot for shoulder width in different correlation structures for non-truck-involved crashes ........................................................................................................ 91

Figure B-11. Cumulative residuals plot for median width in different correlation structures for non-truck-involved crashes ........................................................................................................ 93

Figure B-12. Cumulative residuals plot for terrain width in different correlation structures for non-truck-involved crashes ........................................................................................................ 95
1. Introduction

A vehicle is the most important transportation mode in the modern world. However, the safety of traffic is always a major concern. There were 2,006 fatalities and 166,725 injuries caused by vehicle crashes in Canada in 2012 (Transport Canada, 2015). Also, more than 5.4 million motor vehicle crashes occurred in the United States in 2010 (NHTSA, 2012). Twenty eight percent of those crashes (1.54 million) led to an injury, and less than 1% (30,196) resulted in a death. Various measures have been exploited to make driving safer, such as optimization of road design, more safety features in vehicles, and traffic policies restraining drivers’ behaviour.

To develop countermeasures, several studies have been conducted in the past to evaluate the influence of road geometry, traffic, environment, and driver behavior on crashes. Generally, higher AADT, higher traffic density, higher post speed limit, more number of lanes and bad weather conditions will lead to higher crash frequency. Caliendo et al. (2013) found that the numbers of both severe and non-severe crashes increased when the number of lanes increased. On the other hand, wider shoulder, wider lanes, good lighting conditions, flat and narrow median and continuously reinforced concrete pavement will reduce crash frequency. For instance, Elvik (2004) concluded that fatal traffic crash frequencies were reduced by 34% when mean speed of driving was reduced by 10%.

The trucking industry has played a significant role of moving goods especially after the globalization of trade. It is predicted that the tonnage of goods transported by
domestic freight and logistics will increase by 65-70% by 2020s in the U. S. (Mallet et al., 2004). Similarly, the number of registered large trucks has increased by 63% since 1990 in Ontario. Consequently, as the demand of surface freight transportation system increases, the number of truck-involved crashes will also increase. In particular, a large number of truck-involved crashes lead to fatality and injury. According to the U.S. National Highway Traffic Safety Administration (NHTSA) (2014), large trucks (gross vehicle weight rating greater than 10,000 pounds) were involved in the traffic crashes which led to 3,921 fatalities and 104,000 injuries in 2012 – i.e., 18% and 4% increase compared to 2011, respectively (NHTSA, 2014). Thus, it is important to analyze truck-involved crashes and identify their unique characteristics compared to non-truck-involved crashes.

Most studies on truck-involved crashes focused on identifying the relationship between frequency of truck-involved crashes and the related factors such as truck percentage. Different types of techniques have been adapt to model crash frequency. For instance, the negative binomial (NB) regression model, novel multinomial generalized Poisson (MGP) model and generalized additive model (GAMs) were used in previous studies.

Kotikalapudi and Dissanayake (2013) found that posted speed limit significantly affected truck-involved crashes. They also found that truck-involved crashes are less likely to occur on road segments with more number of lanes and wider lane width. On the other hand, wider median contribute to higher car and car–truck crash frequencies (Dong et al., 2014). These studies mainly assumed that these factors have linear relationships with frequencies of truck-involved crashes. However, this assumption is violated if the
actual relationship between crash frequency and the related factors is not linear. Also, past studies concentrated on only one type of crash: total crashes or truck-involved crash. Thus, there is a lack of studies on the comparison between truck-involved and non-truck-involved crashes.

The objectives of this study are 1) to develop proper methods of predicting frequency truck-involved and non-truck-involved crashes on road segments based on the relationships between crash frequency and the related factors and 2) identify unique characteristics of truck-involved crashes from the comparison between truck-involved and non-truck-involved crashes. To capture the nonlinear effect of variables on crash frequency, the variation of crash rate with the factors will be observed and reflected in the model development. Also, temporal correlations among annual crash frequencies at a given site will be taken into account in the modeling framework.
2. Literature Review

This section reviews various studies that identified the factors contributing to crash frequency. The section also reviews the methods that have been used to predict crash frequency and the transportation policies that aimed at reducing truck-involved crashes.

2.1. Factors Related to Crash Frequency

2.1.1. Speed

Previous studies found that driving speed is a vital factor of road safety. In general, it takes longer time for vehicles to stop at higher speed. If a crash occurs due to driver’s exceeding speed limits or driving too fast, the crash is classified as speed-related crash in accordance with the U.S. NHTSA (2012). Nearly 55 percent of speed-related crashes were attributed to exceeding speed limits (NHTSA, 2012). However, due to lack of information on actual driving speed, the speed limit at the crash location is taken as an approximate speed of the vehicle.

Vernon et al. (2004) analyzed observed crashes on Interstate highways in Utah with speed limits in the range of 60-75 mph. They found that the total crash rate significantly increased after speed limit increased from 60–65 mph on urban road segments. However, there was no significant change in crash frequency and injury severity on rural Interstate segments after speed limit increased to 70–75 mph.

Similar analysis conducted by Elvik (2014) concluded that speed is the most important factor related to the frequency of higher injury severity. Fatal traffic crashes
tend to be reduced by 34% if there is a 10% reduction on mean speed of driving (Elvik, 2004). Jun et al. (2011) compared driving speed patterns between the drivers involved in crashes and the drivers not involved in crashes. The authors found that the drivers involved in crashes are more likely to drive on the road with higher speed than the drivers not involved in crashes. They also found that the drivers the drivers who exceeded the speed limits are more likely to be involved in crashes.

Some studies focused on the effects of speed on truck-involved crashes. For instance, Kotikalapudi and Dissanayake (2013) identified a list of significant factors related to truck crashes using five-year crash data in Kansas, U.S.A. They observed that more than 38% of truck-involved crashes in Kansas occurred on the road segments with 61-70 mph in 2004-2008. They also observed that a driver of large trucks is 1.56 times more likely to have a higher injury severity than the drivers of the other vehicle types if the driver is speeding. Choi et al. (2014) also reported that the effects of speed-related variables on injury severity of truck-involved crashes are more statistically significant than the effects of volume-related variables. Thus, speed management is the most effective way to reduce truck-involved crashes (Choi et al., 2014).

2.1.2. Geometric characteristics

Past studies have found that road geometric characteristics are closely associated with crash frequency. For instance, more number of lanes on the road generally increases the chance of crashes. This is because more number of lanes increases opportunity of changing lanes and the number of conflicts among vehicles (Caliendo et al., 2013). Kim et al. (2006) also found that the numbers of both severe and non-severe crashes increase
as the number of lanes increases. However, Zhu and Srinivasan (2011) found that the crashes were less severe on roads with more number of lanes.

Lane width is also a significant factor affecting crash frequency. Dong et al. (2014) observed that the number of car–truck crashes was higher at intersections with wider lanes of both minor and major roads. According to Elvik and Vaa (2004), larger lane width leads to a decrease in crash rates on rural roads but a slight increase on urban roads. Elvik and Vaa (2004) also reported that crash rate can be reduced by up to 10% if lane width increases within a standard design range. Harnen et al. (2003) found that wider lane was better for reducing motorcycle crashes. They predicted that 0.5 m increase in lane width at intersections can reduce motorcycle crash rate by about 4-6%. They also found that 3.2 m or wider lanes reduced motorcycle-related crashes by 34% compared to lane width less than 3.2 m.

Shoulders on the roadside are reserved lanes which often serve as a stopping lane for emergencies (FHWA, 2007). Shoulders can reduce the chance of rear-end crashes and severe congestion by removing disabled vehicles from high-speed and high-volume highways such as urban freeway. Also, when drivers unintentionally leave their lanes or try to avoid objects and depressions in their present lanes, appropriate shoulders on road side provide safe area to reduce the risk of collision.

Haleem et al. (2013) studied how the change in width of shoulders affects crash frequency using crash modification factors (i.e., the multiplicative factors used to compute the expected number of crashes after implementing a given countermeasure at a specific site). The study found that wider inside and outside shoulder can reduce the number of total crashes because wider shoulder can provide recovery area for errant
vehicles and parking space for disabled vehicles (Bonneson and Pratt, 2009). They also found that the segments with 9 feet or more outside shoulder had lower probability of fatal and injury crashes. The authors explained that this is potentially because shoulders work as buffers and safe space for problematic vehicles and erroneous manipulations. Li et al. (2014) found that right shoulder width also had a negative effect on crash frequency although the effect of right shoulder width was not statistical significant.

Median width is another geometric factor related to crashes. In general, wider median contributes to lower crash frequencies. This is because wider median provides larger recovery area and reduces glare of oncoming vehicle headlights in the opposite direction (Bonneson and Pratt, 2009). Wider median also reduces the likelihood of vehicle’s crossing the median and entering the lanes in the opposite direction. Thus, wider median helps reduce head-on collisions. In addition, wider median can provide temporary parking area for vehicles with mechanical failures. Due to these safety benefits of wider median, Haleem et al. (2013) found that 40-feet medians increase total, fatal and injury crashes by 7% compared to 64-feet medians. Furthermore, 22-feet medians lead to increase in total crashes and fatal and injury crashes by 263% and 223%, respectively, compared to 64-feet median. However, Dong et al. (2014) found that wider median rather increased car and car-truck crash frequency at intersections due to larger area for turning and higher chances of wrong way entries. Hu (2010) investigated how different types of medians affected crashes. It was found that steeper median side-slopes substantially increased the possibility of rollover, and both frequency and severity of median barrier crashes. However, flat and narrow medians are less likely to prevent cross-median
crashes occurrence (Hu, 2010). Hosseinpour et al. (2014) found that presence of median can effectively prevent head-on crashes which are more likely to be fatal crashes.

To prevent cross-median crashes, median barrier has been installed. Thus, selection of an appropriate median type is critical for road safety development (FHWA, 2007). Concrete is rigid material which will not deflect while encountering impact. When impact angle is high, the impact could be more severe if median is made in concrete instead of material with ductility. Although median barriers are more likely to cause median-related crashes, they can help mitigate crashes with higher injury severity - e.g., cross median crashes (Hu, 2010).

Intersection angle is related to truck, truck-car and car crashes at intersections (Dong et al., 2014). For instance, vehicles will have to traverse a longer distance to cross a skewed-angle intersection than a right-angle intersection. This leads to more time to be exposed to the traffic in other approaches and increase collision risk.

