Transportation Research Part A 75 (2015) 30–41

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

Assessment of the effects of highway geometric design features on the frequency of truck involved crashes using bivariate regression

Chunjiao Dong [a,*], Shashi S. Nambisan [b], Stephen H. Richards [a], Zhuanglin Ma [c]

[a] Center for Transportation Research, The University of Tennessee, 600 Henley Street, Knoxville 37996, TN, USA
[b] Department of Civil & Environmental Engineering, The University of Tennessee, 319 John D. Tickle Building, Knoxville, TN 37996-2321, USA
[c] School of Automobile, Chang'an University, Nan Er Huan Zhong Duan, Xi'an 710064, Shaanxi, China

ARTICLE INFO

Article history:
Received 28 July 2014
Received in revised form 4 February 2015
Accepted 5 March 2015
Available online 28 March 2015

Keywords:
Large truck involved crashes
Crash frequency
Geometric design features
Bivariate ZINB model
Bivariate NB model

ABSTRACT

Given the enormous losses to society resulting from large truck involved crashes, a comprehensive understanding of the effects of highway geometric design features on the frequency of truck involved crashes is needed. To better predict the occurrence probabilities of large truck involved crashes and gain direction for policies and countermeasures aimed at reducing the crash frequencies, it is essential to examine truck involved crashes categorized by collision vehicle types, since passenger cars and large trucks differ in dimensions, size, weight, and operating characteristics. A data set that includes a total of 1310 highway segments with 1787 truck involved crashes for a 4-year period, from 2004 to 2007 in Tennessee is employed to examine the effects that geometric design features and other relevant attributes have on the crash frequency. Since truck involved crash counts have many zeros (often 60–90% of all values) with small sample means and two established categories, car-truck and truck-only crashes, are not independent in nature, the zero-inflated negative binomial (ZINB) models are developed under the bivariate regression framework to simultaneously address the above mentioned issues. In addition, the bivariate negative binomial (BNB) and two individual univariate ZINB models are estimated for model validation. Goodness of fit of the investigated models is evaluated using AIC, SBC statistics, the number of identified significant variables, and graphs of observed versus expected crash frequencies. The bivariate ZINB (BZINB) models have been found to have desirable distributional property to describe the relationship between the large truck involved crashes and geometric design features in terms of better goodness of fit, more precise parameter estimates, more identified significant factors, and improved predictive accuracy. The results of BZINB models indicate that the following factors are significantly related to the likelihood of truck involved crash occurrences: large truck annual average daily traffic (AADT), segment length, degree of horizontal curvature, terrain type, land use, median type, lane width, right side shoulder width, lighting condition, rutting depth (RD), and posted speed limits. Apart from that, passenger car AADT, lane number, and indicator for different speed limits are found to have statistical significant effects on the occurrences of car-truck crashes and international roughness index (IRI) is significant for the predictions of truck-only crashes.

* Corresponding author.

http://dx.doi.org/10.1016/j.tra.2015.03.007

1. Introduction

The number of total registered trucks in the United States has grown steadily during the 10-year period, from 94.9 million in 2003 to about 133.1 million in 2012, a 40% increase (FHWA, 2003–2012). Large trucks have made an important contribution to the growth of the national economy by facilitating the distribution of a large portion of the nation's products [more than 13 billion tons per year (USDOT/BTS, 2013)]. However, as truck transport has become more important, the safety issues associated with the large truck traffic, in terms of traffic crashes, injuries, and fatalities have become more evident. Indeed, large trucks are a serious safety concern. More than 3000 people are killed every year in large truck involved crashes. The yearly number of fatalities in large truck involved crashes is on the rise again after nearly a decade of declining. According to the most recent data from the National Highway Traffic Safety Administration, in 2012, 3921 people were killed and 104,000 people were injured in crashes involving large trucks and 83% of those killed and 76% of those injured were not occupants of the trucks (NHTSA, 2012). A significant proportion of deaths occur to passenger car occupants because the large sizes and heavy weights of large trucks transmit most of the crash force to the passenger cars.

Truck issues currently command a large share of the traffic safety interest of researchers, policy makers, and the public. In recent years, some studies have been performed to better identify and understand factors that contribute to the frequencies of large truck involved crashes. Geometric design features, environmental factors, traffic conditions, and driver performances have been investigated to determine how these factors influence the occurrences of truck crashes. However, most of the research focused on modeling truck crash counts as a whole. Given the differences observed in the characteristics associated with car-truck crashes and truck-only crashes, some researchers (Dong et al., 2014) have proposed using separate crash prediction models for these crashes. Separate predictive models have been developed to estimate the safety performances when, instead of total crash counts, one wishes to model specific types of crash counts, such as single-vehicle crashes versus multi-vehicle crashes (Geedipally and Lord, 2010), and crashes resulting in injuries and non-injuries (Khattak et al., 2002). These efforts indicated that developing two distinct models provide better predicting performances than developing models that combine both crash categories together. Though the separate models have been employed as a starting point for modeling the counts of specific types of crashes (as opposed to total crashes), researchers (Geedipally and Lord, 2010; Dong et al., 2014) have found that crash counts exhibit characteristics that make the application of the separate models problematic. Specifically, separate models cannot handle the correlation among two interdependence crash types. In this paper, the relationship between the geometric design features, pavement conditions, traffic factors, and crash frequencies are investigated by using bivariate regression models. Since truck involved crash counts have a significant amount of zeros, the NB and ZINB regressions are developed under the bivariate regression framework for jointly modeling car-truck and truck-only crashes simultaneously.

2. Literature review

Most of the recent research on improving truck traffic safety have focused on reducing the degree of injury sustained by those involved in truck crashes and a wide variety of methodological techniques have been applied to gain a thorough understanding of the factors that affect the degree of injury, given that a crash has occurred (Lyman and Braver, 2003; Chen and Chen, 2011; Lemp et al., 2011; Chang and Chien, 2013; Zhu and Srinivasan, 2011). These studies provide valuable insights into the safety of trucks. However, it is important to note that reducing crash frequencies and reducing crash-injury severities require different strategic approaches. With regards to the crash frequency study, the relationships between total crashes and geometric design features of road segments, such as horizontal curvature, vertical grade, lane width, and shoulder width have been studied in numerous previous studies [a thorough review and assessment of these studies can be found in a recent paper (Lord and Mannering, 2010)]. Compared to the predictive models developed for total crashes, the studies that investigated the effects of highway geometric design features on the occurrences of truck involved crashes are relatively limited. Over the past 20 years, only a few researchers have proposed predictive models specifically to analyze the frequency of crashes involving large trucks.

