Modeling Crash Frequency of Heavy Vehicles in Rural Freeways

Reza Imaninasab
School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran
Email: imany_reza@civileng.iust.ac.ir

Emad Sekhavati and Mansour Hajihoseinloo
Department of Civil and Environmental Engineering, K.N. Toosi University of Technology Tehran, Iran
Email: emad.sekhavati@gmail.org

***Abstract*—In the recent years, Prediction models of crash frequency have been employed to estimate accident rate and have been remarkably taken into account for safety improvement plan. Heavy vehicles are one of the main causes of accidents worldwide as there is one out of 9 accidents that they were involved. In this research, a prediction model was developed using heavy vehicles accident data collected from Karaj-Qazvin freeway. In the model, important factors of occurrence of accidents including climatic condition, road surface condition, natural light condition, average vehicle speed, average daily traffic and 5 dangerous violations commitment were considered. Poison regression and negative binomial regression models were employed for modeling purposes. After evaluation of both models, factors including natural light condition, heavy vehicles daily volume, speed limit violation and vehicle defects, especially in lighting system during darkness, were found to be the most effective factors on heavy vehicles accident.**

***Index Terms*—heavy vehicles accident frequency, poison regression model, negative binomial regression model**

I. Introduction

In the United States, there were 317000 large trucks involved in traffic crashes during 2012. 3921 people were killed and 104000 people injured in such crashes. There is 4-persent increase in fatal crash in which heavy vehicles involved from 2011 to 2012. Moreover, in these crashes, 73% of fatalities are occupant of other vehicles. 63% fatal crashes of this type take place in rural highways. 3802 large trucks were involved in fatal crashes and 77000 in injury crashes. 81% of fatal crashes of large trucks are multiple-vehicle crashes, compared with 58% for fatal crashes involving passenger vehicles [1].

A study comparing the heavy vehicle safety performance of Australia's road found that Australia's heavy vehicle fatalities per kilometer traveled was 46% higher than the US [2]. In the 2006-7, 41 of the 295 working fatalities were in the road transport industry [3]. Statistics for road crashes involving heavy vehicle in the 12 month period to June 2010 indicated total of 160 fatalities were recorded from 130 crashes involving articulated truck, further 79 fatalities from 64 crashes involving rigid truck [4].

In Iran, heavy vehicles defects have caused accidents increase. One type of tractor-trailer, namely Hovo, comprises 13% of truck fleet in Iran. It is the most unsafe heavy vehicle having some safety deficiency. According to the statistics of Iran Center of Traffic Control, it claimed 118 lives from 255 accidents that it was involved within 2008-9. It increased to 203 people out of 456 accidents within the following year [5]. Its gear 4 is 20 km faster than that of conventional ones such as Volvo FH which is not concerned by amateur drivers. Moreover, an inspection of 84 Hovos revealed that there were just two of the vehicle types featuring with standard trailers attached to them while all the others were fabricated in the local workshops [6]. It was announced that 90 percent of the total used trailers are not standard [7].

Such statistics indicate that heavy vehicle involvement in an accident results in severe and fatal injury accidents compared with other types of accidents. Not surprisingly, as the vehicles weight increases the risk of being injured or damage decrease substantially, even though other driver-vehicle units involved in the same crash may be vulnerable to be injured or damaged [8]. Moreover, a higher relative fatality risk is associated with truck crash on high speed facilities like freeway rather than urban roads [9]. These all signify the importance of heavy vehicles accidents in rural highways to be investigated since they are one of the main reasons for accidents.

In this research, crash frequency of heavy vehicles in Karaj-Qazvin freeway was modeled by applying typical regression models that can be used for this purpose. The most accurate model was then selected based on lowest p-value. It should be noted that 50 percent of total freeway traffic volume in Iran travels through Karaj-Qazvin freeway as it connects north-west of Iran to its center and east.

II. Literature Review

There are various models recommended to predict crash frequency of heavy vehicles. Some of them are presented as following:

Manuscript received January 15, 2016; revised June 4, 2016.

doi: 10.18178/jtle.4.2.98-102

A. Analysis of Truck-Involved Freeway Accidents Using Log-linear Modeling

This model was developed according to 9000 truck-involved accidents occurred during a two-years period in south California metropolitan freeway. Log-linear model was used to establish relation among deterministic parameters which quantifies identifications and locations of heavy vehicles accident [10].