2.1.3. Traffic volume

Traffic volume was also found to have significant effects on crash frequency in previous studies. Traffic volume is the most common measures of roadway usage and control factor (Christoforou et al., 2011). It is typically applied to calculate crash rate which reflects the chance of crash occurrence in a certain time period. Different measures of traffic volume such as Annual Average Daily Traffic (AADT), Vehicle-miles Traveled (VMT), and Number of Entering Vehicles (NEV) (intersections only) have been used. AADT represents daily volume of vehicle traffic on a road segment for a specific year (Qin et al., 2006).
A number of studies identified different relationships between AADT and crash frequency. Caliendo et al. (2013) observed that the relationship between AADT and crash is not linear. In free-flow conditions, the number of crashes increased with AADT. However, in congested conditions, the number of crashes decreased with AADT increase. Christoforou et al. (2011) found that the number of two-vehicle sideswipe crashes increased with traffic volume. Hu et al. (2012) also found that crash frequency monotonically increased as AADT increased.

Furthermore, some studies classified traffic volume by vehicle type and used truck percentage in total traffic volume as a control factor. Dong et al. (2014) found that as truck percentage increases, the opportunity of a collision involving with at least one truck would also increase. On the other hand, Kotikalapudi and Dissanayake (2013) observed that angle crashes on the major roads tend to increase as truck percentage increases. This is potentially because higher number of trucks in traffic flow increases likelihood of car driver’s sight obstruction and car driver’s lane change to overtake preceding trucks in free-flow conditions.

2.1.4. Other factors

There are other factors affecting crash frequency such as pavement condition and weather. Li et al. (2013) found the relationship between pavement type and crash severity using the two-year crash data in Texas in 2008-2009. Pavement types include jointed Portland cement concrete pavement (JCP), continuously reinforced concrete pavement (CRCP) and asphalt concrete pavement (ACP). Generally, CRCP has been widely used to pave major highways such as urban Interstates. For other kinds of highways, ACP is more
commonly used. The authors found that JCP is related to more severe crashes than ACP and CRCP. They also suggest that passenger cars are more likely to lose control in poor weather conditions than commercial vehicles (Li et al., 2013).

Besides types of pavement, weather can also affect road surface conditions. Kotikalapudi and Dissanayake (2013) found that icy and slushy road conditions are associated with truck-related crashes. The authors also found that blacktop surface and dry surface are related to more severe truck-related crashes. On the other hand, Zhu and Srinivasan (2011) revealed that crashes that occurred on wet roads tend to be less severe. The authors explained that this is potentially because drivers pay more attention to wet surface conditions (Zhu and Srinivasan, 2011). In this regard, Choi et al. (2014) claimed that crash severity is lower in adverse weather conditions because most crashes occur in lower speed.

**2.1.5. Transportation polices to improve safety**

Various transportation policies have been implemented to reduce conflicts between trucks and other vehicle types. One example is differential speed limits (DSL) which set different speed limits for different vehicle types - a lower speed limit is set for trucks than passenger cars. Lower truck speed can generally reduce crash frequency and injury severity. However, larger speed difference between truck and other vehicle types can rather increase the frequency of rear-end crashes and lane-change crashes (Garber et al., 2006).

Garber et al. (2006) investigated the effects of DSL on crash frequency. They found that changing uniform speed limit (i.e., the same speed limit for all vehicle types) to DSL
changed crash frequency. However, they found that the change in crash frequency was not solely caused by the change of speed limit policy (Garber et al., 2006). The potential reason for this inconclusive result is that the effect of DSL depends on many other factors such as truck percentage, AADT, and lane restriction for trucks.

Restriction of lane usage for trucks on freeways has also been implemented to reduce truck-involved crashes in Louisiana, Tennessee and Texas (Qi, 2009). This policy restricts heavy trucks to use one or more specific lanes so that the interactions among trucks and other vehicle types can be minimized. Borchardt (2002) observed that lane restriction for trucks reduced the total number of crashes by 30%. Zeitz (2003) also found that the rate of truck-involved crashes could be reduced by 78% if lane restriction for trucks were implemented on the road sections of I-85 in South Carolina. Cate et al. (2004) recommended applying lane restriction for trucks on freeway sections with 6 or more lanes and restricting trucks to use more than one lane.

Archer and Young (2009) reported that red-light running at signalized intersections causes serious consequences when crashes occur. In particular, they observed that the probability of red-light running was higher for heavy vehicles than the other vehicle types. Thus, they proposed an all-red extension for potential red-light runners and a green extension for heavy vehicles which are detected at the dilemma zone.

2.2. Crash Frequency Models

Over past decades, different methodologies have been developed to identify the relationship between crash frequency and contributing factors. A majority of the previous studies used the generalized linear models (GLM). As an extension of the traditional
linear model, GLM can fit data with distributions in exponential family and allow independent to be variables linearly related to dependent variables through a nonlinear link function. GLM describes a dependent variable in a function of explanatory variables as follows:

\[ Y = \exp(\alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k) \]

where \( Y \) is the expected crash frequency during a certain time period; \( X_k \) is the explanatory variable related to crash frequency; \( \alpha \) is a constant, and \( \beta_k \) is the coefficients for the explanatory variables \( X_k \). A positive coefficient \( \beta_k \) indicates that as the value of \( X_k \) increases, crash frequency also increases.

GLM can be developed by choosing different distributions of crash frequency such as Poisson on negative binomial (NB) distributions. Poisson regression models have been widely used in predicting crash frequency. Kumara and Chin (2005) used this methodology after modifying the original Poisson regression model. They found that left-turn volume, number of signal phases during one cycle and shorter sight distance were significant factors affecting crash frequency at three-legged signalized intersections. Other researchers also applied the Poisson regression model to analyze crash frequency data (Ye et al., 2013; Li et al., 2013).

The Poisson distribution assumes that the mean and the standard deviation are equal consequently, the distribution is not valid if the variation in crash frequency is larger than the mean crash frequency (i.e., over-dispersion). To account for over-dispersion, the negative binomial distribution in which the error terms following the Gamma distribution has been applied to crash frequency models (Hauer, 2001). Unlike the Poisson
distribution, the negative binomial distribution allows the standard deviation of crash frequency to vary with the mean crash frequency. More specifically, the standard deviation of the crashes equals the square root of the mean + mean² / k where k is the over-dispersion parameter which is determined from the data. For instance, Hu et al. (2012) used the NB regression model to determine the factors that are statistically significant to traffic collisions at highway-railroad grade crossings.

Recently, generalized additive models (GAM) were applied to prediction of crash frequency. GAMs is an extension of GLM with additive terms (Hastie and Tibshirani, 1986; Wood, 2008). The linear predictor in GAM partly depends on some unknown smooth functions. This model maximizes the quality of estimates of a dependent variable by estimating non-parametric function and presents nonlinear relationship between crash frequency and changes in road segment characteristics.

The additive logistic regression model is an example of a generalized additive model. In the additive logistic regression model, the linear term is replaced by general function which could be nonlinear. Ma and Yan (2014) used the additive logistic regression model to examine the effect of drivers’ age on the odds of being at fault in rear-end crashes. They observed the nonlinear effect of age using flexible additive terms.

However, the limitation of GLM is that the model cannot account for temporal and spatial correlations among crash frequencies. For instance, when crash frequencies are repeatedly observed every year at the same location, they are potentially correlated (i.e., temporal correlation effect). Similarly, crash frequencies on different road segments in the same corridor are likely to be correlated (i.e., spatial correlation effect).
In this regard, the random parameter count model considers location-specific
effects of variables to account for correlations in longitudinal data (e.g., annual crash
frequencies at a given location). For instance, Anastasopoulos and Mannering (2009)
used the random parameter count model to predict crash frequency and found that the
model showed better goodness-of-fit than the fixed parameter count models.
Venkataraman et al. (2013) also applied the random parameter negative binomial model
to account for heterogeneity of effects of variables among different locations. However,
the estimation of the random parameter count model is complex as simulation-based
maximum likelihood method is required.

Alternatively, the General Estimating Equations (GEE) can accommodate
correlations in longitudinal crash frequency data in the model. GEE specifies the
correlation structure in crash frequencies unlike GLMs. The main advantage of GEE is
that the model can handle temporal correlation even without knowing the extent and type
of correlation (Liang and Zeger, 1986). Lord and Persaud (2000) found that the GEE with
temporal correlation outperformed the models without temporal correlation using 6-year
crash frequency data in Toronto, Canada. Wang and Abdel-Aty (2006) also found that
there exist temporal and spatial correlations of rear-end crashes at signalized intersections
in Florida based on the results of GEEs. Recently, Mohammadi et al. (2014) found that
the GEE with temporal correlations produced more accurate and less biased estimates
than the models without temporal correlations using 10-year longitudinal crash frequency
data in Missouri, U.S.A.

Another limitation of the GLM is that the model does not reflect nonlinear
relationship between crash frequency and the related variables. Although categorical
variables or dummy variables can be incorporated in the model to capture the nonlinearity, the model cannot fully identify general relationship. In this regard, Lao et al. (2013) applied the generalized nonlinear model (GNM) to account for nonlinear relationship between crash rate and the related factors. They demonstrated that right shoulder width, AADT, grade percentage, and truck percentage had significant nonlinear effects on crashes and GNM showed better model performance than GLM. However, GNM and GEE have not been integrated to address both temporal correlation in longitudinal crash frequency data and nonlinear relationship between crash frequency and the related factors.
3. Data

A list of data including crash, road geometry and traffic information were obtained from the Ontario Ministry of Transportation were used in this study. Seven-year (2004-2010) data were collected from 6,475 roadway segments of Ontario’s highway system. In this study, only crashes that occurred within road segments not influenced by intersections were analyzed. A majority (63%) of the road segments are freeways and arterials. Approximately 8% of total crashes involved heavy trucks. Although a majority of crashes were non-truck-involved crashes, total crashes were also analyzed as total crash frequency has been modelled in the past studies. The variables in the data are listed in Table 3-1.