Blower et al. (1993) analyzed crash rates of heavy truck-tractors using log-linear methods. They investigated the effects of the number of trailers, road type, area type, and time of day on the rates of casualty and property-damage-only crashes. The results showed that differences between tractors with one and two trailers were not significant and characteristics of the operating environment were found to have significant effects on crash rates. In other words, the results showed that operating environment is more important in determining the risk of crash involvements than truck configuration. This study provided considerable interest both in its method and its results. However, log-linear models only demonstrate association between variables and the investigated variables are all treated as "response variables". If one or more variables are treated as explicitly dependent and others as independent, in case of crash frequency analysis, then the log-linear models are not appropriate and the results obtained from these models are questionable. Subsequent studies have used more flexible methods and have achieved slightly different conclusions from Blower et al. (1993). A case-control study was performed to address whether multiple-trailer combination vehicles were overinvolved in crashes on interstate highways in Indiana by Braver et al. (1997). The day of week, time of day, urban/rural area, and specific highway were identified as significant factors using logistic regression model. The results indicated that the crash risk of double-trailer trucks might be greater under less favorable operating conditions and road environments compared to the single-trailer trucks. The major limitation of this study is that the control data were not obtained for the variables that might confound the relationship between truck configuration and crash risk as the authors pointed out at the end of their paper. In addition, using the logistic regression as the investigation method ignores its implicit linear assumption in terms of the logit function versus the independent variables, which is fairly unreasonable for describing the random and discrete crash events on the road.

The unsatisfactory property of logistic regression models has led to the investigation of the Poisson, NB, and zero-inflated Poisson (ZIP) regression models. Miaou (1994) evaluated the performances of those regression models in establishing the relationship between truck crashes and geometric design features. Miaou (1994) also investigated the sensitivity of including short road segment in those three models. The analysis variables included yearly dummy variables, annual average daily traffic (AADT) per lane, horizontal curvature, vertical grade, and deviation of paved inside shoulder width. The estimation results revealed that the NB regression model using the moment estimation is sensitive to the inclusion of short road segment. Though the evaluation results indicated that, under the maximum likelihood (ML) method, no particular model outperformed the other two models in terms of the estimated relative frequencies of truck crashes, the suggestion was that the ZIP regression model appeared to be a potential alternative model when the crash count exhibit excess zeros. Daniel et al. (2002) developed the Poisson and NB models to identify factors that contribute to truck crashes on roadways influenced by signalized intersections. The results indicated that the number of signalized intersections, segment length, AADT, length of horizontal curve, length of vertical curve, number of lanes, number of signals within the segment, crest curve grade rate, and pavement width have significant effects on the frequencies of crashes involving trucks. Schneider et al. (2009) proposed a NB model analyzing the effects of horizontal curvature on truck crashes. The full Bayesian method was used to improve the model performances. The results indicated that a significant increase in truck crashes is related to the increase in the value of the horizontal curve, truck average daily traffic (ADT), passenger vehicle ADT, and the degree of curvature. While these various methodological applications have undoubtedly provided new insights, the impact of highway geometric design features, along with traffic factors and pavement characteristics on the frequencies of truck involved crashes are not fully understood or accounted for. In addition, the relative performances of these models in establishing such relationships have not been fully addressed and evaluated.

From the literature review, we can conclude that the most commonly used predictive models for truck involved crashes are Poisson and NB models. Despite their broad applicability in a variety of situations, the limitation of Poisson models, which requires the mean equals to the variance, makes the application of Poisson regression problematic. Though the NB regression models have desirable distributional property, in terms of better goodness of fit, to describe the relationship between geometric design features and crash frequencies, the models do have limitations, especially, their inability to handle the data characterized by the small simple sizes. Because the occurrences of truck crashes are typically sporadic across road segments, a large number of road segments had no reported truck crashes during the observed period. For instance, in a study by Miaou and Lum (1993), over 80% of the road segments had no reported truck crashes during a one-year period. In most cases, the researchers are faced with a problem of dealing with more zeros than conventional Poisson and NB models can handle. Regarding the observed data that are characterized by a significant amount of zeros, in most of the cases, zero-inflated regressions offer better goodness of fit and provide some new insights. Some researchers have criticized the application of zero-inflated models in highway safety, for instance, Lord et al. (2005, 2007) argued that, because the safe states have a long-term mean equal to zero, these models are not appropriate to descript the crash count generating process. Nonetheless, the zero-inflated regression models, especially, the ZINB models, have been broadly applied to traffic safety analysts (Shankar et al., 1997; Carson and Mannering, 2001; Lee and Mannering, 2002; Kumara and Chin, 2003; Shankar et al., 2003). In such contexts, we employed the ZINB model as the baseline of the proposed model, since our data set includes more zeros than the conventional models can handle. This study examines the impact of geometric design features, as well as traffic factors and pavement conditions on the frequency of car-truck and truck-only crashes using crash counts obtained from Tennessee during the period of 2004 to 2007. The ZINB model is developed under the bivariate regression framework to address the correlation issues, since the car-truck and truck-only crashes are not independent in nature. For comparison, a BNB model and two individual univariate NB models are employed for modeling the same crash counts. The performances of the proposed models are evaluated using AIC, SBC statistics, the number of identified significant variables, and graphs of observed versus expected crash frequency relationship.

3. Methodology

Passenger cars and large trucks share the same driving environment, but differ in size, dimensions, weight, and operating characteristics. Therefore, to provide useful guidelines for crash prevention, it is critically important to examine crash frequencies categorized by collision vehicle types. In this study, truck involved crashes have been distinguished as car-truck crashes and truck-only crashes. Truck-only crashes include single truck collisions and truck to truck collisions and car-truck crashes include collisions involving at least one car and one truck. Since truck involved crash data have a significant amount zeros, the ZINB regressions have been employed as the baseline of the proposed model to identify and understand factors that contribute to truck crashes. Given certain numbers of truck crashes, because the counts of truck-only crashes and the counts of car-truck crashes are interdependent, BZINB models are developed for jointly modeling those crashes. Since the BNB distribution is the element to construct the BZINB regression model, BNB distributions are presented first.

3.1. BNB distributions

The NB models can handle over-dispersion in the crash counts by assuming that the Poisson parameter follows a gamma probability distribution. The particular NB regression model employed in this study has the following form:

$$p(y_{ij}) = \frac{\Gamma(\lambda_{ij}/\sigma + y_{ij})}{\Gamma(\lambda_{ij}/\sigma)\Gamma(y_{ij}+1)}\left(\frac{1}{1+\sigma}\right)^{\lambda_{ij}/\sigma}\left(\frac{\sigma}{1+\sigma}\right)^{y_{ij}} \tag{1}$$

where $\Gamma(\cdot)$ is the gamma function, y_{ij} is the crash number of crash type j for roadway segment i and $E[y_{ij}] = \lambda_{ij} = EXP(\mathbf{\beta X}_i + \varepsilon)$. $EXP(\varepsilon)$ is a gamma-distributed error term with mean 1 and variance α^{-1}. Generally, the parameter α^{-1} is referred to as the over-dispersion parameter.

Though the NB distributions have received considerable attention in the literature of crash frequency modeling, it remains rather inflexible, in the case that there are two crash types need modeling. In such contexts, the BNB models are appropriate since they can explicitly consider the correlation between the interdependent crash counts. The BNB distributions have been proposed under various forms of mechanism. We follow the usual notation for bivariate data let y_{ij}, $i = 1, 2, \ldots, n$, $j = 1, 2$ denotes the crash counts for individual segment i with crash type j. In this specific study, y_{i1} and y_{i2} are car-truck crashes and truck-only crashes, respectively, for segment i. Note that y_{ij} can be represented by $y_{ij} = z_{ij} + u_i$, where z_{ij} and u_i have independent NB distributions with mean λ_{ij} and λ_{i0}, respectively. The property of NB distribution indicates that the sum of two independent NB distributions is again a NB distribution only if the two distributions share the common parameter $\sigma = \lambda_{ij}{\alpha_z}^{-1} = \lambda_{i0}{\alpha_u}^{-1}$, where α^{-1} is the overdispersion parameter. Hence, the joint probability function of the BNB distribution for segment i is obtained after integration (Winkelmann, 2008; Anastasopoulos et al., 2012):

$$p(Y_1 = y_{i1}, Y_2 = y_{i2}) = \sum_{k=0}^{\min(y_{i1},y_{i2})} p(k)\prod_{j=1}^{2} p(y_{ij}-k) \tag{2}$$

where $p(k)$ is the NB probability distribution for k.

