Abstract No. 90

A critical review of regression models for analysing highway crash data

Ms. Sujata Basu
Research scholar
Department of Civil Engineering
Indian Institute of Engineering Science and Technology
Shibpur-711103, INDIA
basusujata29@gmail.com

Dr. Pritam Saha *(Corresponding Author)*
Assistant Professor
Department of Civil Engineering
Indian Institute of Engineering Science and Technology
Shibpur-711103, INDIA
saha.pritam@gmail.com

Abstract: Increasing road traffic over the years has resulted in simultaneous increase in road accidents particularly in most of the developing countries like India. Thus, road accident has become a major concern and analysing accident data has been an important look out to the analysts in order to discern the trends of road crashes. Accordingly, a number of statistical models have been developed over the past few decades for road crash prediction. This paper basically provides a detailed insight of those models and evaluates their compatibility and aptness in predicting crashes on roads if the prevailing traffic is heterogeneous in character.

Keyword: Heterogeneous traffic; road crashes; over-dispersion; regression models

1. Introduction

Since independence, India has witnessed tremendous economic growth which has resulted in expansion of road network and subsequent increase in vehicular population .This had a considerable impact on operation and safety of traffic. With more than one death and four injuries every minute, India, being a developing nation, has the dubious distinction of reporting highest number of road fatalities in the world. Moreover, there have been strong indications over the past decades which placed "road traffic injury" on 9th position among the 10 leading causes of deaths (WHO, 2011). The world health organization (WHO) revealed in the global status report on road safety that fatal accidents is more on Indian roads than anywhere else in the world (WHO, 2013) . This calls for a comprehensive analysis of highway cashes at the stage of planning and designing.

Crash prediction is a crucial step in the road safety management process. The Highway Safety Manual (HSM, 2010) suggests the use of safety performance functions (SPF) while predicting crash frequencies on different types of roads and also, considers the fact that it would vary significantly with the change of road environment. Accordingly, the suggested model warrants calibration for the local conditions at the time of application. HSM provides guidelines for the calibration process wherein it incorporates Empirical Bayes (EB) method for the regression of SPF model. However, HSM is applied only for the road segments of homogeneous characteristics; this is expressed in terms of traffic volumes, roadway design characteristics, and also, traffic control features. Thus, there is a need to develop an indigenous model, aimed at predicting crash frequencies in developing countries where heterogeneity in traffic composition is prevalent.

There have been a number of regression models, proposed over the past few decades, which are applied in modelling crash data. Typically, the Binomial, Poisson, Poisson lognormal, Zero-inflated models exhibit their appropriateness because of their flexibility and ease in estimating the parameters. However, conventional regression analysis cannot be used as it assumes dependent variable is continuous and normally distributed with a constant variance. Thus, application of such regression is impractical in predicting crashes; this attributes to the fact that crash frequencies is discrete non negative variable and its variance depends on its mean (Hadi et al, 1995). Further, poisson regression is considered as one of the most basic models which offers significant ease in estimating the parameters, however, it cannot handle over and under dispersion (Lord et al 2005). In dealing with the over-dispersion in crash data, Negative Binomial Regression (NBR), an alternative to poisson regression, has been used in accident modelling (Lord, 2006; Cheng, 2008). Poisson lognormal and COM poisson distribution model, on the other hand, is more flexible than NBR (Lord et al, 2005).