Table 3-1. Descriptive statistics of continuous explanatory variables

<table>
<thead>
<tr>
<th>Numeric Variables</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>AADT (veh/day)</td>
<td>1</td>
<td>44290</td>
<td>24801.34</td>
<td>45469.52</td>
</tr>
<tr>
<td>Truck Percentage (%)</td>
<td>0</td>
<td>78.12</td>
<td>16.25</td>
<td>11.21</td>
</tr>
<tr>
<td>Truck AADT (veh/day)</td>
<td>0</td>
<td>26875</td>
<td>2885.72</td>
<td>4364.65</td>
</tr>
<tr>
<td>Length (km)</td>
<td>0.01</td>
<td>59</td>
<td>6.32</td>
<td>6.55</td>
</tr>
<tr>
<td>Posted speed limit (km/h)</td>
<td>50</td>
<td>100</td>
<td>87.81</td>
<td>10.63</td>
</tr>
<tr>
<td>Number of lanes</td>
<td>1</td>
<td>14</td>
<td>3.39</td>
<td>2.30</td>
</tr>
<tr>
<td>Lane width (m)</td>
<td>1.825</td>
<td>7.9</td>
<td>3.52</td>
<td>0.38</td>
</tr>
<tr>
<td>Surface width (m)</td>
<td>3.8</td>
<td>51.2</td>
<td>11.25</td>
<td>8.68</td>
</tr>
<tr>
<td>Streams</td>
<td>1</td>
<td>2</td>
<td>1.43</td>
<td>0.66</td>
</tr>
<tr>
<td>Median shoulder width (m)</td>
<td>0</td>
<td>4.8</td>
<td>0.58</td>
<td>1.02</td>
</tr>
<tr>
<td>Median width (m)</td>
<td>0</td>
<td>30.5</td>
<td>4.36</td>
<td>8.50</td>
</tr>
<tr>
<td>Shoulder width (m)</td>
<td>0</td>
<td>5.5</td>
<td>2.26</td>
<td>0.87</td>
</tr>
</tbody>
</table>
Table 3-2. Descriptive statistics of crash frequency

<table>
<thead>
<tr>
<th>Annual number of crashes by injury severity level</th>
<th>Minimum</th>
<th>Minimum</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>0</td>
<td>627</td>
<td>26.11</td>
<td>50.12</td>
</tr>
<tr>
<td>Injury and Fatal</td>
<td>0</td>
<td>392</td>
<td>7.88</td>
<td>20.24</td>
</tr>
<tr>
<td>Fatality</td>
<td>0</td>
<td>3</td>
<td>0.09</td>
<td>0.32</td>
</tr>
<tr>
<td>Major</td>
<td>0</td>
<td>7</td>
<td>0.27</td>
<td>0.64</td>
</tr>
<tr>
<td>Minor</td>
<td>0</td>
<td>40</td>
<td>2.34</td>
<td>4.61</td>
</tr>
<tr>
<td>Minimal</td>
<td>0</td>
<td>53</td>
<td>2.44</td>
<td>5.50</td>
</tr>
<tr>
<td>PDO</td>
<td>0</td>
<td>517</td>
<td>20.92</td>
<td>40.58</td>
</tr>
</tbody>
</table>

The database consists of three different data sets: road geometry data, crash frequency data and traffic volume data. These data were combined by matching LHRS (Linear Highway Referencing System) numbers which are the identification number of each road segment. Road geometry is a unique characteristic of each segment whereas crash frequency and traffic volume change every year. Each road segment has a different length.

Since crash frequency was generally higher for longer segment, crash rate (i.e., crash frequency divided by length of segment) was computed for each segment. Injury severity of crashes was classified into five levels: fatal, major, minor, minimal, and property damage only (PDO).
Table 3-3. List of categorical geometric variables

<table>
<thead>
<tr>
<th>Categorical variables</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road Classification</td>
<td>Freeway (33.5 %)</td>
</tr>
<tr>
<td></td>
<td>Arterial (29.9%)</td>
</tr>
<tr>
<td></td>
<td>Collector (16.1%)</td>
</tr>
<tr>
<td></td>
<td>Local (20.5%)</td>
</tr>
<tr>
<td>Shoulder type</td>
<td>Gravel</td>
</tr>
<tr>
<td></td>
<td>Paved</td>
</tr>
<tr>
<td></td>
<td>Partially paved</td>
</tr>
<tr>
<td>Divided</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Terrain</td>
<td>Flat</td>
</tr>
<tr>
<td></td>
<td>Mountainous</td>
</tr>
<tr>
<td></td>
<td>Rolling</td>
</tr>
<tr>
<td>Road surface type</td>
<td>Asphalt</td>
</tr>
<tr>
<td></td>
<td>Not Asphalt</td>
</tr>
<tr>
<td>Median type</td>
<td>Barrier Curb</td>
</tr>
<tr>
<td></td>
<td>Box Beam Guide Rail</td>
</tr>
<tr>
<td></td>
<td>Grass Depressed</td>
</tr>
<tr>
<td></td>
<td>Granular Filled Steel Barrier</td>
</tr>
<tr>
<td></td>
<td>Painted</td>
</tr>
<tr>
<td></td>
<td>Raised, Guide Rail with AntiGlare Screen</td>
</tr>
<tr>
<td></td>
<td>Raised, Steel Flex Beams Guide Rail</td>
</tr>
<tr>
<td></td>
<td>Standard Concrete Barrier/New Jersey Barrier</td>
</tr>
<tr>
<td></td>
<td>Six Cable Rail Wire Steel Post</td>
</tr>
<tr>
<td></td>
<td>Raised, Six Cable Guide Wire Wood Post</td>
</tr>
<tr>
<td></td>
<td>Singing Strip</td>
</tr>
<tr>
<td></td>
<td>Tall Wall Concrete Barrier i.e. High Mast Lighting</td>
</tr>
</tbody>
</table>
Figure 3-1 shows the comparison of crash rates on different road classifications for total crashes and heavy truck-involved crashes. Heavy truck-involved crashes are defined as the crashes involving at least one heavy truck. The figure shows that total crash rate was highest on freeways among the four road classifications. Heavy truck-involved crash rate was also highest on freeways. Similar patterns were observed for fatal and injury crash rates.

![Crash rates by road classification](image)

**Figure 3-1. Crash rates by road classification**
Figure 3-2 shows that crash rates for all crash types were higher for divided roads than undivided roads. This is because divided road segments generally have higher AADT and posted speed limit, which will increase the crash rate.

![Crash rates by median type]

**Figure 3-2. Crash rates by median type**
Figure 3-3 demonstrates that crash rates of total and fatal/injury crashes were highest for the segments with 100 km/h of the posted speed limit. However, although it is expected that crash rate generally increases as speed limit increases, the figure shows that the relationship between crash rate and speed limit is not linear.

![Crash rates by posted speed limit](image)

**Figure 3-3. Crash rates by posted speed limit**

Figure 3-4(a) shows the trend of total crash rates for different truck percentages. The crash rate was highest for 5-10% for both total crashes and fatal/injury crashes. Similar pattern was observed for truck-involved crashes as shown in Figure 3-4(b).
Figure 3-4. Crash rates by truck percentage
Figure 3-5 shows trends of annual crash frequencies for a sample of three road segments. As shown in the figure, crash frequencies of the LHRS section 10130 were always higher than crash frequencies of the other two segments (24150, 36340) in all seven years. This pattern indicates that some segments have consistently higher crash frequency than the other segments every year. This demonstrates that there are potential temporal correlations among annual crash frequencies on the same segment. Thus, these correlations must be considered in development of crash frequency models.

![Figure 3-5. Crash Frequency Trend for Different Road Segments](image)

Since crash frequency is closely related to weather conditions, annual weather data for Ontario were also obtained from Environment Canada (2013). Figure 3-6 shows the annual precipitation in Ontario from 2004 to 2010. The annual precipitation was calculated as an average of annual precipitations observed at 358 weather stations in Ontario. For snowfall, density corrections based upon coincident ruler and Nipher measurements were applied to all snow ruler measurements (Mekis and Brown, 2010). It
was observed that average precipitation was highest in 2005 and lowest in 2007 among 7 years. However, average precipitation did not significantly vary across years.

Figure 3-6. Average annual precipitation of Ontario
4. Methods

To identify the significant factors contributing to truck-involved crashes and the relationships between crash frequency and the related factors, statistical models were applied. Generalized nonlinear models and generalized estimating equations were used to capture nonlinear effects of variables on crash rate and temporal correlation among crash frequencies, respectively. These models are explained in detail in this section.

4.1. Generalized Nonlinear Models

GLMs assume linear monotone effect (positive or negative) of variables on crash frequency. Thus, the model cannot account for nonlinear effects of variables on crash frequency although the effect can change as the value of variable changes.

To overcome the limitation of GLMs, the generalized nonlinear model (GNM) was developed (Lao et al., 2013). GNM relax the assumption of linear relationships in GLMs and captures nonlinear effect of variables on a dependent variable using nonlinearizing link functions. The functional specification of the GNM is as follows:

\[ Y_t = \exp(\alpha + \beta_k x_k + \gamma_l U(z_l)) \]  

(4-1)

where

- \( Y_t \) = the expected crash frequency during a certain time period \( t \);
- \( x_k \) = vectors of linear predictors related to crash frequency \( (k = 1, \ldots, K) \);
- \( z_l \) = vectors of nonlinear predictors related to crash frequency \( (l = 1, \ldots, L) \);
- \( K, L \) = numbers of linear and nonlinear predictors, respectively;
\[ U(z_i) = \text{the nonlinearizing link function which varies with the value of } z_i; \]
\[ \alpha = \text{a constant}; \]
\[ \beta_k, \gamma_l = \text{the coefficients}. \]

Among different forms of the nonlinearizing link function (e.g., quadratic, exponential), the function that can best fit the observed relationship between the logarithm of crash rate and the variable \( z_i \) is selected (Lao et al., 2013). If a single function cannot fit the observed relationship for all values of \( z_i \), different functions can be introduced to achieve the better fit for different ranges of the value of \( z_i \) separately.

For example, Lao et al. (2013) developed nonlinear link functions based on the observed nonlinear relationship between crash rate and roadway grade from five-year crash data in Washington State as shown in Figure 4-1.

![Figure 4-1. Rear-end crash rate (crash frequency per mile) in 5 years (2002–2006) from 10 highways in Washington State by grade
[Source: Lao et al., 2013]](image-url)
The figure shows that the observed crash rate fluctuates for different grades. The linear predictor $L(x)$ used in the GLM which assumes monotonic relationship between crash rate and grade was fit to the observed data. The linear predictor shows that crash frequency consistently decreases as grade increases in the linear regression. However, this predictor could not reflect the actual relationship. Instead, the nonlinear predictor $U(x)$ was used in the GNM which defines different relationships for different grades better reflects the actual relationship.

Lao et al. (2013) further compared the model fit between the linear and nonlinear predictors based on $R^2$ as follows:

$$R^2 = \frac{SS_{ssr}}{SS_{sst}} \quad (4-2)$$

where $SS_{ssr}$ is the explained variation and $SS_{sst}$ is the total variation. They found that $R^2$ for $U(x)$ and $L(x)$ are 0.68 and 0.06, respectively, which indicates the nonlinear predictor. In this study, the models were developed using the GENMOD procedure in SAS 9.2 (SAS Institute, 2012).