B. Developed Models for U.S. Highways

In these models, mutual effect of number of lane, direction-number of lane, median-number of lane, PHF[1] -lane width, refuge-ADT[2], speed-ADT, ADT-PHF, lane width-PHF and lane width are important parameters and effect of other parameters were distinguished negligible. Relation between accident rate and effective parameters was developed by the presented models in the best way since accident rate is predicted more accurately by multiple variables models. Based on the collected data, following multiple variables regression models were so calibrated to relate effective factors to accident rate. Coefficients of these models were computed by statistic software of SPSS. The models are classified into two levels: level one was spread models that relate total accident to the total models, and level two was accident rates categorized according to number of lanes (one lane to six lanes), ADT (more or less than 4000 vehicles daily), speed variation (more or less than 50 km/hr) [11].

C. Basic Model Presented by FHWA[3]

This model was developed by HSIS[4] of FHWA based on negative binomial regression analysis and accident data of 616 sections of rural two-lane highways of Minnesota and 712 sections of rural two-lane highways of Washington. These sections comprise of 1130 km and 850 km of rural two-lane highways in Minnesota and Washington, respectively. Available database on accidents was collected within 5 years (from 1985 to 1989) for Minnesota and 3 years for Washington (from 1993 to 1995) [12].

D. General Linear Models

General linear modeling techniques are linear modeling techniques for multiple variables functions that relates several effective parameters to crash frequency. Abnormal distribution is assumed in such techniques.

One of the most typical general linear models is Poison Regression Model (PRM). Poison distribution is applicable for statistical variables whose mean and variance must be equal. Eq. (1) shows poison regression model:

$$ln\, \mu = \beta_0 + \beta_1 x \tag{1}$$

In the equation above:

μ is both the mean and variance which is a limitation for Poison distribution

β_0 and β_1 are calibration parameters in the form of vectors

x is independent variable

Inequality of mean and variance is common for accident data and, as a result, this leads to overdispersion. It declines one of the model's assumptions and therefore Poison model would not be suitable in such cases [13].

Since overdispesion of count data for accidents is usual, negative binomial regression model is recommended to deal with this condition. It allows the variance to exceed the mean. This distribution model is composed of count and adjunct data of poison regression model. Eq. (2) shows negative binomial regression mode.

$$\ln \lambda = \beta_0 + \beta_1 x + \varepsilon_i \tag{2}$$

In equation above:

λ is the mean value of the response variable

β_0 and β_1 are calibration parameters in the form of vector

x is independent variable

ε_i is error.

It should be noted that $\exp(\varepsilon_i)$ is standard gamma distribution with mean equal to one and variance of σ^2. In this model the response (dependent) variable distribution is also named poison gamma which is negative binomial distribution [14].

III. Data Collection

In this research, different types of data sets were collected from three different resources. Light, road surface, climatic condition at the moment of accident, and heavy vehicles-related data were obtained from traffic police of Tehran and Alborz provinces. Iran Road Maintenance and Management Organization is the source of volume and speed data. Data regarding driving violation is supplied by information technology organization of traffic police. All data is from March 21, 2011 to March 21, 2012. Heavy vehicles accident frequency data in Karaj-Qazvin freeway is presented in Table I.

TABLE I. Heavy Vehicles Accidents Frequency in Karaj-Qazvin Freeway

Number of heavy vehicle accident	Frequency	Percent of frequency (%)	Cumulative percent (%)
0	138	38.12	38.12
1	122	33.70	71.82
2	73	20.17	91.99
3	22	6.08	98.07
4	5	1.38	99.45
5	2	0.55	100
Sum		100	

A. Light Condition Data

This type of data is referred to the goodness of natural lighting condition at the moment of accident. Similar to any other corridor, there are four lighting conditions for Karaj-Qazvin freeway including daylight, dark night; dawn and dusk. Light condition at dawn and dusk are nearly the same and number of accidents at these periods

[1] Peak Hour Factor
[2] Average Daily Traffic
[3] Federal Highway Administration
[4] Highway Safety Information system

is low, so for modeling purposes, they are assumed as one light condition [15].