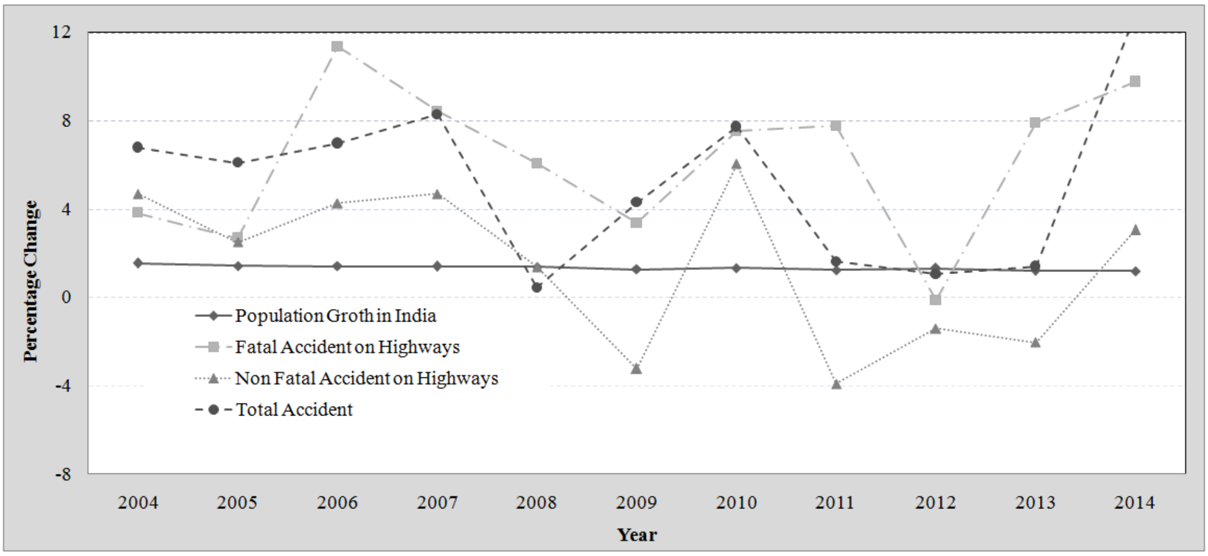
Over the years, several regression techniques have been experimented while modelling the crash data in order to investigate their compatibility in terms of goodness of fit and the most common techniques include negative binomial and poisson lognornal. Considering the Indian traffic, which exhibits heterogeneity in its composition and experiences higher rate of fatality, the present paper explicates the need of proposing an appropriate regression technique in calibrating crash prediction model. Further, in the process of arriving at an appropriate technique, it was felt imperative to study the inherent strengths and weaknesses of the models developed over the years and also their external opportunities and threats while analysing current traffic safety situation of roads. Accordingly, an attempt was made to analyse the SWOT aimed at systematic identification of the appropriate approach for the development of prediction model; the logic behind this was to maximize the strength and opportunities of a model on the one hand and minimize its weaknesses of on the other hand.

2. Analysing highway crash data: A state-of-the art review

A look into global statistics of road accidents makes it evident that the trend of road crashes over the past few decades is exponential in most of the developing countries where traffic is mostly composed of a wide variety of vehicle categories, thereby, making it a serious concern to the traffic analysts; they are in search of having a unified method of assessing safety performance of roads having mixed traffic. Thus, the premise on which the present study is based considers development of an indigenous model for road crash prediction if the prevailing traffic is mixed in character and displays a range of vehicle categories in its composition.