### 4.2. Generalized Estimating Equation

Compared to the GLM, generalized estimating equation (GEE) is a more flexible approach to estimate the mean and analyze the within-subject association structure (Fitzmaurice et al., 1993). GEE accommodates the GLM to correlated data and relax the strict distribution assumption of variables (Ghisletta and Spini, 2004).

Since crash frequencies are repeatedly observed every year at each location (i.e., subject), they are likely to be correlated for the same location. Hence, when longitudinal
crash frequency data are used, temporal correlation should be taken into consideration for model development.

Assume that the observed crash frequency at location \( i \) in year \( t \) is \( y_{it} \) where \( i = 1, \ldots, I \) (number of locations) and \( t = 1, \ldots, T \) (number of years). GEE estimates the coefficient \( \beta \) by solving the quasi-score differential function as follows (Liang and Zeger, 1986):

\[
U_k(\beta) = \sum_{i=1}^{k} \frac{\partial \mu_i}{\partial \beta} V_i^{-1} (Y_i - \mu_i) = 0
\]  
(4-3)

where

\( U_k(\beta) \) = the quasi-score differential function;

\( Y_i \) = a vector of the crash frequency at location \( i \), \([y_{i1}, \ldots, y_{iT}]\);

\( \mu_i \) = a vector of the expected value of the crash frequency at location \( i \), \([\mu_{i1}, \ldots, \mu_{iT}]\);

\( V_i \) = the covariance matrix of \( Y_i \) which specifies types of temporal correlations of crash frequency as follows:

\[
V_i = \phi A_i^{1/2} R_i(\alpha) A_i^{1/2}
\]  
(4-4)

where

\( \phi \) = a scale parameter;

\( A_i^{1/2} \) = a \( T \times T \) diagonal matrix with the variance function of \( Y_i \), \( v(\mu_{ij}) \), as the \( j \)th diagonal element;

\( R_i(\alpha) \) = the working correlation matrix of \( Y_i \) with a vector of parameters \( \alpha \).
The value of $\alpha$ determines the structure of correlation. Since 7-year crash data are used in this study, there are 7 annual crash frequencies and the working correlation matrix is expressed as a $7 \times 7$ matrix. There are four types of correlation structures as follows (Liang and Zeger, 1986):

(a) Independent $R_i(\alpha)$

The independent structure assumes that the correlation between two crash frequencies is independent. This implies that there is no correlation between two different annual crash frequencies. As a result, the $R_i(\alpha)$ is expressed as follows:

$$
R_{7 \times 7} = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
$$

This matrix is symmetrical because the correlation between crash frequencies in any two years is zero regardless of their chronological order.

(b) Exchangeable $R_i(\alpha)$

The exchangeable structure assumes that the correlation between any two years is constant ($\alpha$). The maximum value of $\alpha$ is 1.
(c) Autoregressive $R_i(\alpha)$

The autoregressive structure assumes that the correlation between two observed frequencies in successive years is stronger than the correlation between two observed frequencies in two years with a gap.

\[
R_{7\times7} = \begin{bmatrix}
1 & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha \\
\alpha & 1 & \alpha & \alpha & \alpha & \alpha & \alpha \\
\alpha & \alpha & 1 & \alpha & \alpha & \alpha & \alpha \\
\alpha & \alpha & \alpha & 1 & \alpha & \alpha & \alpha \\
\alpha & \alpha & \alpha & \alpha & 1 & \alpha & \alpha \\
\alpha & \alpha & \alpha & \alpha & \alpha & 1 & \alpha \\
\alpha & \alpha & \alpha & \alpha & \alpha & \alpha & 1
\end{bmatrix}
\]

Since $\alpha$ is smaller than one, the correlation decreases as the difference in years between the observed crash frequencies increases.

(d) Unstructured $R_i(\alpha)$

In the unstructured correlation matrix, there is no specific relationship among annual crash frequencies as follows:
\[
R_{7 \times 7} = \begin{bmatrix}
1 & \alpha_{12} & \alpha_{13} & \alpha_{14} & \alpha_{15} & \alpha_{16} & \alpha_{17} \\
\alpha_{21} & 1 & \alpha_{23} & \alpha_{24} & \alpha_{25} & \alpha_{26} & \alpha_{27} \\
\alpha_{31} & \alpha_{32} & 1 & \alpha_{34} & \alpha_{35} & \alpha_{36} & \alpha_{37} \\
\alpha_{41} & \alpha_{42} & \alpha_{43} & 1 & \alpha_{45} & \alpha_{46} & \alpha_{47} \\
\alpha_{51} & \alpha_{52} & \alpha_{53} & \alpha_{54} & 1 & \alpha_{56} & \alpha_{57} \\
\alpha_{61} & \alpha_{62} & \alpha_{63} & \alpha_{64} & \alpha_{65} & 1 & \alpha_{67} \\
\alpha_{71} & \alpha_{72} & \alpha_{73} & \alpha_{74} & \alpha_{75} & \alpha_{76} & 1 
\end{bmatrix}
\]

However, the correlation type for repeated measurements is not always known in many situations (Mohammadi et al., 2014). Thus, the working correlation matrix is estimated using iterative fitting process in the following steps (SAS, 2014):

1) Compute an initial estimate of \( \beta_0 \) using an ordinary GLM.

2) Compute the working correlation matrix \( R \) based on the following Pearson residuals and the current \( \beta_n \):

\[
e_{ij} = \frac{y_{ij} - \mu_{ij}}{\sqrt{v(\mu_{ij})}}
\]

(4-5)

3) Compute an estimate of the covariance matrix using Equation 4-4.

4) Update \( \beta \) using the following equation:

\[
\beta_{n+1} = \beta_n + \left[ \sum_{i=1}^{K} \frac{\partial \mu_i}{\partial \beta} V_i^{-1} \frac{\partial \mu_i}{\partial \beta} \right]^{-1} \left[ \sum_{i=1}^{K} \frac{\partial \mu_i}{\partial \beta} V_i^{-1} Y_i - \mu_i \right]
\]

(4-6)

5) Repeat steps 2-4 until the solution converges.

The goodness-of-fit of the GEE is evaluated based on quasi-likelihood. The QIC (Quasi-likelihood under the Independent model Criterion) developed by Pan (2001) is a
modification of Akaike's Information Criterion (AIC) and it can be used to assess
goodness-of-fit of the GEE. The QIC is expressed as follows:

\[
QIC(R) = -2Q(\hat{\beta}(R), \phi) + 2\text{trace}(\hat{\Omega}_l, \hat{V}_R)
\]  

(4-7)

where

\( \hat{\beta}(R) \) = a parameter estimate of GEE under the working correlation structure \( R \);

\( Q(\hat{\beta}(R), \phi) \) = the quasi-likelihood function under the independent working correlation assumption;

\( \hat{V}_R \) = the robust covariance estimate;

\( \hat{\Omega}_l \) = the inverse of the model-based covariance estimate under the independent working correlation assumption as follows (Pan, 2001):

\[
\hat{\Omega}_l = -\left. \frac{\partial^2 Q(\beta, \phi)}{\partial \beta \cdot \partial \beta'} \right|_{\beta = \hat{\beta}}
\]  

(4-8)

The above QIC can also be used to select the working correlation structure. Similar
to AIC, smaller QIC represents better model fit accounting for the number of explanatory
variables. The models were developed using the GENMOD procedure with the
5. Results and Discussion

5.1. Nonlinearizing Link Functions

The first step in the analysis focused on exploring the non-linear link function that describes the relationship between the explanatory variables and the crash rate. Figure 5-1 shows the observed relationship between lane width and crash rate. It was observed that crash rate does not consistently increase or decrease as lane width increases. Each dotted line represents the relationship between lane width and crash rate for each year from 2004 to 2010. Since all the lines have a similar trend, general relationship can be represented by the red solid line which shows annual average crash rate for each lane width. It is critical to check that the relationships between explanatory variables and crash rate are similar in all years so that a single nonlinearizing link function can be developed to describe the nonlinear effects of a given geometric factors on crash rate for all years.
Figure 5-1. Observed relationship between total crash rate and lane width

Based on the plot of lane width and crash rate, lane width can be used as a nonlinear predictor in the model. Lane width was classified into one of the following three ranges: less than or equal to 3.4 m, 3.4-3.7 m and greater than 3.7 m. Nonlinearizing link functions for lane width were developed using the actual lane width of all the segments within each range of lane width.

Nonlinearizing link functions describe logarithm of 7-year crash rates in a function lane width ($LW$) as follows (Figure 5-2):

Total crash:

$$U_{LW} \begin{cases} 
= -0.5861LW + 1.7962 & \text{if } LW \leq 3.4m \\
= 28.317LW^2 - 190.01LW + 318.64 & \text{if } 3.4m < LW \leq 3.7m \\
= -6.7579LW + 27.038 & \text{if } LW > 3.7m
\end{cases}$$
Truck-involved crash:

\[U_{\text{LW}}\left\{\begin{array}{ll}
0.3945LW - 2.2750 & \text{if } LW \leq 3.4m \\
38.418LW^2 - 261.62LW + 444.4741 & \text{if } 3.4m < LW \leq 3.7m \\
-6.0376LW + 23.5241 & \text{if } LW > 3.7m
\end{array}\right.\]

Non-truck-involved crash:

\[U_{\text{LW}}\left\{\begin{array}{ll}
-0.0407LW - 0.5350 & \text{if } LW \leq 3.4m \\
39.403LW^2 - 267.7LW + 453.9941 & \text{if } 3.4m < LW \leq 3.7m \\
-7.2683LW + 28.4551 & \text{if } LW > 3.7m
\end{array}\right.\]

The figure shows that the crash rate is highest for 3.7-m lane width and the patterns of relationship were similar for all crash types. Similar concave downward functions for lane width were also found in the previous studies (Xie et al., 2007; Lee et al., 2015).
Similarly, relationships between crash rate and the other factors such as AADT, speed limit, truck percentage and shoulder width were also explored. Figure 5-3 shows that crash rate generally increases as AADT increases but it slightly decreases from $9,000 < \text{AADT} \leq 12,000$ to $12,000 < \text{AADT} \leq 15,000$ then it increases again. Therefore, the observed relationship between crash rate and AADT is also nonlinear. Again, the patterns of the relationship were similar in all years.