The parameters of BNB model can be estimated by ML method, the moment estimation method, and Bayesian method. In this study, the ML estimation method is adopted for parameter estimation, since it is commonly used for estimating the parameters in the BNB model. The detailed information related to the ML estimation of the BNB regression model and the calculation of associated statistics is found in Famoye (2010).

3.2. BZINB regression model

Apart from BNB regression model, we proposed a BZINB model for car-truck and truck-only crash analysis. The BZINB is developed as a mixture of two univariate NB distributions, a BNB distribution, and a point mass at (0, 0) (Wang et al., 2003). Then

$$\begin{aligned}
(Y_1, Y_2) &\sim (0,0) && \text{with probability } p_0 \\
&\sim (NB(\lambda_1), 0) && \text{with probability } p_1 \\
&\sim (0, NB(\lambda_2)) && \text{with probability } p_2 \\
&\sim BNB(\lambda_{10}, \lambda_{20}, \lambda_{00}) && \text{with probability } 1 - p_0 - p_1 - p_2
\end{aligned} \tag{3}$$

The probability mass function of the BZINB is given by

$$P(Y_1 = 0, Y_2 = 0) = p_0 + p_1\exp(-\lambda_1) + p_2\exp(-\lambda_2) + (1 - p_0 - p_1 - p_2)\exp(-\lambda) \tag{4}$$

$$P(Y_1 = y_{i1}, Y_2 = y_{i2}) = (1 - p_0 - p_1 - p_2)\exp(-\lambda)\sum_{j=0}^{\min(y_{i1},y_{i2})}\frac{\lambda_{10}^{y_{i1}-j}\lambda_{20}^{y_{i2}-j}\lambda_{00}^{j}}{[(y_{i1}-j)!(y_{i2}-j)!j!]} \qquad y_{i1}\neq 0 \text{ or } y_{i2}\neq 0 \tag{5}$$

where λ_{10}, λ_{20}, and λ_{00} are the parameters of BNB distribution and $\lambda = \lambda_{10} + \lambda_{20} + \lambda_{00}$; λ_1 and λ_2 are the means of independent univariate NB distributions, and $\lambda_1 = \lambda_{10} + \lambda_{00}$, $\lambda_2 = \lambda_{20} + \lambda_{00}$. To describe the relationship between the crash frequency and relevant attributes, we define the following regression relationship to link the BZINB parameters to the exploratory variables

$$\lambda_{ij0} = \exp(\beta_j \mathbf{X}_i + \varepsilon) \tag{6}$$

$$p_{ij} = \exp(\gamma_j \mathbf{X}_i) \Bigg/ \left[1 + \sum_{j=0}^{2}\exp(\gamma_j \mathbf{X}_i)\right] \tag{7}$$

where $\beta_j = (\beta_{j0}, \beta_{j1}, \ldots, \beta_{jK})$, and $\gamma_j = (\gamma_{j0}, \gamma_{j1}, \ldots, \gamma_{jK})$ are vectors of coefficients, $j = 0,1,2$; $\mathbf{X}_i = (1, x_{i1}, x_{i2}, \ldots, x_{iK})'$, and x_{ik} is the indicator variable.

The ML estimation and other possible variations of the BZINB models have been discussed in Wang (2003). In this study, the ML method is also used to estimate the unknown parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ and their standard deviations of the BZINB models. Akaike's information criterion (AIC) and Schwarz Bayesian criterion (SBC) [also known as Bayesian information criterion

(BIC)] are employed to identify whether the BNB or BZINB offers a better statistical fit. Since the penalties are assigned to models with large numbers of parameters in AIC and BIC, the smaller the value of AIC and BIC, the better the model. In addition, goodness of fit of the investigated models is evaluated using graphs of observed versus expected crash frequencies.

4. Data description

The crash data collected from Tennessee state routes through the crash record information system are used to examine the effects of geometric design features and other relevant attributes on the frequency of large truck involved crashes. Since the crash data provide details about the type of vehicles as well as the number of vehicles involved in the collision, collisions involving at least one car and one truck are considered for car-truck crashes and only large truck involved collisions are considered for truck-only crashes. Information on roadways includes geometric design features such as horizontal curves, shoulder width, median type, posted speed limits, and traffic characteristics such as traffic volume and percentage of trucks are obtained from the road inventory records maintained by the Tennessee Department of Transportation (TDOT). Detailed roadway pavement information such as international roughness index (IRI) and rutting depth (RD) is obtained from the Pavement Management System (PMS). The geometric, traffic, and pavement characteristics are linked to the crash data through the common variable *id_number*. Extensive data processing including cleaning, consistency checks, and data reduction is undertaken. Finally, a total of 1,310 highway segments (with an average segment length of 0.795 miles and a total of 1041.637 miles) are available for analysis. Those segments have homogenous characteristics because the segments begin and end when characteristics change. The data set has a total of 14,428 reported crashes for a 4-year period, from 2004 to 2007. We use two years as the analysis period, since the sample size of one year truck-only crashes is very small. In addition, the pavement data are measured once every two years on Tennessee state routes. It is useful to note that there is no significant difference between two years or one year as the analysis period based on the computing results of correlation coefficients between crash frequencies. To examine the influence of every-two-year changes on highway geometric design features, and especially, yearly traffic condition changing, we use two years as the analysis period, which means the same road segments, even if nothing had changed, is considered as two independent segments—one for every two years from 2004 to 2007. Consequently, the investigated sample contains 1364 car-truck crashes and 423 truck-only crashes with an approximately 76.33–23.67% split. Individual segment experienced crashes involving large truck from 0 to 9 crashes in two years. Regarding car-truck crashes, of the sampled segments, 1766 (67.40%) had zero crash, 597 (22.79%) had one crash, 159 (6.07%) had two crashes, and 98 (3.74%) had more than two crashes. Regarding truck-only crashes, in the 2620 segment samples, 2317 (88.44%) had zero crashes, 225 (8.59%) had one crash, 46 (1.76%) had two crashes, and 32 (1.22%) had more than two crashes. Table 1 provides relevant descriptive statistics for the crash data and key explanatory variables.

Table 2 provides the correlation matrix between car-truck and truck-only crashes and the zero sample proportion. As seen in Table 2, though the car-truck crashes and truck-only crashes are not highly correlated, they cannot be assumed to be independent either. The correlation between car-truck and truck-only crashes could come from several possible sources,

Table 1
Descriptive statistics of independent variables and crash data.