2.1 Trends in highway crashes: A global challenge

Gradual increase in automobile growth over the last few decades has resulted simultaneous growth of vehicle crashes. A look into the global status on road safety indicates that every year more than a million deaths are caused by road traffic. Over the last decade the pressing need of providing safe roads for automobiles has been a major look out of the traffic safety researchers. Accordingly, they are in the process of establishing relationships between the level of safety and the quality of service offered to the users. Several studies indicate a number of parameters which have significant impact on road accidents; they are lane and shoulder width, horizontal and vertical alignment, roadside condition, traffic control, exclusive left- and right turn lanes, sight distance and driveway density (HSM, Vogt and Bared, 1988; Harwood et al, 2000; Brimley et al, 2012). Most of the parameters, however, exhibit a nonlinear relationship with traffic safety. At the same time, access density, sight distance, speed limit, and proportions of no-passing zones were found to have high correlation with crash rates (Mayora and Rubio, 2003). Several international studies indicate an increase in overall crashes as speed increases (Moreno et al, 2014; Harwood et al, 2000). Also, impact of speed limit in increasing injury severity is significantly high (Garber and Ehrhart, 2000; Griffin et al, 1998; Matírnez el at, 2013; Elvik, 2009; Allpress and Leland, 2009). A study, however, reported quite the opposite; this tells about no such relationships between crashes and speed (Vogt and Bared, 1988). Moreover, various geometric elements and traffic variables have significant influence on crash occurrences. Wider lanes help in lowering crash rates as it provides a buffer against driver mistakes or distraction (Harwood et al, 2007). Truck lane restriction, on the other hand, also reduces road crashes; couple of international studies ascertained approximate 4% and 17.6% reduction by implementing such restrictions (Kobelo, 2008; Reddy, 2008). Thus, it could be reasonably concluded that truck lane restriction improves traffic operations by eliminating slower vehicles in the traffic stream and thereby, reducing the potential for auto-truck conflicts (Middleton et al, 2003; Khoury and Hobeika, 2007). Furthermore, passing operations on two-lane roads have important effects on road safety; drivers' characteristics, roadway geometry, traffic flow and composition affect such operations. The current passing sight distance operational criteria satisfy the needs of acceptable risk levels (Llorca et al, 2013). A study, on the other hand, compared the passing operations under daytime and night time conditions. Observations indicate that passing at night time is relatively safer; this attributes to the fact that headlights of an opposing vehicle allow a driver to anticipate the vehicle's position before it becomes visible (Cheng-cheng et al, 2007). Access density is another major contributor to road mishaps. Possibility of rear-end collision and also, road side accidents increase as access density increases. Study indicates 4.36% increase in total accidents and 11.9% increase in rear-end accidents with one unit increase of the access density on two-lane highways in mountainous terrain (Persaud et al, 2000). Another major concern to the traffic analysts is crashes on horizontal curves on undivided roads. Several models developed in the past have delineated crashes on horizontal curves on the basis of some parameters which include annual average daily traffic, length of curve, and degree of curvature (Schneider et al, 2009; Torbic et al, 2004). Research indicates that both crash rates and severity of crashes are quite high at horizontal curves (Khan et al, 2013;). This is attributable to either unawareness or underestimation of the driver population about the approaching horizontal curve or the radius or sharpness of the curve (Torbic et al, 2004). Thus, operating speeds on horizontal curves should be made considerably lower than the design speed of the highways (Mclean, 2000; Oña et al, 2013).



(Source: Road Accidents in India, 2013, Open Government Data Platform India)

Fig. 1: Trends in percentage change of population growth, fatal and non-fatal accidents during 2004-14

A number of studies have reported that, in India, about 5 lakh road accidents were fatal out of a total of 10 lakh such road mishaps in 2010, indicating about 56 numbers of accidents per hour which means one accident occurred per minutes. Further, 53% of the people who die in road accident are the most productive age group of 20 to 50 years (Sivakumar and Krishnaraj 2012). The recent edition of *Road Accidents in India* indicates that Indian highways accounts for a share of 53.7% in total road accidents and 62.8% in total number of persons killed in road accidents (Road Accidents in India, 2013). *Figure 1* displays that road accidents have considerably increased over the past few years and significantly, most of those resulted fatalities. Although, percentage change in population growth passed through an insignificant change, it appeases to be alarming if safety aspects are looked into.

2.2 Modelling crash data: A glimpse of statistical fits

Reliance on past data of traffic crashes has been the basis of developing statistical models for crash prediction. Accordingly, a number of regression models have been suggested by the analysts for such modelling; typically, the poisson, poisson lognormal, zero-inflated, negative binomial exhibit appropriateness owing to their flexibility and ease in estimating the parameters. The following section provides a detailed insight of these models in context to regression of crash data.

Poisson regression is a form of regression analysis used to model count data and assume that the response variable has a Poisson distribution. A Poisson regression model is sometimes known as a log-linear model (Washington et al., 2003; Lord et al., 2004).

Negative binomial regression is a generalization of Poisson regression. This regression model is commonly based on the Poisson-gamma mixture distribution. This model is popular because its model the Poisson heterogeneity with an overdispersion (Maycock and Hall, 1984; Miaou, 1994).