![Figure 5-3. Observed relationship between total crash rate and AADT](image)

Similar to lane width, nonlinearizing link functions for AADT were developed using the actual AADT of all the segments. Nonlinearizing link functions for AADT are described in quadratic functions as follows (Figure 5-4):
Total crash: \[ U_{AADT} = -10^{-10} AADT^2 + 4 \times 10^{-5} AADT + 0.4645 \]

Truck-involved crash: \[ U_{AADT} = -6 \times 10^{-11} AADT^2 + 3 \times 10^{-5} AADT - 0.07941 \]

Non-truck-involved crash: \[ U_{AADT} = -6 \times 10^{-11} AADT^2 + 3 \times 10^{-5} AADT + 0.22859 \]

Figure 5-4. Nonlinearizing link function for AADT

Figure 5-5 shows that the observed relationship between crash rate and speed limit (\( SL \)) is also nonlinear. To plot the relationship, speed limit was categorized into 4 ranges which are \( SL \leq 70 \text{ km/h}, 70 \text{ km/h} < SL \leq 80 \text{ km/h}, 80 \text{ km/h} < SL \leq 90 \text{ km/h} \) and \( 90 \text{ km/h} < SL \leq 100 \text{ km/h} \). The crash rate decreases from 50-70 km/h to 70-80 km/h of speed limit but then gradually increases as speed limit increases.
Figure 5-5. Observed relationship between total crash rate and speed limit

Based on this observed U-shape relationship, nonlinearizing link functions for speed limit were developed using the actual speed limit of all the segments. Nonlinearizing link functions for speed limit ($SL$) are described in quadratic functions as follows (Figure 5-6):

Total crash:  \[ U_{SL} = 0.0062SL^2 - 0.9121SL + 33.225 \]

Truck-involved crash:  \[ U_{SL} = 0.0059SL^2 - 0.8641SL + 30.718 \]

Non-truck-involved crash:  \[ U_{SL} = 0.0066SL^2 - 0.98SL + 35.37109 \]
Figure 5-6. Nonlinearizing link function for speed limit

Figure 5-7 shows that the observed relationship between crash rate and truck percentage is also nonlinear. The total crash rate and non-truck-involved crash rate increase from 0-5% to 5-10% of trucks and then gradually decreases as truck percentage increases. However, truck-involved crash rate increases from 0-5% to 10-15% of trucks and then gradually decreases as truck percentage increases.

Figure 5-7. Observed relationship between crash rate and truck percentage
Based on this observed relationship, nonlinearizing link functions for truck percentage were developed using the actual truck percentage of all the segments. Nonlinearizing link functions for truck percentage \((TP)\) are described as follows (Figure 5-8):

- Total crash: \(U_{TP} = -0.0053TP^2 + 0.1043TP + 0.9168\)
- Truck-involved crash: \(U_{TP} = -0.005TP^2 + 0.1154TP - 0.1757\)
- Non-truck-involved crash: \(U_{TP} = -0.0055TP^2 + 0.1026TP + 0.4381\)

![Figure 5-8. Nonlinearizing link function for truck percentage](image)

Figure 5-9(a) shows the observed relationship between right shoulder width and annual crash rates. Similar to lane width and AADT, the patterns of crash rate were similar among different years. However, the relationship appears to be linear because crash rate consistently increases as right shoulder width increases. This is potentially because drivers tend to be careless and drive faster on the road with wider shoulder. Thus, logarithm of total 7 year crash rates can be fit to a linear function of shoulder width as shown in Figure 5-9 (b). Since a linear function reasonably fits the observed relationship, a nonlinearizing link function for right shoulder width was not developed.
Figure 5-9. Observed relationship between total crash rate and right shoulder width
Similarly, the relationship between number of lanes and logarithm of total 7 year crash rates can also be fit to a linear function as shown in Figure 5-10. The figure shows that the crash rate linearly increases with number of lanes (R-squared value was higher for the linear function (0.9512) than the nonlinear function (0.9401)). This is potentially because drivers tend to feel more comfortable on wider roadways and travel at higher speed. Based on this observed relationship, number of lanes is considered as a linear predictor.

![Graph showing relationship between number of lanes and logarithm of crashes per kilometer.](image)

**Figure 5-10. Observed relationship between total crash rate and number of lanes**

In addition to independent effects of each variable, interaction effects of two variables were also investigated. Figure 5-11 shows that relationships between crash rate and truck percentage are nonlinear for all ranges of speed limits. The figure also shows that the segments with speed limit higher than 80 km/h have the higher crash rate than speed limits lower than or equal to 80 km/h. However, crash rate is higher for speed limit between 60 and 80 km/h than speed limit lower than 60 km/h only for 0-5% of trucks.
This indicates that crash rate is not consistently higher for higher speed limits. This implies that driver’s speeds are affected by speed limits and truck percentage concurrently.

**Figure 5-11. Observed relationships between total crash rate and different combination of speed limit and truck percentage (Total Crash)**

The relationships between truck percentage and crash rate at different speed limit ranges were also compared for truck-involved crashes as shown in Figure 5-12. The figure shows that the relationship between truck percentage and truck-involved crash rate is also nonlinear similar to total crashes. However, crash rates were consistently lower for speed limit between 60 and 80 km/h than speed limit lower than 60 km/h. Also, truck-involved crash rates become similar for speed limit higher than 80 km/h and speed limit lower than 60 km/h as truck percentage increases. This is potentially because as truck percentage increases, truck drivers become more cautious regardless of speed limits. Also, they are less likely to change their lanes at higher truck percentage due to less available gaps and
spacing. Thus, this truck driver’s behaviour helps reduce truck-involved crashes. However, such behaviour is less likely to be affected by speed limits if truck percentage is higher. On the other hand, the trends of non-truck-involved crashes were similar to the trends of total crashes as shown in Figure 5-13.

![Figure 5-12. Relationships between total crash rate and different combination of speed limit and truck percentage (Truck-involved crash)](image)

![Figure 5-13. Relationships between total crash rate and different combination of speed limit and truck percentage (Non-truck-involved crash)](image)
Based on these observed relationships, nonlinearizing link functions for the interaction between speed limit and truck percentage ($U_{SL*TP}$) were developed as follows (Figure 5-14):

Total crash:

$$U_{SL*TP} = \begin{cases} 
-0.0056TP^2 + 0.1577TP + 0.6642 & \text{SL} \leq 60 \text{ km/h} \\
-0.0049TP^2 + 0.0937TP + 0.994 & 60 \text{ km/h} < \text{SL} \leq 80 \text{ km/h} \\
-0.0211TP^2 + 0.6672TP + 2.3373 & \text{SL} > 80 \text{ km/h}
\end{cases}$$

Truck-involved crash:

$$U_{SL*TP} = \begin{cases} 
-0.0083TP^2 + 0.2107TP - 0.2874 & \text{SL} \leq 60 \text{ km/h} \\
-0.0053TP^2 + 0.0935TP - 1.172 & 60 \text{ km/h} < \text{SL} \leq 80 \text{ km/h} \\
-0.0027TP^2 - 0.0057TP + 1.9122 & \text{SL} > 80 \text{ km/h}
\end{cases}$$

Non-truck-involved crash:

$$U_{SL*TP} = \begin{cases} 
-0.0059TP^2 + 0.1904TP + 0.7062 & \text{SL} \leq 60 \text{ km/h} \\
-0.0053TP^2 + 0.1012TP + 1.0713 & 60 \text{ km/h} < \text{SL} \leq 80 \text{ km/h} \\
-0.0203TP^2 + 0.5672TP + 1.5073 & \text{SL} > 80 \text{ km/h}
\end{cases}$$
Figure 5-14. Nonlinearizing link functions for interaction between truck percentage and speed limit
5.2. Result of GNM

To compare truck-involved crashes with non-truck-involved crashes, GNM{s} were separately developed for each crash type using the nonlinearizing link functions in Section 5.1. The GNM describes 7-year crash frequency in a function of explanatory variables. The AADT and truck percentage in the GNM are averages of annual AADT and truck percentage in 2004-2010. Since the speed limit was correlated with many geometric factors, it was removed from the models to capture the effects of road geometry. The result of GNM for total crashes is shown in Table 5-1. The nonlinear predictors of truck percentage, AADT, and interaction between speed limit and truck percentage were statistically significant at a 95% confidence level. Besides, the segments with wider shoulder, longer length and rolling or mountain terrain are more likely to have higher crash frequency. However, this result is inconsistent with Peng et al. (2012) and Park et al. (2014) which found that wider shoulder decreased crash frequency. The positive effect of shoulder width is potentially due to higher speed limits on the segments with wider shoulder increases vehicle speed. The positive effect of non-flat terrain on crash frequency is mainly due to driver’s limited sight. This result is consistent with the finding of Hosseinpour et al. (2014).

Table 5-1. Estimated parameters of Generalized Nonlinear Models for total crashes

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.0743</td>
<td>0.1027</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>$U_{TP}$ (truck percentage)</td>
<td>0.2013</td>
<td>0.0519</td>
<td>0.0001</td>
</tr>
<tr>
<td>$U_{AADT}$ (AADT)</td>
<td>0.6527</td>
<td>0.0300</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>$U_{SL-TP}$ (speed limit*truck %)</td>
<td>-0.0641</td>
<td>0.0172</td>
<td>0.0002</td>
</tr>
<tr>
<td>Length (km)</td>
<td>0.0208</td>
<td>0.0049</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Shoulder width (m)</td>
<td>0.5453</td>
<td>0.0361</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Terrain (1 = rolling and</td>
<td>0.1175</td>
<td>0.0602</td>
<td>0.0509</td>
</tr>
<tr>
<td>mountainous, 0 = flat)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Deviance = 1021.7004
Full Log Likelihood = -4957.6325
GNMs were also developed for truck-involved and non-truck-involved crashes as shown in Tables 5-2 and 5-3. Unlike total crashes, the nonlinear predictor of lane width have significant effects on crash frequency in truck-involved and non-truck involved. Also, differences in significant variables were observed between truck-involved and non-involved crashes. The effect of interaction between speed limit and truck percentage was significant in truck-involved crashes, but not in non-truck-involved crashes. The effect of terrain was significant in non-truck-involved crashes, but not in truck involved crashes.