Variable name	Variable description	Mean	Std. Err.	Min.	Max.
Car-truck	Number of car-truck crashes in an segment per two years	0.52	0.77	0	9
Truck-only	Number of truck-only crashes in an segment per two years	0.16	0.38	0	4
ThAADT_C	Thousand passenger car AADT per lane	7.27	6.59	0.73	29.98
ThAADT_T	Thousand truck AADT per lane	0.68	0.47	0.12	2.50
Segment_L	Segment length (miles)	0.80	1.43	0.01	12.71
Horizontal_C	Degree of horizontal curvature	1.57	3.34	0	14
Terrain_T	Indicator for terrain type (2 for mountainous; 1 for rolling; 0 for flat)	0.99	0.31	0	2
Land_U	Indicator for the predominate land use (3 for industrial; 2 for residential; 1 for commercial; 0 for rural)	0.74	0.69	0	3
Meidan_T	Indicator for median type (2 for raised median; 1 for two-way left turn lane (TWLTL); 0 for no medians)	0.61	0.80	0	2
Lane_N	Number of lanes (2 for 6 lanes; 1 for 4 lanes; 0 for 2 lanes)	1.14	1.07	0	3
Lane_W	Indicator for lane width (1 for 12 ft; 0 for less than 12 ft)	0.50	0.67	0	2
Shoulder_T	Indicator for shoulder type (2 for pavement; 1 for gravel; 0 for dirt)	1.59	0.67	0	2
Shoulder_Wl	Left side shoulder width (ft)	3.07	3.09	0	10
Shoulder_Wr	Right side shoulder width (ft)	3.77	3.54	0	11
ILLUM	Indicator of roadway lighting on the roadway segment (1 have lighting; 0 no lighting)	0.63	0.48	0	1
RD	Rutting depth (in.)	0.13	0.06	0.04	0.53
IRI	International roughness index (in./mile)	94.39	37.57	27.55	295.45
Speed_L	Posted speed limit (mph)	43.58	8.87	2	65
Truck_SL	Truck speed limit (mph)	42.80	8.63	2	65
DSL	Indicator for different speed limits (1 the posted truck speed limit is different from the posted passenger car speed limit; 0 no differences)	0.09	0.29	0	1
No. of crashes		1787			

Table 2
Characteristics of the car-truck and truck-only crash data.

Crash types	Correlation coefficient		Zero samples
	Car-truck crashes	Truck-only crashes	
Car-truck crashes	1	0.534	1766 (67.40%)
Truck-only crashes	0.534	1	2317 (88.44%)

e.g. omitted variables, sharing the same geometric design features, and homogeneous highway environment. In cases of correlation, the bivariate regression model typically fits better than a standard univariate regression model. Because of this, the bivariate regression models need to be used for analyzing two crash types simultaneously. Table 2 also shows that of the sampled 2620 segments, 1766 had no car-truck crashes and 2317 had no truck-only crashes. Since the data contains a significant amount of zeros, the models that can handle crash counts characterized by a significant amount of zeros, such as ZINB, should be performed under the bivariate regression framework. In addition, in cases of the preponderance of zeros observed, the main concern is the overdispersion issues. The literature (Lord and Mannering, 2010) has indicated that, if the data exhibit overdispersion characteristics, the NB model provides sufficient fit to the data. Hence, the BNB model was proposed as an alternative model to describe the relationship between the geometric design features and occurrences of large truck crashes as well.

5. Modeling results

5.1. Model performance evaluation

The performances of BZINB and BNB models are presented in Table 3. To evaluate the performances of the proposed models, two individual univariate ZINB models were estimated as the comparison methods. The AIC tests indicate that the BZINB models provide the best overall fit relative to the BNB model and two individual univariate ZINB models. Also, the SBC tests comparing the investigated models indicate that the BZINB models are statistically superior. Though the BZINB model is more flexible than the BNB model, the interpretation of the BZINB model can be difficult. For instance, the expected number of crashes is related to the covariates $\mathbf{X}_i$ in a much more complicated way in the BZINB model than that in the BNB model, and it is not easy to see how an increase in $\mathbf{X}_i$ would increase or decrease the mean in the BZINB model. Apart from the AIC and SBC statistics, we count the number of identified significant variables in each investigated model and the results show that the BZINB model can identify more significant variables with better goodness of fit than the other models. Compared to the individual ZINB models, the BZINB model can identify more significant variables than summing up the results from two individual univariate ZINB models. This is a very important and helpful feature of the BZINB model, since the BZINB model can improve our understanding of factors affecting the crash frequency.

To further evaluate the performances of proposed models, predictions from the investigated models are compared for the different crash groups categorized by the collision vehicle type. Our data set contained 422 roadway segment samples, and the analysis period was two years (2008–2009). Of the sampled segments, 140 (33.18%) had no truck involved crash at all, 185 (43.84%) had no car-truck crashes, and 310 (73.46%) had no truck-only crashes. Of the 282 truck involved crash segment samples, just 45 (15.96%) segment samples have truck-only crashes and 170 (60.28%) have car-truck crashes. Both car-truck crashes and truck-only crashes occurred exclusively in only 67 (23.76%) of the 282 sampled segments.

Results in Table 3 show that the predicted number of the crash-free segments from the BZINB model is very close to the observed value of 140. A very important characteristic of the BZINB model is that it can provide a good estimation of the probability that the segment is in a certain zero group. For the crash-prone propensity states in terms of car-truck and

Table 3
Model comparison between the BZINB and BNB regressions.

Model	Goodness of fit		Identified significant variables		Crash predictions (expected versus observed)			MAPE (%)
	AIC	SBC	Car-truck model	Truck-only model	Crash free*	Car-truck crashes	Truck-only crashes	
BZINB	1405.82	1494.72	18 (11)**	16 (10)	163 (16.43%)***	448 (−1.63%)	166 (−0.62%)	14.09
BNB	1840.92	1874.73	12	9	178 (27.14%)	465 (2.27%)	170 (1.56%)	17.59
ZINB model for car-truck crashes	2082.31	2111.11	12 (7)	–	224 (21.08%)	432 (−5.05%)	–	24.13
ZINB model for truck-only crashes	2295.32	2319.11	–	8 (5)	377 (21.61%)	–	153 (−8.38%)	29.14

* Values in the column represent the number of segments which are crash free.
** Values in parentheses represent the number of significant variables in the Logistic regression part.
*** Values in parentheses represent the percentage difference (%) between the observed and predicted values.

truck-only crashes, Table 3 suggests that the prediction biases are greater for the BNB analysis. This result may be caused by the low sample-mean value problem that haunts many studies using the NB model. Though BNB models in many cases fit the data well, sometimes zero-inflated regression models provide a better fit. The two individual univariate ZINB models, one for car-truck crashes and one for truck-only crashes, have the greatest prediction bias. The results are expected, since those two individual univariate models cannot handle the correlation issues. In the estimated BZINB and BNB models, the correlation coefficients between car-truck and truck-only crash counts are 0.140 and 0.164, which, unfortunately, are not statistically significant. Based on the descriptions of the correlation effects earlier in the study, the bivariate regressions are expecting to yield a superior outcome because the crash counts by collision vehicle types on the same segment of roadway are found to be correlated with one another as shown in Table 2. Note that this is a theoretical point, but rather an empirical one. In other words, where potential correlation exists, it should be addressed and modeled. The last column of Table 3 gives the Mean Absolute Percentage Error (MAPE) estimations of each model. Based on the MAPE, the BZINB model predicts better than the BNB and two individual univariate ZINB models. We hypothesize that the BZINB model better addresses the issue of excess zero counts in correlated data. Since the BZINB and BNB models have better performances, the estimated results of these two models are further investigated and discussed. The mean-predicted over the actual values for the two models are shown in Fig. 1. The results indicate that the BZINB model provides better predictions relative to the BNB model. The predicted values are closer to the actual values in the BZINB model.