For the variable related to light condition, three factors of daylight, dark night, dusk or dawn were employed for modeling purposes. Each variable that relates to any of the light conditions is equal to 1 if that condition exists when accident takes place, otherwise it is equal to 0.

B. Road Surface Condition Data

Road surface condition is affected by weather type, moisture level, and the temperature. It is divided into the dry, wet and icy surface. For countries with tropical region climate like Taipei or Singapore there are just two road surface condition including wet and dry as snowing is rare [8], [15]. In the developed models in this study, since there are all the three surface conditions probable, they are employed in the form of three bainary variables equaling 1 in the case of being true and 0 otherwise.

Surface condition is considered significantly effective in term of accident modeling due to the highly impact of sliding surface on accident occurrence, especially on heavy vehicles in the recent years.

C. Seasonal and Climatic Conditions Data

Another effective factor on prediction model of accident frequency is seasonal and climatic conditions. Clearly, bad climatic condition not only does increase accident probability, but it also affects volume and speed of vehicles. Moreover, volume, as the most important factor of accident, changes with season change. Weather can be considered as fine or not [8]. It is also accounted when it is raining or snowing [16]. In this research, there were clear, raining and snowing climatic conditions. For seasonal variables, 0 was attributed to the warm seasons, namely spring and summer, and 1 to the cold seasons, namely fall and winter.

D. Data Related to Mean Speed of Vehiclesd

In the recent years, variation of mean speed of vehicles is identified as one of the most sensitive item with respect to accident frequency for different corridors. In this research, effect of this factor on heavy vehicles accident was taken into account .

E. Volume Data

Another important factor on accident is volume of a corridor. Undoubtedly, low volume increases mean speed and there is mean speed reduction by volume increase. The relation between these two parameters results in accident reduction or increase. In this modeling, impact of volume was considered by heavy vehicles and overall average annual daily traffic (AADT) as two independent variables in modeling.

F. Traffic Violation Data

Traffic violation data was provided by Information Technology Organization of Traffic Police. There are 20 items of Traffic violation that causes driving license-point reduction. Among them, 5 most effective on heavy vehicles accidents were selected for modeling purposes. They are Passing in no passing zone, speed limit violation, vehicle defect especially in lighting system at night, failure to yield and spiral movement.

IV. MODELLING CRASH FREQUENCY OF HEAVY VEHICLES IN KARAJ-QAZVIN FREEWAY

After modeling using both Poison and negative binomial regression models, the appropriate model was selected. Since likelihood ratio Statistics was nearly zero, P-value of 0.489 was obtained. As a result, Poison model was more accurate than negative binomial. Calculated coefficient for Poison model is presented in Table II.

TABLE II. RESULT OF POISON REGRESSION MODEL

Variable	coefficient	p-value
Light condition		
Daylight	2.218098	0.008
Dark night	2.409442	0.012
Dawn-sunset	2.105092	0.191
Road surface condition		
Dry	1.663033	0.427
Wet	2.096273	0.568
Icy	-2.591775	0.437
Seasonal and climatic condition		
Clear	-1.648118	0.236
Raining	-2.008221	0.079
Snowing	2.399593	0.541
Season	0.076938	0.394
Mean speed of vehicles		
Pickup and passenger car	0.0047989	0.456
2 axles light truck and minibus (dual rear tire)	0.0067719	0.561
2 0r 3 axle truck	-0.005362	0.174
Bus	0.0130862	0.221
Over 3 axle truck and trailer	-0.0021515	0.384
AADT		
Overall AADT	-4.93e-06	0.095
Heavy vehicle AADT	6.07e-06	0.021
Traffic violation		
Passing in no passing zone	-0.0023659	0.425
Speed violation	0.0031765	0.036
Vehicle defect	-0.0095022	0.062
Yield failure	0.0025269	0.514
Spiral movement	0.0043577	0.185

To improve the model accuracy, variables with P-value greater than 0.2 were eliminated from the model and modeling process was repeated by the remaining variables once again. Eq. (3) is the final model:

$$\begin{cases} \lambda_i = e^{(-2.215946+\sum_i A_i x_i)} \\ , \\ 0 \leq i \leq 9 \end{cases} \tag{3}$$