**Table 5-2. Estimated Parameters of Generalized Nonlinear Models for truck-involved crashes**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.4303</td>
<td>0.1435</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>( U_{TP} ) (truck percentage)</td>
<td>-0.1748</td>
<td>0.0491</td>
<td>0.0004</td>
</tr>
<tr>
<td>( U_{LW} ) (lane width)</td>
<td>0.0471</td>
<td>0.0136</td>
<td>0.0006</td>
</tr>
<tr>
<td>( U_{AADT} ) (AADT)</td>
<td>0.7531</td>
<td>0.0458</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>( U_{SL\times TP} ) (speed limit\times truck %)</td>
<td>0.1090</td>
<td>0.0400</td>
<td>0.0064</td>
</tr>
<tr>
<td>Length (km)</td>
<td>0.0281</td>
<td>0.0047</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Number of streams</td>
<td>-0.2643</td>
<td>0.0449</td>
<td>0.0059</td>
</tr>
<tr>
<td>Shoulder width (m)</td>
<td>0.5349</td>
<td>0.0265</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Deviance = 1034.4640
Full Log Likelihood = -4168.1459

**Table 5-3. Estimated parameters of Generalized Nonlinear Models for non-truck-involved crashes**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.8634</td>
<td>0.2231</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>( U_{TP} ) (truck percentage)</td>
<td>0.0656</td>
<td>0.0154</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>( U_{AADT} ) (AADT)</td>
<td>-0.7066</td>
<td>0.0490</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>( U_{LW} ) (lane width)</td>
<td>0.0026</td>
<td>0.0012</td>
<td>0.0255</td>
</tr>
<tr>
<td>Length (km)</td>
<td>0.0172</td>
<td>0.0052</td>
<td>0.0009</td>
</tr>
<tr>
<td>Shoulder width (m)</td>
<td>0.5048</td>
<td>0.0432</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Number of streams</td>
<td>0.4352</td>
<td>0.1059</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Terrain (1 = rolling and mountainous, 0 = flat)</td>
<td>0.1645</td>
<td>0.0638</td>
<td>0.0099</td>
</tr>
</tbody>
</table>

Deviance = 1049.0406
Full Log Likelihood = -4389.8066
Effects of some nonlinearizing link functions are negative although they already captured nonlinear effects. Since nonlinearizing link functions were derived using the relationship between the crash rate and single variable only, they may not capture the correlations of the variable with the other variables.

5.3. Result of GLM

To evaluate the performance of GNMs in Section 5.2, GLMs were also developed for total, truck-involved and non-truck-involved crashes. Unlike GNMs, GLMs contain only linear predictors. The results of GLM are shown in Tables 5-4 to 5-6. Model fits were compared between GLM and GNM as shown in Table 5-7.

Table 5-4. Estimated parameters of Generalized Linear Models for total crashes

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.6951</td>
<td>0.1049</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Truck Percentage</td>
<td>-0.0448</td>
<td>0.0177</td>
<td>0.0114</td>
</tr>
<tr>
<td>AADT</td>
<td>0.0142</td>
<td>0.0007</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>speed limit*truck %</td>
<td>0.0005</td>
<td>0.0002</td>
<td>0.0142</td>
</tr>
<tr>
<td>Length (km)</td>
<td>0.0195</td>
<td>0.0048</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Shoulder width (m)</td>
<td>0.5873</td>
<td>0.0381</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Deviance = 1023.0279
Full Log Likelihood = -4959.2365

Table 5-5. Estimated Parameters of Generalized Linear Models for truck-involved crashes

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.5559</td>
<td>0.2257</td>
<td>0.0138</td>
</tr>
<tr>
<td>Truck percentage</td>
<td>0.0073</td>
<td>0.0024</td>
<td>0.0025</td>
</tr>
<tr>
<td>Lane width</td>
<td>0.3330</td>
<td>0.0631</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>AADT</td>
<td>0.0137</td>
<td>0.0046</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Length (km)</td>
<td>0.0265</td>
<td>0.0046</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Shoulder width (m)</td>
<td>0.5518</td>
<td>0.0349</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Deviance = 1035.0728
Full Log Likelihood = -4182.2624
Table 5-6. Estimated parameters of Generalized Linear Models for non-truck-involved crashes

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.3985</td>
<td>0.2608</td>
<td>0.1265</td>
</tr>
<tr>
<td>Truck percentage</td>
<td>-0.0161</td>
<td>0.0027</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>AADT</td>
<td>0.0117</td>
<td>0.0677</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Lane width</td>
<td>0.3017</td>
<td>0.0677</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Length (km)</td>
<td>0.0182</td>
<td>0.0052</td>
<td>0.0004</td>
</tr>
<tr>
<td>Shoulder width (m)</td>
<td>0.5598</td>
<td>0.0412</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Number of streams</td>
<td>0.6430</td>
<td>0.0990</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Terrain (1 = rolling and mountainous, 0 = flat)</td>
<td>0.6335</td>
<td>0.0307</td>
<td>0.0053</td>
</tr>
</tbody>
</table>

Deviance = 1048.7472  
Full Log Likelihood = -4390.0273

Table 5-7 shows that the values of full log likelihood were consistently higher for GNM than GLM. This indicates that GNM generally provides slightly better model fit than GLM. In spite of a small difference in model fit between GLM and GNM, GNM can better capture nonlinear effects of variables on crash frequency than GLM.

GNM also identified more statistically significant variables than GLM. For total crashes, terrain is not statistically significant in GLM unlike GNM. For truck-involved crashes, number of streams and the interaction of speed and truck percentage are not significant in GLM. Only for non-truck-involved crashes, GLM has the same number of significant variables as GNM.

Table 5-7. Comparison of model fit between GLM and GNM

<table>
<thead>
<tr>
<th>Full log likelihood</th>
<th>Total crashes</th>
<th>Truck-involved crashes</th>
<th>Non-truck-involved crashes</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLM</td>
<td>-4959.2365</td>
<td>-4182.2624</td>
<td>-4390.0273</td>
</tr>
<tr>
<td>GNM</td>
<td>-4957.6325</td>
<td>-4168.1459</td>
<td>-4389.8066</td>
</tr>
</tbody>
</table>
5.4. Result of GEE

In Generalized Estimating Equation (GEE) model with nonlinearizing link functions, both nonlinear effects of explanatory variables and temporal correlation among annual crash frequencies were taken into consideration unlike GLM and GNM. Thus, GEE is the considered as the best model among the three models. The model imports the crash frequency and traffic data for each year separately in 2004-2010.

Separate GEEs were developed using four correlation structures – independent, exchangeable, autoregressive and unstructured. Truck percentage, AADT, speed limit and lane width were included in the model as nonlinear predictors. However, GEEs with autoregressive and unstructured correlation structures could not be developed for truck-involved crashes. This is because relatively lower sample size of truck-involved crashes makes it difficult to estimate more complex correlation structures unlike total and non-truck-involved crashes. Thus, for truck-involved crashes, GEEs were developed using only simpler correlation structures (independent and exchangeable).

Table 5-8 shows the result of GEEs with the exchangeable correlation structure for total crashes. It was found that segment length, rolling and mountainous terrain, shoulder width and surface width have positive estimates. This implies the segments with longer length, wider shoulder and wider surface increase crash frequency. This is because segments with higher speed limit usually have wider shoulder and surface. Additionally, flat road segments have lower crash frequency compared to rolling and mountainous segments.
Table 5-8. Estimated parameters of Generalized Estimating Equation for total crashes (Exchangeable correlation structure)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Pr &gt;</th>
<th>Z</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.2719</td>
<td>0.0896</td>
<td>0.0024</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( U_{TP} ) (truck percentage)</td>
<td>0.1387</td>
<td>0.0343</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( U_{AADT} ) (AADT)</td>
<td>0.4321</td>
<td>0.0146</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( U_{SL<em>TP} ) (speed limit</em>truck %)</td>
<td>-0.0390</td>
<td>0.0100</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length (km)</td>
<td>0.0220</td>
<td>0.0010</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terrain (1 = rolling and</td>
<td>0.1065</td>
<td>0.0138</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mountainous, 0 = flat)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shoulder width (m)</td>
<td>0.4508</td>
<td>0.0266</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface width (m)</td>
<td>0.0408</td>
<td>0.0028</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

QIC= -644054.8896

In the GEE with independent correlation structure for truck-involved crashes, three linear and three non-linear predictors are statistically significant as shown in Table 5-9. Similar to total crashes, wider shoulder increases truck-involved crash frequency whereas wider median shoulder decreases truck-involved crash frequency. Unlike total crashes, terrain type was not significant for truck-involved crashes similar to the result of GNM (Table 5-2). It indicates that truck-involved crash frequency is not significantly affected by terrain of road segment. This is potentially because truck drivers ususally drive for a long time during a day and they tend to pay more attention to geographical conditions and become more cautious when segment is not flat than passenger car drivers.

Median width is also significant for truck-involved crashes unlike total crashes. Since segments with wider median have higher AADT, speed limit and truck volume, truck-involved crashes are more likely to occur on these segments. The positive effect of median width on crash frequency is not consistent with Bonneson and Pratt (2009), and Haleem et al. (2013) which found that wider median decreased crash frequency. This is potentially because higher speed limits on the segments with wider median in Ontario are associated with higher likelihood of crash occurrence.
On the other hand, wider median shoulder decreases truck-involved crash frequency. This is potentially because vehicles can overtake trucks more safely when there is an extra space in the median shoulder. A negative effect of wider median shoulder on crash frequency was also found in Bonneson and Pratt (2009), and Haleem et al. (2013).

Table 5-9. Estimated parameters of Generalized Estimating Equation for truck-involved crashes (Independent correlation structure)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Pr &gt;</th>
<th>Z</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.1295</td>
<td>0.0997</td>
<td>0.1937</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U_{TP}$ (truck percentage)</td>
<td>-0.2084</td>
<td>0.0178</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U_{SL}$(speed limit)*</td>
<td>-0.1170</td>
<td>0.0077</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U_{LW}$ (lane width)</td>
<td>0.0206</td>
<td>0.0055</td>
<td>0.0002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U_{SL<em>TP}$ (speed limit</em>truck %)</td>
<td>0.2042</td>
<td>0.0135</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td>-0.0047</td>
<td>0.0021</td>
<td>0.0229</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median width (m)</td>
<td>0.0087</td>
<td>0.0017</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shoulder width (m)</td>
<td>0.8977</td>
<td>0.0263</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median shoulder width (m)</td>
<td>-0.0459</td>
<td>0.0133</td>
<td>0.0005</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

QIC= -105752.0054

*Since speed limit was not significantly correlated with geometric factors in truck-involved crashes, it was included in the model.

However, a negative effect of segment length is counter-intuitive. The model result shows that truck-involved crash frequency is higher on shorter segments. This is potentially because shorter segments are associated with frequent change in AADT and geometry (Hauer et al., 2004). This indicates that truck drivers are more likely to be confused and make errors compared to car drivers in more complex traffic and road conditions.

Unlike truck-involved crashes, terrain is significant for non-truck-involved crashes as shown in the result of GEE with exchangeable correlation structure (Table 5-10). For non-truck-involved crashes, segments with wider surface, shoulder and median have higher crash frequency.
From Tables 5-8 to 5-10, additional significant variables were found to be significant in GEE compared to GNM. This is because GEE can capture temporal correlation among annual crash frequencies and annual variations in crash frequency and traffic volume.