Poisson lognormal regression is Full Bayes (FB) hierarchical models as a better way to handle low sample mean, especially in comparison with the traditional Poisson–gamma or negative binomial approaches(Lord and Miranda-Moreno, 2008).

A zero-inflated regression model is a statistical model based on a zero-inflated probability distribution, where the distribution allows frequent zero-valued observations (Washington et al, 2003, 2010).

Conway–Maxwell–Poisson is a generalization of the Poisson distribution and was first introduced by Conway and Maxwell (1962) for modelling queues and service rates.

Gamma regression model is a model based on Gamma distribution, and similar to linear regression. The values are always positive and will not always be integer numbers, so they do not follow a Poisson distribution or Negative Binomial distribution based process (Oh et al, 2006).

Random effects model, also called a variance components model, is a kind of hierarchical linear model. It assumes that the data being analysed is drawn from a hierarchy of different populations whose differences relate to that hierarchy (Johansson, 1996; Shankar et al.1998).

Generalized estimating equation is not actually a regression model but a method used to estimate models with data characterized by serial correlation (Liang and Zeger, 1986).

Negative multinomial model is similar to negative binomial model. But this model is more flexible then negative binomial(Guo, 1996).

Poisson – Weibull model is a mixture of Poisson and Weibull distributions. This model is Similar to most Poisson-based distributions. The PW model is also designed to accommodate the over-dispersion (McCullagh and Nelder, 1993; Raghavachari et al, 199

3. Evaluation of the regression models

There have been a range of possible modelling approaches, which makes an intelligent choice for modelling vehicle crash data challenging. Therefore, the present study calls for an initiative for providing a defensible guidance on application of appropriate regression model. The following section provides detailed insights on different statistical modelling approaches keeping in view the characteristics of crash data: over-dispersion and under-dispersion.

3.1 Over-dispersion in crash data: methodological issues

Poisson regression model is considered fundamental in crash data evaluation (Madanat and Ibrahim 1995). However, researchers have reported that crash data display characteristics that make the function of the simple Poisson regression difficult. Distinctively, Poisson models cannot handle over and under dispersion and they can be disapprovingly affected by low sample-means and can generate biased results in small samples. The negative binomial or Poisson-gamma model is an extension of the Poisson model to overcome possible over-dispersion in the crash data (Maycock and Hall, 1984; Miaou, 1994). In count-data, actual estimates of over-dispersion can be influenced by a variety of factors, such as the clustering of data, unaccounted temporal correlation, and model miss-specification (Cameron and Trivedi, 1998). Recently, some researchers have reported that model-estimated over-dispersion can be greatly minimized by improving the model specification (Miaou and Song, 2005; Mitra and Washington, 2007).Though, this model does have its limitations, particularly in its incapability to handle under-dispersed data and dispersion- parameter-estimation problems when the data are categorized by the low sample-mean values and small sample sizes (Oh, 2006).

At the same time, a few researchers have suggested the use of the Poisson-lognormal model as an alternative to the negative binomial model for modelling crash data (Miaou et al., 2003; Lord and Miranda-Moreno, 2008;). The Poisson-lognormal model acts similar to the negative binomial model, however, the distributional assumption gamma in case of later one. Although Poisson-lognormal potentially offers more flexibility compared to the negative binomial, but there are instances when it exhibit certain limitations; the model estimation is more complex because the poisson lognormal distribution does not have a closed form and it can be negatively affected by small sample sizes and low sample-mean values (Miaou et al, 2003; Lord and Miranda-Moreno 2008).

Negative multinomial, on the other hand, can handle over-dispersion; this model is similar to the negative binomial but more flexible than negative binomial model (Shankar and Ulfarsson 2003). At times, negative binomial experiences correlation problem with observations, which could perhaps be alleviated by the negative multinomial approach (Guo, 1996). However, such models cannot handle under-dispersion and are susceptible to problems in the presence of low sample-means and small sample sizes.