In equation above:

$$x_1 : \begin{cases} 1 \text{ if accident occurs during daylight} \\ 0 \text{ otherwise} \end{cases}$$

$$x_2 : \begin{cases} 1 \text{ if accident occurs during dark night} \\ 0 \text{ otherwise} \end{cases}$$

$$x_3 : \begin{cases} 1 \text{ if accident occurs at dawn or sunset} \\ 0 \text{ otherwise} \end{cases}$$

$x_4:\begin{cases} 1 \text{ for rainy wether} \\ 0 \text{ otherwise} \end{cases}$

x_5: number of speed limit violations

x_6: number of defected vehicle

x_7: number of spiral movement

x_8: heavy vehicle AADT

x_9: AADT

Modified coefficient of the model is presented in Table III.

TABLE III. MODIFIED COEFFICIENT OF ACCIDENT FREQUENCY PREDICTION MODEL OF KARAJ-QAZVIN FREEWAY

Coefficient	Value
a_1	1.79221
a_2	1.64763
a_3	0.738807
a_4	0.170789
a_5	0.007012
a_6	0.007831
a_7	0.004545
a_8	1.01e-06
a_9	0.000024

V. MODEL VALIDATION

To validate the model, different methods are recommended. Among them, prediction range method was chosen for this research. Data related to the heavy vehicles accidents from November 21, to December 21, is used for verification.

According to prediction range method, prediction range with the level of reliability of 95% is considered acceptable for the calculated models. By putting data into the model, number of heavy vehicles accidents can be estimated.

Number of heavy vehicles accidents, predicted number of heavy vehicles accidents by the model, and level of reliability of 95% for each day [17] all are depicted in Fig. 1. Comparing graphs of real accident number with the predicted one reveals that they are compatible with each other. There are 8 outliers in predicted accidents value for Karaj-Qazvin freeway, which results in 26.67% error in the model.

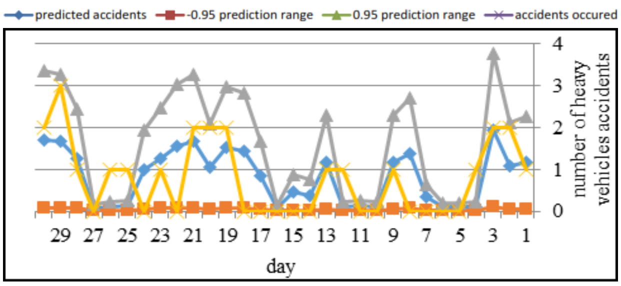

Figure 1. Validation graph for the model.

VI. CONCLUSION

This study focuses on the frequency of heavy vehicles-involved accidents in rural freeway. Karaj-Qazivin freeway was selected as case study and a prediction model for heavy vehicles accident frequency was developed which was the main result of this research (Eq. (3)). Moreover, following conclusions can be inferred from the research:

- Poison Regression model better estimates heavy vehicle-involved accidents compared to Negative Binomial.
- According to Table II., it can be concluded that mean speed of vehicle has no slight effect on crash frequency of heavy vehicles in the developed model. Speed limit violation, however, has significant effect on heavy vehicles accident increase.
- Among binary variables condition of being rainy and night, are more effective than others.
- Variables relating to light conditions, rainy weather, heavy vehicles volume, overall volume, speed limit violation, spiral movement and vehicle defect are the most effective parameters on accident occurrence.
- According to p-values presented in Table. II, impact of variables relating to rainy weather, heavy vehicles volume, overall volume, speed limit violation and vehicle defects on model are more than other variables.

REFERENCES

[1] 2012 data: Large Truck-Traffic Safety Facts Washington, D.C., National Highway Traffic Safety Administration (NHTSA), Washington, DC, 2014.

[2] N. Haworth, P. Vulcan, and P. Sweatman, “Benchmarking truck safety in Australia,” presented at the Road Safety Research, Adelaide, South Australia, Nov. 3-5, 2002.

[3] S. J. Raftery, J. A. Grigo, and J. E. Woolley, “Heavy vehicle road safety: A scan of recent literature,” *Journal of the Australasian College of Road Safety*, vol. 22, no. 3, pp 18-24, Aug. 2011.

[4] Truck Productivity: Sources, Trends and Future Prospects, Report 123, Bureau of Infrastructure, Transport and Regional Economics (BITRE), Canberra, Australia, 2011.