Table 5-10. Estimated parameters of Generalized Estimating Equation for non-truck-involved crashes (Exchangeable correlation structure)

| Parameter | Estimate | Standard Error | Pr > |Z|
|-----------|----------|----------------|------|
| Intercept | 1.5774   | 0.0546         | <.0001|
| $U_{TP}$ (truck percentage) | 0.0568 | 0.0044         | <.0001|
| $U_{AADT}$ (AADT) | -0.7088 | 0.0184         | <.0001|
| $U_{LW}$ (lane width) | 0.0007 | 0.0003         | 0.0161|
| Length (km) | 0.0187 | 0.0010         | <.0001|
| Shoulder width (m) | 0.4513 | 0.0248         | <.0001|
| Median width (m) | 0.0160 | 0.0007         | <.0001|
| Surface width (m) | 0.0101 | 0.0030         | 0.0006|
| Terrain (1 = rolling and mountainous, 0 = flat) | 0.1255 | 0.0212         | 0.0011|

QIC= -337031.9587

Table 5-11 compares significant variables among total, truck-involved and non-truck-involved crashes. It was found that nonlinearizing link functions of truck percentage and length, and shoulder width were significant for all crash types.

Table 5-11. Statistically significant variables for different crash types

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total Crashes</th>
<th>Truck-involved crashes</th>
<th>Non-truck-involved crashes</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U_{TP}$ (truck percentage)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>$U_{SL}$ (speed limit)</td>
<td></td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>$U_{LW}$ (lane width)</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>$U_{AADT}$ (AADT)</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>$U_{SL \times TP}$ (speed limit*truck %)</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Length (km)</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Terrain</td>
<td></td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Median width (m)</td>
<td></td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Shoulder width (m)</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Surface width (m)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median shoulder width (m)</td>
<td></td>
<td>✔</td>
<td></td>
</tr>
</tbody>
</table>
However, there were some differences in significant variables between truck-involved and non-truck-involved crashes. For instance, median shoulder width was only significant for truck-involved crashes, but not for total and non-truck-involved crashes. This is potentially because wider median shoulder helps vehicles overtake trucks more safely. On the other hand, terrain type and surface width were not significant for truck-involved crashes unlike total and non-truck-involved crashes. This reflects that truck drivers tend to pay more attention to geographical conditions of roads than passenger car drivers and truck-involved crash frequency is not significantly different between flat and non-flat terrains.

However, annual precipitation was not significant in any of the GEE models. This is mainly because precipitations were not significantly different among different years.

The results of GEEs for the other correlation structures are shown in Appendix A. It was found that significant variables and their effects (positive and negative) were generally similar among different correlation structures. However, there were a few exceptions. In total crashes, surface width was not statistically significant for unstructured correlation unlike the other correlation structures. Also, in non-truck-involved crashes, surface width has negative effect for unstructured correlation whereas it has positive effect for the other correlation structures. It appears that the results are more consistent among independent, exchangeable and autoregressive correlation structures.

5.5. Model Fit of GEE

To identify the best correlation structure, goodness-of-fits of different GEEs were compared. The goodness-of-fit of GEE was determined based on QIC (Quasi-Akaike
Information Criterion) instead of full log likelihood, log likelihood and AIC (Akaike Information Criterion). The GEE with the best correlation structure has the lowest QIC value. Table 5-12 compares the QIC values among different GEE correlation structures and three crash types. The table shows that the exchangeable correlation structure has the lowest QIC value for total and non-truck-involved crashes whereas the independent correlation structure has the lowest QIC values for truck-involved crashes. This means that temporal correlation among annual total and non-truck-involved crash frequencies at a given segment exists and the correlation between crash frequencies in any two years is constant. In other words, the temporal correlation does not change over time. However, the temporal correlation among annual truck-involved crashes did not exist.

Table 5-12. QIC values of Generalized Estimating Equation for different crash types

<table>
<thead>
<tr>
<th>Structure</th>
<th>Total</th>
<th>Truck-Involved</th>
<th>Non-Truck-Involved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exchangeable</td>
<td>-644054.8896</td>
<td>-101986.6194</td>
<td>-337031.9587</td>
</tr>
<tr>
<td>Independent</td>
<td>-621831.1449</td>
<td>-105752.0054</td>
<td>-318706.6999</td>
</tr>
<tr>
<td>Autoregressive</td>
<td>-620153.6426</td>
<td>-*</td>
<td>-318309.7637</td>
</tr>
<tr>
<td>Unstructured</td>
<td>-516848.3628</td>
<td>-*</td>
<td>-229774.1795</td>
</tr>
</tbody>
</table>

*The models with complex correlation structures could not be developed due to insufficient data.

Beside the QIC examination, the model fit can be evaluated based on cumulative residual plot. Residuals are defined as the differences between the observed and fitted values of the response which have long been used for graphical and numerical examinations of the adequacy of regression models (SAS Institute, 2014). The residual plot is drawn for one independent variable at a time and then compared against the zero-residual line (Mohammadi et al., 2014).

The residual plots are drawn using the observed data and the simulated data. Due to variations in the simulated data, the residual plots also vary in each simulation run. The
model has a good model fit if the cumulative residual plots drawn using the observed and simulated data are similar.

Surface width in non-truck-involved crashes was taken as an example to demonstrate the relationship between cumulative residual and model fit. In Figure 5-15, the cumulative residuals of the observed data and the simulated data were shown in solid lines and dotted lines, respectively.
Figure 5-15. Cumulative residuals plot for surface width in different correlation structures for non-truck-involved crashes
Figure 5-15. Cumulative residuals plot for surface width in different correlation structures for non-truck-involved crashes (Continued)
The solid lines in Figures 5-15(a) are closer to the dotted lines than the solid lines in Figure 5-15(b)-(d). This indicates that exchangeable structure is the most suitable structure for non-truck-involved crashes. The cumulative residual plots for the other significant variables in GEEs with different correlation structures for total, truck-involved and non-truck-involved crashes are shown in Appendix B.
6. Conclusions and Recommendations

6.1. Conclusions

The objectives of this study are to develop the methods of predicting frequency of truck-involved and non-truck-involved crashes on road segments and identify unique characteristics of truck-involved crashes based on the comparison between truck-involved and non-truck-involved crashes. To capture nonlinear effects of variables, the variation of crash rate with the values of factors was observed and Generalized Nonlinear Models (GNMs) were developed. To capture temporal correlations among annual crash frequencies at a given site, Generalized Estimating Equation models (GEEs) with different correlation structures were also developed.

First, GLMs were developed and compared with GNMs. It was found that GNMs consistently provided slight better model fit than GLMs for three types of crashes - total, truck-involved and non-truck-involved crashes. The main advantage of GNMs is that the model can identify more statistically significant variables than GLMs. The result of GNMs shows that the significant variables are different for different type of crashes. Terrain type is statistically significant for non-truck involved crashes, but not for truck-involved crashes. This indicates that truck drivers are more cautious on non-flat terrains and trucks are less likely to be involved in crashes. The interaction between speed limit and truck percentage was also significant for total crashes and truck-involved crashes. This implies that the effect of truck percentage on crash frequency varies across different
speed limits. It appears that driver’s speed patterns are concurrently affected by truck percentage and speed limits.

GEE models were developed with 4 correlation structures - independent, exchangeable, autoregressive type 1 and unstructured. Truck percentage, AADT, speed limit and lane width were included in the model as nonlinear predictors. Therefore both nonlinear effects of explanatory variables and temporal correlation among annual crash frequencies were taken into consideration. The results of GEE models show that statistically significant variables and their effects on crash frequency were generally similar for different correlation structures. Among different correlation structures, exchangeable correlation structure showed the best model fit for total and non-truck-involved crashes whereas independent correlation structure showed the best model fit for truck-involved crashes as indicated by the lowest QIC value. This implies that total and non-truck-involved crash frequencies in two successive years at a given location are correlated but the correlation does not significantly vary over time. Thus, the effect of temporal correlation among annual crash frequencies must be considered in crash prediction using GEE. However, the correlation did not exist for truck-involved crashes.

The result of GEE shows that nonlinearizing link functions of lane width and truck percentage, and shoulder width were significant for all crash types. In particular, wider shoulder increases crash frequency since segments with higher speed limit and AADT generally have wider shoulder. For a similar reason, wider median also increases truck-involved and non-truck-involved crash frequency.
6.2. Contributions

This study contributes to better understanding of truck-involved crashes which have occurred less frequently (e.g., 8% of total crashes on provincial highways in Ontario) but caused relatively higher number of fatalities and severe injuries. In this regard, the study identified unique characteristics of truck-involved crashes from comparison of the statistically significant variables among different crash types. The result shows that median shoulder width was only significant for truck-involved crashes, but not for total and non-truck-involved crashes. On the other hand, terrain type and surface width were not significant for truck-involved crashes unlike total and non-truck-involved crashes.

The study also contributes to improvement of methodology for predicting crash frequency. The study considered nonlinear or non-monotonic effects of explanatory variables on crash frequency, which could not be reflected in conventional GLM. The study also accounted for temporal correlations among annual crash frequencies observed at the same location, which normally exist due to similarity of road geometric and traffic conditions. The study demonstrated that these temporal correlations are significant for non-truck-involved crashes, but not truck-involved crashes.

Based on the findings in this study, some countermeasures to reduce truck-involved crashes are suggested. For instance, the width of median shoulder is increased and speeding is more strictly regulated on the segments with wider shoulder and median. Also, given that crash rate significantly varies with truck percentage at different speed limits, speed limits are determined considering truck percentage.
6.3. Limitations and Recommendations

There are some limitations in this study. First, more general nonlinear relationships between the variables (e.g., lane width, truck percentage) and crash rate could not be derived due to a lack of crash frequency for some ranges of variables. Second, changes in factors not related to traffic and road geometry (e.g., legislative changes) could not be considered. These changes potentially affect crash frequencies and their temporal correlation. Third, the precipitation data obtained from weather stations may not reflect actual weather conditions at the locations of road segments. Thus, the study could not accurately capture the effect of weather on crash frequency. Lastly, due to complexity of identifying geographical locations of the roadway segments, spatial correlation could not be considered in this study.

In future studies, it is recommended that GEEs with nonlinearizing link functions are applied to the prediction of crash frequency for the other roadway types such as intersections and interchanges. It is also recommended that the conditions contributing to truck-involved crashes be investigated using disaggregate data including driver characteristics and vehicle performance.
References


Hu, W., 2010. Factors associated with median related crash frequency and severity. The Pennsylvania State University, University Park, Pennsylvania, USA.