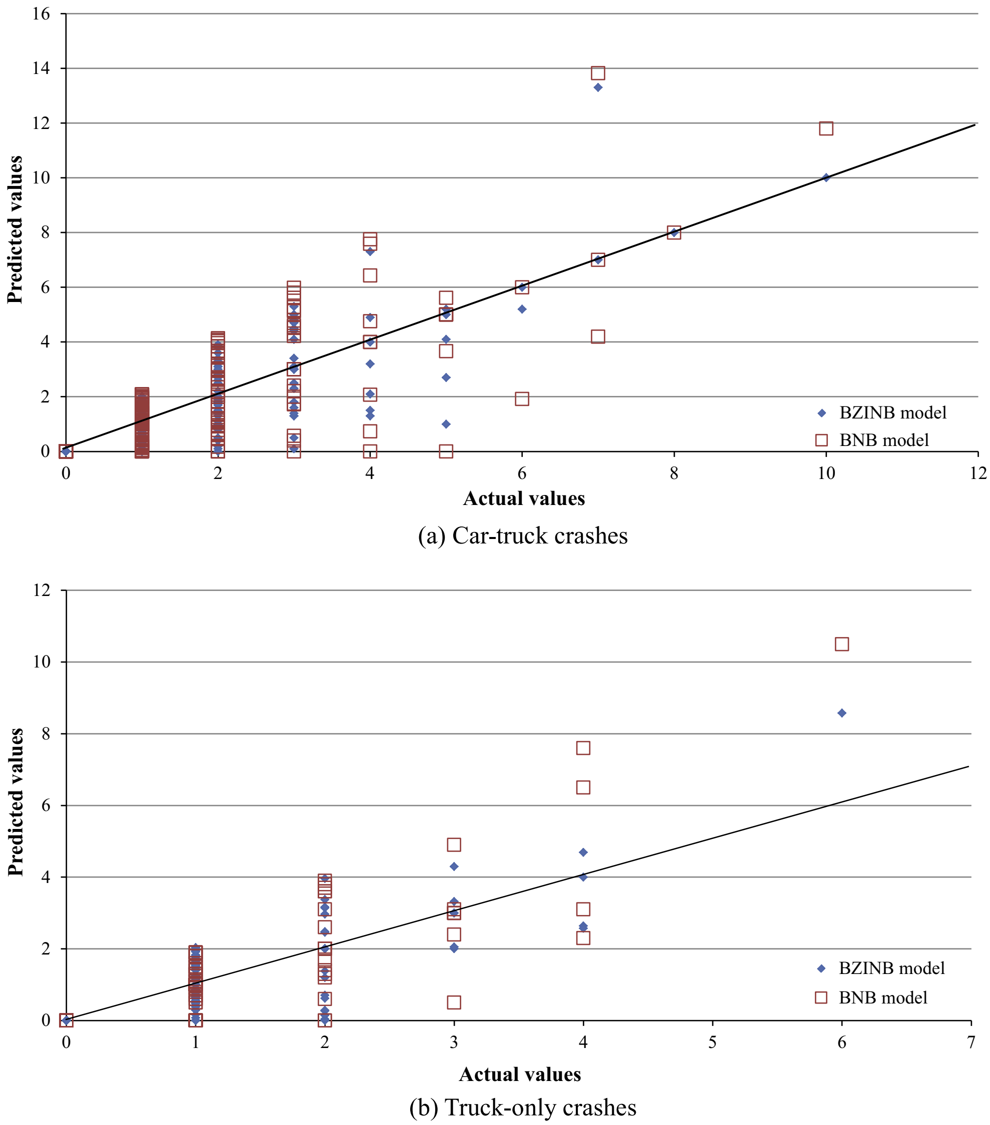

Fig. 1. Comparison between the predictions and the observations (the straight line indicates the equivalence of mean-predicted and actual value).

5.2. Results and discussion

Tables 4 and 5 provide the parameter estimates, stand errors, and *t*-statistics for the car-truck and truck-only crash count models under the BZINB and BNB frameworks, respectively. The estimates of overdispersion parameters in Tables 4 and 5 are positive and statistically significant at the 5% confidence level, indicating that the crash counts are overdispersed. This confirms that the appropriateness of the NB and ZINB models compared with the Poisson and ZIP regression models, respectively. The results show that BZINB model and BNB model provide consistent results in terms of the direction of the significant variables. The comparison of the estimated parameters and *t*-statistics in Tables 4 and 5 for the BZINB and BNB models suggests not only that the conclusions reached regarding the significance level of the relationships between large truck crash involvements and the traffic, pavement condition, and geometric design feature variables are quite consistent, but also that most of the estimated coefficients have the expected algebraic sign. However, the uncertainty estimates from the BZINB model are smaller than those from the BNB model. This finding confirms that BZINB model leads to more precise parameter estimates compared with BNB model by accounting for extra zeros in bivariate crash counts. In addition, the BZINB model identified more significant factors than the BNB model did. Hence, the BZINB model provides more useful information than the BNB regression model does.

Table 4 shows that the coefficients sometimes change from significant to insignificant for different crash types. This result indicates that some variables may have a different effect given the involved collision vehicle types. Though the significant characteristics of effect change for different crash types, the sign of significant coefficients are consistent throughout the models, indicating robust directions of effect for the significant variables. The result is expected since car-truck crash count and truck-only crash count are closely correlated. Because the BZINB model provides a better model goodness of fit, the results of the BZIBN model are further interpreted in the following sections.

Regarding the interpretation of the models' output, the coefficient estimates of the model parameters show how these independent variables are associated with the occurrences of large truck crashes. For the BZINB model, a positive sign of the estimated parameters in the NB regression part indicates increased crash frequency with increase in the value of the independent variables, or a decrease if the coefficient is negative. For instance, the estimated parameters of passenger car

Table 4
Parameter estimations of the BZINB model.