[5] Return of Death Chariot to Roads, SHARGHDaily Newspaper, Margin Newspaper Database, Sep. 19, 2010.

[6] Investigation of Heavy Vehicles Accident Considering Technical Defects, Transportation Industry News Network (TINN), Nov. 22, 2011.

[7] 90 Percent of Trailers are not Standard in Iran, Tabnak News Database, Nov. 1, 2011.

[8] H. Huang, H. C. Chin, and M. M. Haque, “Severity of driver injury and vehicle damage in traffic crashes at intersections: a Bayesian hierarchical analysis,” *Accident Analysis & Prevention*, vol. 40, no. 1, pp. 45-54, Jan. 2008.

[9] S. M. Rifaat and H. C. Chin, “Analysis of severity of single-vehicle crashes in Singapore,” presented at TRB 2005 Annual Meeting CD-ROM, Transportation Research Board, National Research Council, Washington DC, Jan. 9-13, 2005.

[10] T. F. Golob and W. W. Recker, “An analysis of truck-involved freeway accidents using log-linear modeling,” *Journal of Safety Research*, vol. 18, no. 3, pp. 121-136, Autumn 1987.

[11] X. Li, D. Lord, Y. Zhang, and Y. Xie, “Predicting motor vehicle crashes using support vector machine models,” *Accident Analysis & Prevention*, vol. 40, no. 4, pp. 1611-1618, April 2008.

[12] A. Vogt and J. Bared, “Accident models for two-lane rural segments and intersections,” *Transportation Research Record: Journal of the Transportation Research Board*, no. 1635, pp. 18-29, Nov. 1998.

[13] S. Vavilikolanu, “Crash prediction models on truck-related crashes on two-lane rural highways with vertical curves,” M.S. thesis, Graduate Faculty, University of Akron, Akron, OH, 2008.

[14] G. Berhanu, “Models relating traffic safety with road environment and traffic flows on arterial roads in Addis Ababa,” *Accident Analysis & Prevention*, vol. 36, no. 5, pp. 697-704, Dec. 2004.

[15] L. Y. Chang and H. W. Wang, "Analysis of traffic injury severity: An application of non-parametric classification tree techniques," *Accident Analysis & Prevention*, vol. 38, no. 5, pp. 1019-1027, June 2006.
[16] A. Khorashadi, D. Niemeier, V. Shankar, and F. Mannering, "Differences in rural and urban driver-injury severities in accidents involving large-trucks: An exploratory analysis," *Accident Analysis & Prevention*, vol. 37, no. 5, pp. 910-921, Sep. 2005.
[17] Y. J. Kweon and K. M. Kockelman, "Overall injury risk to different drivers: combining exposure, frequency, and severity models," *Accident Analysis & Prevention*, vol. 35, no. 4, pp. 441-450, June 2003.

Reza Imaninasab was born in Bandarabbas, Iran, in 1988. In 2013, he received his MS degree in highway and transportation engineering from Iran University of Science and Technology, Tehran, Iran, where he is currently PhD student.
His main research interests are traffic safety, pavement materials characterization and evaluation, and pavement work planning optimization. He is now working on collision prevention in order to reduce number of accidents.
Mr. Imaninasab was the reviewer of 14th International Conference on Traffic and Transportation Engineering (ICTTE). He won best paper award in traffic safety section of 14th ICTTE entitled "Modeling crash frequency of heavy vehicles in rural freeways".

Emad Sekhavati was born in Mashhad, Iran, in 1988. He was graduated from K.N. Toosi University of Technology, Tehran, Iran, in highway and transportation engineering in 2012.
He used to cooperate with traffic police and investigated impact of various factors on accident rate which was also related to his MS thesis.
Mr. Sekhavati was the co-author of the best paper award of traffic safety section of 14th ICTTE entitled "Modeling crash frequency of heavy vehicles in rural freeways".

Mansour Hajihoseinloo received his MS and PhD from Hokkaido University, Japan, in 1998.
He joined K.N. Toosi University of Technology from 1998 and is now assistant professor of Civil Engineering faculty there. His field of interest includes Transportation Planning, Traffic Safety, and airport design and planning.
Dr. Hajihoseinloo has supervised several dissertations on traffic and transportation engineering.