3.2 Under-dispersion in crash data: methodological issues

Oh et al. (2006) has proposed the gamma model which can handle both, over-dispersion and under-dispersion and reduces to the Poisson model when the variance is roughly equal to the mean (Cameron and Trivedi, 1998). Though this model performs well statistically, it is a dual-state model, leading to one of the states, having a long-term mean equal to zero. Shmueli et al. (2005) explored the statistical properties of the Conway–Maxwell–Poisson distribution (Conway and Maxwell 1962), and Kadane et al. (2006) developed conjugate distributions for the parameters. This model can also handle both, over-dispersion and under-dispersion and accordingly, applied in highway-safety research wherein it was found to be comparable to the Poisson gamma model for data characterized by over-dispersion (Lord et al; 2007). However, it is more advantageous to data that are characterized by under-dispersion. (Guikema and Coffelt; 2007)

Furthermore, couple of international studies indicates Zero-inflated models to handle data characterized by a significant amount of zeros or more zeros than the one would expect in a traditional Poisson or negative binomial. Since its inception, the zero-inflated model, both for the Poisson and negative binomial models has been popular among transportation safety analysts (Shankar et al., 1997; Carson and Mannering, 2001).

Most of above studies dealt with the key issues associated with highway crash data as well as the strengths and weaknesses of the various regression models. While the steady march of methodological innovation has substantially improved our understanding of the factors that affect crash-frequencies, it is the prospect of combining evolving methodologies with far more detailed vehicle crash data that holds the greatest promise for the future.

4. SWOT analysis and choice of model form under mixed traffic

On the basis of substantial empirical evidence derived from the work of researchers across the globe, it was evident that for accident data over-dispersion is very common. It, however, aggravates further in the event of heterogeneity in traffic mix. Several studies, over the years, reported the use of Negative Binomial, poisson lognormal and negative multinominal as alternatives to Poisson regression in modelling over-dispersed crash data, of which, negative binomial was observed to meet all the required traits as it effectively account for the extra variation in crash counts; in a way that the variance of the data is greater than the mean.

Table 1: Type of highway crash data and the associated problems with examples

Type of highway crash data	Associated problems	Examples
Over-dispersion	Can violate some of the basic count-data modeling assumptions of some modeling approaches	When the variance exceeds the mean of the highway crash count (Park and Lord, 2007)
Under-dispersion	As with over-dispersion, can violate some of the basic count-data modeling assumptions of some modelling approaches	When the mean of the crash counts on roadway entities is greater than the variance, especially when the sample-mean value is very low. (Oh et al., 2006)
Time-varying explanatory variables	Averaging of variables over studied time intervals ignores potentially important variations within time intervals – which can result in erroneous parameter estimates	The aggregation of data over time periods can lead to a bias, especially in non-linear models. (Washington et al, 2010)
Temporal and spatial correlation	Correlation over time and space causes losses in estimation efficiency	Multiple observations will be correlated over time because many of the unobserved effects associated with a specific roadway entity will remain the same over time (Washington et al 2010).
Low sample-mean and small sample size	Causes an excess number of observations where zero crashes are observed which can cause errors in parameter estimates	Because of the large costs associated with the data collection process, crash data are often characterized by a small number of observations. In addition, crash data for some roadway entities may have few observed crashes which results in a preponderance of zeros (Lord and Bonneson 2005).
Injury severity and crash-type correlation	Correlation between severities and crash types causes losses in estimation efficiency when separate severity-count models are estimated	The most common modelling approach is to consider the frequency of all crashes and deal with the injury severities or crash types separately once the total number of crashes is determined (Lee and Mannering 2002).
Under-reporting	Under-reporting can distort model predictions and lead to erroneous inferences with regard to the influence of explanatory variables	Incomplete reporting of crash data has been known to be a major problem in highway safety analysis (Yamamoto et al. 2008)
Omitted-variables bias	If significant variables are omitted from the model, parameter estimates will be biased and possibly erroneous inferences with regard to the influence of explanatory variables will result	It is often persuasive to develop a simplified model with few explanatory variables. All traditional statistical estimation methods, leaving out important explanatory variables results in biased parameter estimates that can produce erroneous inferences and crash-frequency forecasts (Washington et al., 2003, 2010).
Endogenous variables	If endogenous variables are included without appropriate statistical corrections parameter estimates will be biased and erroneous inferences with regard to the influence of explanatory variables may be drawn	The frequency of ice-related accidents and the effectiveness of ice-warning signs in reducing crashes frequency, this is the endogeneity problem (Carson and Mannering, 2001)
Functional form	If incorrect functional for is used, the result will be biased parameter estimates and possibly erroneous inferences with regard to the influence of explanatory variables	Count-data models assume that explanatory variables influence the dependent variables in some linear manner but non-linear functions better characterize the relationships between crash-frequencies and explanatory variables (Bonneson and Pratt, 2008)
Fixed parameters	If parameters are estimated as fixed when they actually vary across observations, the result will be biased parameter estimates and possibly erroneous inferences with regard to the influence of explanatory variables	The effect of an exposure variable such as the number of vehicle miles travelled over the time period being considered is the same across all roadway segments.(El-Basyouny and Sayed, 2009)