Parameter	Car-truck crash				Truck-only crash			
	Estimate	Std. Err.	*t* Value	Pr > |*t*|	Estimate	Std. Err.	*t* Value	Pr > |*t*|
Negative binomial regression part								
Intercept	−11.84	0.95	−12.48	<.0001	−12.29	1.04	−11.76	<.0001
ThAADT_C	0.37	0.11	3.18	0.0015				
ThAADT_T	3.29	0.34	9.73	<.0001	3.03	0.32	9.33	<.0001
Segment_L	0.94	0.13	7.26	<.0001	1.03	0.04	23.15	<.0001
Horizontal_C	0.22	0.04	5.41	<.0001	0.29	0.03	10.06	<.0001
Terrain_T_mount	1.84	0.17	10.63	<.0001	0.76	0.07	11.26	<.0001
Terrain_T_rolling	0.47	0.07	7.01	<.0001	0.33	0.07	4.75	<.0001
Land_U_indust	0.48	0.06	7.65	<.0001	4.36	0.34	12.98	0.0000
Land_U_commer	0.79	0.05	14.89	<.0001	0.85	0.06	14.99	<.0001
Meidan_T_raised	−0.44	0.03	−13.69	<.0001	−0.14	0.05	−2.74	0.0061
Meidan_T_TWLTL	0.51	0.09	5.50	<.0001	1.02	0.10	9.78	<.0001
Lane_N_six	0.71	0.06	11.72	<.0001				
Lane_N_four	0.51	0.07	7.60	<.0001				
Lane_W	−0.45	0.03	−13.64	<.0001	−0.34	0.03	−10.41	<.0001
Shoulder_Wr	−0.09	0.01	−13.09	<.0001	−0.21	0.01	−13.94	<.0001
ILLUM	−0.42	0.03	−13.21	<.0001	−0.46	0.07	−6.82	<.0001
RD	1.12	0.15	7.23	<.0001	1.71	0.11	15.70	<.0001
IRI					0.01	0.00	5.83	<.0001
Speed_L	0.02	0.01	3.80	0.0001	0.03	0.01	4.20	<.0001
DSL	0.13	0.01	11.60	<.0001				
Logistic regression part								
Inf_Intercept	4.16	1.11	3.75	0.0002	0.21	0.02	12.87	<.0001
Inf_ThAADT_C	−0.14	0.01	−15.19	<.0001				
Inf_ThAADT_T	−1.60	0.60	−2.68	0.0073	−1.84	0.59	−3.11	0.0019
Inf_Segment_L	−0.19	0.05	−3.68	0.0002	−0.58	0.06	−10.29	<.0001
Inf_Land_U_indust	−0.46	0.13	−3.63	0.0003	−0.69	0.09	−8.00	<.0001
Inf_Land_U_commer	−0.62	0.04	−14.67	<.0001	−0.35	0.06	−5.78	<.0001
Inf_Meidan_T_raised	0.57	0.07	8.15	<.0001	4.83	0.91	5.30	<.0001
Inf_Meidan_T_TWLTL	−0.54	0.06	−9.09	<.0001	−0.29	0.03	−8.58	<.0001
Inf_Lane_W	0.07	0.00	−15.77	<.0001	0.87	0.13	6.89	<.0001
Inf_ILLUM	0.48	0.03	14.57	<.0001	0.45	0.05	8.56	<.0001
Inf_RD	−2.81	0.34	−8.24	<.0001	−0.83	0.07	−11.96	<.0001
Inf_Speed_L	−0.01	0.00	−8.18	<.0001	−0.02	0.00	−15.07	<.0001
_Alpha	2.788	0.195	14.33	<.0001	1.955	0.227	8.63	<.0001

Table 5
Parameter estimations of the BNB model.

Parameter	Car-truck crash				Truck-only crash			
	Estimate	Std. Err.	*t* Value	Pr > |*t*|	Estimate	Std. Err.	*t* Value	Pr > |*t*|
Intercept	−7.60	1.05	−7.26	<.0001	−7.82	2.00	−3.90	0.0001
ThAADT_C	0.35	0.11	3.03	0.0024				
ThAADT_T	2.48	0.50	4.95	<.0001	3.10	0.82	3.80	0.0001
Segment_L	0.90	0.18	4.98	<.0001	0.28	0.06	5.07	<.0001
Horizontal_C	0.56	0.05	10.56	<.0001	0.48	0.07	6.74	<.0001
Land_U_indust	1.16	0.14	8.53	<.0001	2.03	0.54	3.74	0.0002
Land_U_commer	1.52	0.57	2.69	0.0072	1.25	0.09	13.67	<.0001
Meidan_T_raised	−0.39	0.07	−5.69	<.0001				
Meidan_T_TWLTL	0.63	0.09	6.69	<.0001				
Shoulder_Wr	−0.19	0.06	−3.02	0.0025	−0.16	0.03	−6.30	<.0001
ILLUM	−0.48	0.15	−3.26	0.0011	−0.49	0.13	−3.67	0.0002
RD	4.17	0.30	14.07	<.0001	7.14	0.50	14.34	<.0001
Speed_L	0.02	0.01	3.37	0.0007	0.03	0.01	3.44	0.0006
_Alpha	0.890	0.201	4.44	<.0001	0.860	0.172	4.99	<.0001

AADT, large truck AADT, segment length, degree of horizontal curvature, terrain type, land use, TWLTL median, lane number, RD, posted speed limits, and indicator for different speed limits are positive and statistically significant (at the 5% level) for the car-truck crash count model under the BIZNB framework. The results indicate that higher passenger car AADT, higher large truck AADT, longer segment length, greater degree of horizontal curvature, rolling and mountainous terrain, industrial and commercial land use, TWLTL median, greater lane number, greater RD, higher speed limit, and presence of different speed limit are associated with higher car-truck crash occurrences. The estimated parameters of raised median, lane width, right side shoulder width, and lighting conditions are negative and statistically significant, revealing that the presence of raised median, greater lane width, greater right side shoulder width, and the presence of lighting appear to be negatively related to the likelihood of car-truck crash occurrences. Compared to the results of car-truck crash count model, passenger car AADT, lane number, and indicator for different speed limits are insignificant in the truck-only crash count model analyzing by the BZINB regression and IRI is found to have significant effect on the frequency of truck-only crashes. The insignificant variables for both car-truck and truck-only crashes, including residential land use, shoulder type, left side shoulder width, and posted truck speed limit are not presented in Table 4 for the purpose of saving space.

The logistic regression part of the BZINB model predicts the likelihood of crash free occurrences. The coefficient estimates of the parameters reflect how the independent variables are associated with the odds of being a certain zero. Regarding the parameter coefficient estimates, the results of car-truck crash count model show that the higher the traffic exposure (passenger car AADT, large truck AADT, and segment length), the lower the possibility of crash free occurrences. For instance, if a segment were to increase its percentage of truck value by one unit, the odds that it would be in a certain zero group would decrease by a factor of $\exp(-1.60) = 0.20$. In other words, the more trucks on the segment, the less likely the crash frequency of segment is a certain zero. Apart from the traffic exposure variables, industry and commercial land use, TWLTL median, greater RD, and a higher speed limit are associated with a lower probability of a zero crash. Raised median, wider lanes, and better lighting conditions are associated with a higher probability of zero crashes. The similar results are obtained for the truck-only crash count model, except that the effect of passenger car AADT is insignificant.

Most of our findings are consistent with numerous previous studies (Anastasopoulos and Mannering, 2009; Haque et al., 2010; Donnell et al., 2010; Anastasopoulos et al., 2012) in terms of the effects of passenger car AADT, segment length, degree of horizontal curvature, terrain type, median type, lane number, right side shoulder width, lighting condition, RD, IRI, and posted speed limit on the occurrence of crashes involving large trucks. The following discussions of the results focus on the effect of truck AADT, land use, lane width, and the presence of different speed limits, since these debates are of greatest concern to agencies and policymakers and the conclusions are still controversial.