Offei, E. and Young, R., 2014. Quantifying the Impact of Large Percent Trucks Proportion on Rural Freeways. Dept. of Civil and Architectural Engineering, Univ. of Wyoming, Laramie, W.Y., U.S.A.


Qi, Y., 2009. Effect of Truck Lane Restriction and Differential Speed Limit on Traffic Characteristics of Four-Lane Rural Freeways. Dept. of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA.


Sun, X. and Garber, N. J., 2002. Determining the Safety Effects of Differential Speed Limits on Rural Interstate Highways Using Empirical Bayes Method. No. UVACTS-14-5-36 Dept. of Civil Engineering, University of Virginia, Charlottesville, VA.


### Appendix A. Results of GEE Models

#### Table A-1. Estimated parameters of Generalized Estimating Equation for total crashes (Independent correlation structure)

| Parameter | Estimate | Standard Error | Pr > |Z| |
|-----------|----------|----------------|-------|---|
| Intercept | 0.2450   | 0.0944         | 0.0094|
| $U_{TP}$ (truck percentage) | 0.1392 | 0.0342 | <.0001|
| $U_{AADT}$ (AADT) | 0.4295 | 0.0134 | <.0001|
| $U_{SL*TP}$ (speed limit*truck %) | -0.0396 | 0.0100 | <.0001|
| Length (km) | 0.0208 | 0.0007 | <.0001|
| Terrain (1 = rolling and mountainous, 0 = flat) | 0.1108 | 0.0144 | <.0001|
| Shoulder width (m) | 0.4557 | 0.0284 | <.0001|
| Surface width (m) | 0.0411 | 0.0026 | <.0001|

QIC= -621831.1449

#### Table A-2. Estimated parameters of Generalized Estimating Equation for total crashes (Autoregressive correlation structure)

| Parameter | Estimate | Standard Error | Pr > |Z| |
|-----------|----------|----------------|-------|---|
| Intercept | 0.4137   | 0.0724         | <.0001|
| $U_{TP}$ (truck percentage) | 0.1220 | 0.0325 | 0.0002|
| $U_{AADT}$ (AADT) | 0.4238 | 0.0157 | <.0001|
| $U_{SL*TP}$ (speed limit*truck %) | -0.0360 | 0.0096 | 0.0002|
| Length (km) | 0.0221 | 0.0007 | <.0001|
| Terrain (1 = rolling and mountainous, 0 = flat) | 0.0642 | 0.0119 | <.0001|
| Shoulder width (m) | 0.3847 | 0.0200 | <.0001|
| Surface width (m) | 0.0430 | 0.0027 | <.0001|

QIC= -620153.6426

#### Table A-3. Estimated parameters of Generalized Estimating Equation for total crashes (Unstructured correlation structure)

| Parameter | Estimate | Standard Error | Pr > |Z| |
|-----------|----------|----------------|-------|---|
| Intercept | 1.9163   | 0.0489         | <.0001|
| $U_{TP}$ (truck percentage) | 0.1279 | 0.0112 | <.0001|
| $U_{AADT}$ (AADT) | 0.5615 | 0.0015 | <.0001|
| $U_{SL*TP}$ (speed limit*truck %) | -0.0397 | 0.0034 | <.0001|
| Length (km) | 0.0255 | 0.0003 | <.0001|
| Terrain (1 = rolling and mountainous, 0 = flat) | 0.0602 | 0.0098 | <.0001|
| Shoulder width (m) | 0.5379 | 0.0087 | <.0001|

QIC= -543001.0918
Table A-4. Estimated parameters of Generalized Estimating Equation for truck-involved crashes (Exchangeable correlation structure)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Pr &gt;</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.1194</td>
<td>0.0910</td>
<td>0.1895</td>
<td></td>
</tr>
<tr>
<td>$U_{TP}$ (truck percentage)</td>
<td>-0.2124</td>
<td>0.0177</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>$U_{SL}$ (speed limit)</td>
<td>-0.1192</td>
<td>0.0075</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>$U_{LW}$ (lane width)</td>
<td>0.0202</td>
<td>0.0056</td>
<td>0.0003</td>
<td></td>
</tr>
<tr>
<td>$U_{SL<em>TP}$ (speed limit</em>truck %)</td>
<td>0.2083</td>
<td>0.0136</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td>-0.0047</td>
<td>0.0021</td>
<td>0.0246</td>
<td></td>
</tr>
<tr>
<td>Median width (m)</td>
<td>0.0088</td>
<td>0.0017</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Shoulder width (m)</td>
<td>0.8947</td>
<td>0.0249</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Median shoulder width (m)</td>
<td>-0.0447</td>
<td>0.0136</td>
<td>0.0010</td>
<td></td>
</tr>
</tbody>
</table>

QIC= -101986.6194

Table A-5. Estimated parameters of Generalized Estimating Equation for non-truck-involved crashes (Independent correlation structure)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Pr &gt;</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.5260</td>
<td>0.0596</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>$U_{TP}$ (truck percentage)</td>
<td>0.0561</td>
<td>0.0043</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>$U_{AADT}$ (AADT)</td>
<td>-0.7055</td>
<td>0.0177</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>$U_{LW}$ (lane width)</td>
<td>0.0008</td>
<td>0.0003</td>
<td>0.0129</td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td>0.0176</td>
<td>0.0008</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Shoulder width (m)</td>
<td>0.4571</td>
<td>0.0267</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Median width (m)</td>
<td>0.0159</td>
<td>0.0007</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Surface width (m)</td>
<td>0.0107</td>
<td>0.0029</td>
<td>0.0002</td>
<td></td>
</tr>
<tr>
<td>Terrain (1 = rolling and mountainous, 0 = flat)</td>
<td>0.1292</td>
<td>0.0220</td>
<td>&lt;.0001</td>
<td></td>
</tr>
</tbody>
</table>

QIC= -318706.6999

Table A-6. Estimated parameters of Generalized Estimating Equation for non-truck-involved crashes (Autoregressive correlation structure)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Pr &gt;</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.7228</td>
<td>0.0557</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>$U_{TP}$ (truck percentage)</td>
<td>0.0472</td>
<td>0.0044</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>$U_{AADT}$ (AADT)</td>
<td>-0.7194</td>
<td>0.0186</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>$U_{LW}$ (lane width)</td>
<td>0.0005</td>
<td>0.0003</td>
<td>0.0926</td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td>0.0211</td>
<td>0.0006</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Shoulder width (m)</td>
<td>0.3741</td>
<td>0.0206</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Median width (m)</td>
<td>0.0168</td>
<td>0.0008</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Surface width (m)</td>
<td>0.0127</td>
<td>0.0028</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Terrain (1 = rolling and mountainous, 0 = flat)</td>
<td>0.0913</td>
<td>0.0185</td>
<td>&lt;.0001</td>
<td></td>
</tr>
</tbody>
</table>

QIC= -318309.7637
Table A-7. Estimated parameters of Generalized Estimating Equation for non-truck-involved crashes (Unstructured correlation structure)

| Parameter                   | Estimate  | Standard Error | Pr > |Z| |
|-----------------------------|-----------|                |      |   |
| Intercept                   | 3.7672    | 0.0216         | <.0001|
| $U_{TP}$ (truck percentage) | 0.0465    | 0.0011         | <.0001|
| $U_{AADT}$ (AADT)           | -0.7775   | 0.0028         | <.0001|
| $U_{LW}$ (lane width)       | 0.0017    | 0.0001         | <.0001|
| Length (km)                 | 0.0150    | 0.0002         | <.0001|
| Shoulder width (m)          | 0.4810    | 0.0120         | <.0001|
| Median width (m)            | 0.0189    | 0.0001         | <.0001|
| Surface width (m)           | -0.0196   | 0.0006         | <.0001|
| Terrain (1 = rolling and mountainous, 0 = flat) | 0.0938 | 0.0100 | <.0001 |

QIC = -229774.1795
Appendix B. Cumulative Residual Plots of GEE Models

![Cumulative Residual Plots](image)

(a) Exchangeable error structure

(b) Independent error structure

Figure B-1. Cumulative residuals plot for shoulder width in different correlation structures for total crashes
Figure B-1. Cumulative residuals plot for shoulder width in different correlation structures for total crashes (Continued)
Figure B-2. Cumulative residuals plot for terrain in different correlation structures for total crashes
(c) Autoregressive error structure

(d) Unstructured error structure

Figure B-2. Cumulative residuals plot for terrain in different correlation structures for total crashes (Continued)
Figure B-3. Cumulative residuals plot for terrain in different correlation structures for total crashes
Figure B-3. Cumulative residuals plot for terrain in different correlation structures for total crashes
Figure B-4. Cumulative residuals plot for length in different correlation structures for total crashes
Figure B-4. Cumulative residuals plot for length in different correlation structures for total crashes (Continued)
Figure B-5. Cumulative residuals plot for shoulder width in different correlation structures for truck-involved crashes.
Figure B-6. Cumulative residuals plot for median shoulder width in different correlation structures for truck-involved crashes
Figure B-7. Cumulative residuals plot for median width in different correlation structures for truck-involved crashes
Figure B-8. Cumulative residuals plot for length in different correlation structures for truck-involved crashes
(a) Exchangeable error structure

(b) Independent error structure

Figure B-9. Cumulative residuals plot for length in different correlation structures for non-truck-involved crashes
(c) Autoregressive error structure

(d) Unstructured error structure

Figure B-9. Cumulative residuals plot for length in different correlation structures for non-truck-involved crashes (Continued)
(a) Exchangeable error structure

(b) Independent error structure

Figure B-10. Cumulative residuals plot for shoulder width in different correlation structures for non-truck-involved crashes
(c) Autoregressive error structure

(d) Unstructured error structure

Figure B-10. Cumulative residuals plot for shoulder width in different correlation structures for non-truck-involved crashes (Continued)
Figure B-11. Cumulative residuals plot for median width in different correlation structures for non-truck-involved crashes
Figure B-11. Cumulative residuals plot for median width in different correlation structures for non-truck-involved crashes (Continued)
Figure B-12. Cumulative residuals plot for terrain width in different correlation structures for non-truck-involved crashes
Figure B-12. Cumulative residuals plot for terrain width in different correlation structures for non-truck-involved crashes (Continued)
NAME: Ran Ran

PLACE OF BIRTH: Chengdu, China

YEAR OF BIRTH: 1990

EDUCATION: Southwest Jiaotong University, Chengdu, China
2008-2012, Bachelor of Engineering

University of Windsor, Windsor, ON, Canada
2012-2015, Master of Applied Science