Source: After Lord and Mannering 2010

The present study, thus, made an attempt to identify all the models that could be applied while analysing crash data. Accordingly, an evaluation was made to assess their suitability in context to over-dispersed data (see Table 2). A careful examination of the strengths and weaknesses of these models reveals that Negative binomial, Poisson lognormal, Conway-maxwell poisson, Negative multinomial and Poisson wiebull could be chosen as most promising models for handling over-dispersed data. The negative binomial, however, is frequently used because of its simplicity and ease in the analysis.

A polynomial function is considered as an appropriate model form (Eq. 1) and thus, the present study creates a starting point of a systematic investigation aimed at fitting the function applying negative binomial as an effective regression for crash models.

$$\mu = x^{\beta^0}(1 + x^{\beta^1} + x^{\beta^2} + \cdots + x^{\beta^n}) \qquad [1]$$

Where, x = Independent variables, μ = estimated number of accidents on a roadway segment, β = regression parameters

Typically, crash data of mixed traffic exhibits over-dispersion (Gardner et al., 1995; Hilbe, 2007). Thus, the 'β' parameters for the highways could be estimated by maximizing the log-likelihood function of the negative binomial distribution (Eq. 2 & 3), which takes a dispersion parameter 'α' into account while modelling (Kononov and Allery 2003).

$\mu \epsilon$ Negative binomial,

$\therefore \mathrm{Var} > \mu$,

$\mathrm{Var}(y) = \mu(1+\alpha\mu)$,

$\therefore \sigma = \sqrt{(\mu + \alpha\mu^2)}$

$$L(\alpha,\mu) = \prod_{i=1}^{n} \frac{\tau(\alpha^{-1}+y_i)}{\tau(\alpha^{-1})y_i!} \left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i} \left(\frac{1}{1+\alpha\mu_i}\right)^{\alpha^{-1}} \qquad [2]$$

$$\ln[L(\alpha,\mu)] = \sum_{i=1}^{n} \left\{ \left[\frac{\tau(\alpha^{-1}+y_i)}{\tau(\alpha^{-1})y_i!}\right] + y_i \ln\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right) + \alpha^{-1}\ln\left(\frac{1}{1+\alpha\mu_i}\right) \right\} \qquad [3]$$

Where, y_i = observed number of accidents on roadway segment *i* over a period of a year; $L(\alpha,\mu)$ = Likelihood function

5. Conclusions

Several regression techniques have been proposed by a number of researchers for modelling the crash data. Indian traffic exhibits heterogeneity in its composition and experiences higher rate of road mishaps. Thus, there is aneed of developing an appropriate crash prediction model for analysing the safety performances of roads; this warrants an appropriate regression technique in calibrating crash prediction model. However, in the process of arriving at an appropriate technique, it was felt imperative to study the inherent strengths and weaknesses of the models developed over the years and also their external opportunities and threats while analysing current traffic safety situation of roads.