A previous study (Jovanis and Chang, 1986) suggested that as the percentage of trucks increases, the crash rates of truck involved crashes decrease. The crash rates were computed with crash count as numerator and traffic volume and segment length as denominator, which can cause potential estimation bias since crash counts and traffic volume are not linearly correlated. In other words, two times the traffic volume does not correspondingly result in double the number of crashes. The results of other studies (Daniel et al., 2002) showed that the effects of the percentage of trucks on the large truck crash involvements are insignificant. This finding appeared to be counterintuitive. Given the limited number of observations available in their study, it is possible that confounding factors, such as the AADT where large truck traffic volume is included, may explain this outcome. In our study, truck AADT was taken as an explanatory variable to estimate crash frequencies. The modeling results indicate that the occurrences of car-truck and truck-only crashes are significantly positively related to the truck AADT. One possible explanation is that the chance of a collision with a truck increases as the truck traffic volume increases. In addition, the frequency of lane changing and overtaking movements by vehicles increases as the truck volume increases, which has negative effects on safety.

Turning to the effects of land use, there are no consistent results in the literature. Our findings suggest that a higher number of car-truck crashes are likely to occur on commercial zone roadways. Industrial land use is found to associate with a higher number of truck-only crashes. The coefficient for the industrial land use is 0.48 and is 0.79 for the commercial land use in car-truck crash count model, which indicates that commercial zone roadways are associated with higher car-truck crash frequencies compared with industrial zone roadways. The opposite results are obtained in the truck-only crash count model, which show that more truck-only crashes are likely to occur on the road segment that belongs to industrial zones. This result does not agree with most of the previous findings, which suggested that rural areas are associated with higher crash frequencies. There are also some studies (Jiang et al., 2014; Donnell et al., 2010) that suggests that urban locations are associated with higher crash frequencies. The results of Kim et al. (2007) showed that there were no significant differences between the different land development characteristics. However, the results are expected because industrial and commercial zones involve more complex traffic conditions and higher truck traffic volume. In addition, industrial and commercial driveways are likely to have slow moving trucks entering and leaving, increasing the likelihood for vehicle conflicts. Further research is needed with larger data sets to verify the repeatability of this finding.

Regarding the lane width, some studies (Lord and Bonneson, 2007; Chimba et al., 2010) have shown crash reductions for every foot of lane narrowing; other studies (Martens, 1997) showed a slight crash increase for per foot of lane narrowing or no significant effect at all. Our finding is consistent with previous analysis of Garber and Ehrhart (2000), which shows large truck crash involvements decrease with wider lanes regardless of the crash types. Since passenger cars measure about 6 feet in width, while large trucks are often about 8.5 feet wide and side-view mirrors usually add between 6 and 12 in. to the vehicles' total width, the presence of wider lane widths reduces the frequency of large truck involved crashes. Though 10 ft. wide lanes are generally accepted in the practice, we recommend providing wider curb lanes to ease large truck operations, separate large truck traffic from roadside drainage and drainage features, and better accommodate on-street bicycles. In addition, we believe that the wider lane can reduce truck rear-end collisions, especially underride crashes (a car slide underneath the trailer).

Creating a differential speed limit (DSL) is considered to be an effective countermeasure to promote road safety. However, the safety effects of providing different speed limits for cars and trucks traveling on the same roadway have been inconclusive in previous studies. Some studies found no difference between uniform speed limit (USL) and DSL. Other studies found one or the other to be a better policy choice. Our finding shows that the occurrences of car-truck crashes are significantly related to DSL and the effects of DSL on the truck-only crashes are insignificant. The results are consistent with the findings of Garber and Gadiraju (1991). They analyzed the crashes that results in the adjacent states of Virginia (DSL) and West Virginia (USL). The results showed that a higher proportion of car-into-truck and truck-into-car crashes occurred in DSL states than in USL states, especially for rear end crashes where more car-into-truck collisions occurred in the DSL group. In addition, a simulation study by Garber and Gadiraju (1989) reported a potential for an increase in crash rates for facilities using DSL, especially in the case of higher traffic volumes and higher percentage of trucks.

In summary, regarding the effects of passenger car AADT, segment length, degree of horizontal curvature, terrain type, median type, lane number, right side shoulder width, lighting condition, RD, IRI, and posted speed limits on the occurrences of large truck involved crashes, our findings are consistent with the previous studies. Regarding the effects of truck AADT, land use, lane width, and the presence of different speed limits on the occurrences of large truck involved crashes, our findings reveal some new insights that would influence policy making. More specifically, our findings suggest that: (1) the occurrences of car-truck and truck-only crashes are significantly positively related to the truck AADT; (2) a higher number of car-truck crashes is associated with commercial land use and industrial land use is found to associate with a higher number of truck-only crashes; (3) large truck crash involvements decrease with wider lanes regardless of the crash types; (4) the occurrences of car-truck crashes are significantly related to DSL and the effects of DSL on the truck-only crashes are insignificant.

6. Conclusions

Freight trucks are an important component of the nation's highway traffic. Due to their physical and operational characteristics, they can significantly affect the traffic system performance, safety, and the travel experience of non-truck drivers. In this study, we use BZINB and BNB regression modeling of the car-truck and truck-only crash counts simultaneously to investigate the effects of highway geometric design features on the crash occurrences. The results indicate that the BZINB model appears to produce more precise parameter estimates with additional variables found to be statistically significant and better predictive accuracy. In other words, the BZINB models were found to possess most of the desirable statistical properties in describing large truck involved crashes. The prime justification for the use of BZINB models is that this type of model provides improved statistical fit for modeling correlated crash counts characterized by a preponderance of zeros. However, the model does have its limitations, most notably its logic problem to describe crash data using a dual-state process (i.e., the mixture of true safety with unsafe roadway entities), since the baseline of the proposed model is the zero-inflated regression model. In addition, the BZINB model that applies a logistic splitting model for zeros as a function of covariates adds additional model parameters, which means additional technical complexity. Thus, the additional complexity added by the BZINB models needs to be taken into account when applying them.

Compared to the BZINB models, BNB models could not distinguish between the two processes causing an excessive number of zeroes. The results indicate that, if overdispersion in raw data is caused by the zero-inflation, the BZINB model provides better statistical fit to the data. One of the limitations of the study is that we did not include the time variables in the proposed model, since the assumption of the proposed model is that if crashes occurring on the same segment after a two year period are considered as independent due to the changes can potentially occur. A comparative analysis of alternate model structures, subject to the availability of adequate data, is identified as an area of future research.

Regarding the parameter estimation results of the BZINB models, our findings suggest that the passenger car AADT, large truck AADT, segment length, degree of horizontal curvature, terrain type, land use, median type, lane number, lane width, right side shoulder width, lighting condition, RD, posted speed limits, and indicators for DSL are found to have statistical significant effects on the occurrences of car-truck crashes. Compared to the results of car-truck crash count models, among those significant variables, passenger car AADT, lane number, and DSL are insignificant in the truck-only crash count model and one additional factor, IRI, is found to have significant effects on the occurrences of truck-only crashes. The study reveals some new or more comprehensive findings that have not been covered in previous studies, including (1) the occurrences of car-truck crashes are significantly positively related to truck AADT, commercial land use, and DSL; (2) the occurrences of truck-only crashes are significantly positively associated with truck AADT and industrial land use; (3) both the car-truck and truck-only crashes decrease with wider lanes.

Acknowledgements

The authors would like to thank anonymous reviewers for their useful suggestions and comments to improve the paper. Special thanks to TDOT for providing the TRIMS data. This research is supported by funding provided by the Southeastern Transportation Center – a Regional UTC funded by the USDOT – Research and Innovative Technology Administration. Additional funding was provided by the Natural Science Foundation of China (No. 51208052).

References

Anastasopoulos, P.C., Mannering, F.L., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. Accident Anal. Prevent. 41 (1), 153–159.

Anastasopoulos, P.C., Shankar, V.N., Haddock, J.E., Mannering, F.L., 2012. A multivariate tobit analysis of highway accident-injury-severity rates. Accident Anal. Prevent 45 (1), 110–119.

Blower, D., Campbell, K.L., Green, P.E., 1993. Accident rates for heavy truck-tractors in Michigan. Accident Anal. Prevent. 25 (3), 307–321.

Braver, E.R., Zador, P.L., Thum, D., Mitter, E.L., Baum, H.M., Vilardo, F.J., 1997. Tractor-trailer crashes in Indiana: a case-control study of the role of truck configuration. Accident Anal. Prevent. 29 (1), 79–96.

Carson, J., Mannering, F., 2001. The effect of ice warning signs on accident frequencies and severities. Accident Anal. Prevent. 33 (1), 99–109.

Chang, L., Chien, J., 2013. Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model. Safety Sci. 51 (1), 17–22.

Chen, F., Chen, S., 2011. Injury severities of truck drivers in single- and multi-vehicle accidents on rural highways. Accident Anal. Prevent. 43 (5), 1677–1688.

Chimba, D., Sando, T., Kwigizile, V., 2010. Effect of bus size and operation to crash occurrences. Accident Anal. Prevent. 42 (6), 2063–2067.

Daniel, J., Tsai, C., Chien, S., 2002. Factors in truck crashes on roadways with intersections. Transport. Res. Record 1818, 54–59.

Dong, C., Clarke, D.B., Richards, S.H., Huang, B., 2014. Differences in passenger car and large truck involved crash frequencies at urban signalized intersections: an exploratory analysis. Accident Anal. Prevent. 62, 87–94.

Donnell, E.T., Porter, R.J., Shankar, V.N., 2010. A framework for estimating the safety effects of roadway lighting at intersections. Safety Sci. 48 (10), 1436–1444.

Famoye, F., 2010. On the bivariate negative binomial regression model. J. Appl. Statistics 37 (6), 969–981.

FHWA, 2003–2012. Highway Statistics 2003–2012. U.S. Department of Transportation, Federal Highway Administration, <http://www.fhwa.dot.gov/policyinformation/statistics.cfm>.

Garber, N.J., Ehrhart, A.A., 2000. Effect of speed, flow, and geometric characteristics on crash frequency for two-lane highways. Transport. Res. Record 1717, 76–83.

Garber, N.J., Gadiraju, R., 1989. Effect of truck strategies on traffic flow and safety on multilane highways. AAA Found. Traffic Safety.

Garber, N.J., Gadiraju, R., 1991. Impact of differential speed limits on highway speeds and accidents. AAA Found. Traffic Safety.

Geedipally, S.R., Lord, D., 2010. Investigating the effect of modeling single-vehicle and multi-vehicle crashes separately on confidence intervals of Poisson–gamma models. Accident Anal. Prevent. 42 (4), 1273–1282.

Haque, M.M., Chin, H.C., Huang, H., 2010. Applying Bayesian hierarchical models to examine motorcycle crashes at signalized intersections. Accident Anal. Prevent. 42 (1), 203–212.

Jiang, X., Yan, X., Huang, B., Richards, S.H., 2014. Influence of curbs on traffic crash frequency on high-speed roadways. Traffic Injury Prevent. 12 (4), 412–421.

Jovanis, P.P., Chang, H.L., 1986. Modeling the relationship of accidents to miles traveled. Transport. Res. Record 1068, 42–51.

Khattak, A.J., Khattak, A.J., Council, F.M., 2002. Effects of work zone presence on injury and non-injury crashes. Accident Anal. Prevent. 34 (1), 19–29.

Kim, J., Kim, S., Ulfarsson, G.F., Porrello, L.A., 2007. Bicyclist injury severities in bicycle-motor vehicle accidents. Accident Anal. Prevent. 39 (2), 238–251.

Kumara, S.S.P., Chin, H.C., 2003. Modeling accident occurrence at signalized tee intersections with special emphasis on excess zeros. Traffic Injury Prevent. 3 (4), 53–57.

Lee, J., Mannering, F., 2002. Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accident Anal. Prevent. 34 (2), 149–161.

Lemp, J.D., Kockelman, K.M., Unnikrishnan, A., 2011. Analysis of large truck crash severity using heteroskedastic ordered probit models. Accident Anal. Prevent. 43 (1), 370–380.

Lord, D., Bonneson, J.A., 2007. Development of accident modification factors for rural frontage road segments in Texas. Transport. Res. Record 2023, 20–27.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transport. Res. Part A 44 (5), 291–305.

Lord, D., Washington, S.P., Ivan, J.N., 2005. Poisson, Poisson-gamma and zero inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accident Anal. Prevent. 37 (1), 35–46.

Lyman, S., Braver, E.R., 2003. Occupant deaths in large truck crashes in the United States: 25 years of experience. Accident Anal. Prevent. 35 (5), 731–739.

Martens, M., 1997. The effects of road design on speed behavior: a literature review. MASTER. The European Commission under the Transport RTD Programme of the fourth Framework programme. Deliverable D1, University of Leeds, UK.

Miaou, S., 1994. The relationship between truck accidents and geometric design of road sections: poisson versus negative binomial regressions. Accident Anal. Prevent. 26 (4), 471–482.

Miaou, S., Lum, H., 1993. A statistical evaluation of the effects of highway geometric design on truck accident involvements. Transport. Res. Record 1407, 11–23.

NHTSA, 2012. Traffic Safety Facts. U.S. Department of Transportation, National Highway Traffic Safety Administration.

Schneider IV, W., Zimmerman, K., Van Boxel, D., Vavilikolanu, S., 2009. Bayesian analysis of the effect of horizontal curvature on truck crashes using training and validation data sets. Transport. Res. Record 2096, 41–46.

Shankar, V.N., Milton, J., Mannering, F.L., 1997. Modeling accident frequency as zero-altered probability processes: as empirical inquiry. Accident Anal. Prevent. 29 (6), 829–837.

Shankar, V.N., Ulfarsson, G.F., Pendyala, R.M., Nebergall, M.B., 2003. Modeling crashes involving pedestrians and motorized traffic. Safety Science 41 (7), 627–640.

USDOT/BTS, 2013. Freight Facts and Figures 2013. U.S. Department of Transportation, Bureau of Transportation Statics, <http://www.rita.dot.gov/bts/data_and_statistics/by_subject/freight.html>.

Wang, P., 2003. A bivariate zero-inflated negative binomial regression model for count data with excess zeros. Econom. Lett. 78 (3), 373–378.

Wang, K., Lee, A.H., Yau, K.K.W., Carrivick, P.J.W., 2003. A bivariate zero-inflated Poisson regression model to analyze occupational injuries. Accident Anal. Prevent. 35 (4), 625–629.

Winkelmann, R., 2008. Econometric Analysis of Count Data, Fifth ed. Springer, Germany.

Zhu, X., Srinivasan, S., 2011. A comprehensive analysis of factors influencing the injury severity of large-truck crashes. Accident Anal. Prevent. 43 (1), 49–57.