Notably, a number of statistical models have been developed over the past few decades for the prediction highway crash data. However, several studies have reported that those models largely dependent on the choice of the regression technique applied on the data. Even till last decade, Poisson regression was considered as most widespread and promising probabilistic model. Based on the suggestions made in current literature and also, initiatives of present study, it could be well concluded that Poisson regression fails in the event of over dispersion and calls for an alternative approach in such situation. In an attempt to investigate the compatibility of Negative binomial, it was observed that over-dispersion basically results from specification errors which it reasonably takes into account. Under such data structure, Poisson regression model eventually makes things worse by giving erroneous results particularly when the fundamental errors in the model are coincided.

Table 2: Characteristics of regression models used for analyzing highway crash data

Model type	Advantages	Disadvantages	Suitability for over-dispersed data
Poisson	Most basic model; easy to estimate	Cannot handle over- and under-dispersion; negatively influenced by the low sample-mean and small sample size bias	Not applicable
Negative binomial/ Poisson-gamma	Easy to estimate and can account for over-dispersion	Cannot handle under-dispersion; can be adversely influenced by the low sample-mean and small sample size bias	Applicable
Poisson-lognormal	More flexible than the Poisson-gamma to handle over-dispersion	Cannot handle under-dispersion; can be adversely influenced by the low sample-mean and small sample size bias cannot estimate a varying dispersion parameter	Applicable
Zero-inflated Poisson and Zero-inflated negative binomial	Handles datasets that have a large number of zero-crash observations	Can create theoretical inconsistencies; zero-inflated negative binomial can be adversely influenced by the low sample-mean and small sample size bias	Not applicable
Conway–Maxwell–Poisson	Can handle under- and over-dispersion or combination of both using a variable dispersion parameter	Could be negatively influenced by the low sample-mean and small sample size bias; no multivariate extensions available to date	Applicable
Gamma	Can handle under-dispersed data	Dual-state model with one state having a long-term mean equal to zero	Not applicable
Generalized estimating equation	Can handle under-dispersed data	Dual-state model with one state having a long-term mean equal to zero	Not applicable
Random-effects	Can handle temporal correlation	Determine or evaluate the type of temporal correlation a priori; results sensitive to missing values	Not applicable
Negative multinomial	Can account for over-dispersion and serial correlation; panel count data	Cannot handle under-dispersion; can be adversely influenced by the low sample-mean and small sample size bias	Applicable
Poisson – Weibull	It account for over-dispersion	Cannot handle under-dispersion; can be adversely influenced by the low sample-mean and small sample size bias	Applicable

(Source: After Lord and Mannering 2010; https://victorfang.wordpress.com)

Acknowledgements

The authors are grateful to the anonymous referees for their suggestions to improve the paper. The authors, also, express their gratitude to Manoj Muthai Periyasamy and Janardan Kundu for providing statistical oversight and guidance.

References

Wee BV, Banister D (2016) How to write a literature review paper?, Transport Reviews, 36(2): 278-288, DOI: 10.1080/01441647.2015.1065456

Global status report on road safety time for action (2011). World Health Organization. http://whqlibdoc.who.int/publications/2009/9789241563840_eng.pdf

Global status report on road safety supporting a decade of action (2013). World Health Organization. http://www.who.int/about/licensing/copyright_form/en/index.html

AASHTO, (2010) Highway Safety Manual. American Association of State Highway Transportation Officials, 1500p.

Hadi M A, Aruldhas J, Chow L F, Wattleworth JA (1995) Estimating safety effects of cross- section design for various highway types using negative binomial regression. Transport. Res. Rec. 1500:169-177. http://onlinepubs.trb.org/Onlinepubs/nchrp/cd-22/references/anovasafety.pdf

Lord D, Washington SP, Ivan JN (2005) Poisson, poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accident Anal. Prev. 37:35–46. DOI:10.1016/j.aap.2006.06.004

Lord D (2006) Modeling motor vehicle crashes using poisson-gamma models: examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accident Anal. Prev. 38 (4): 751–766. DOI:10.1016/j.aap.2006.02.001

Cheng Zhong (2008) Wyoming rural road safety program. MS Thesis submitted in The Graduate School of The University of Wyoming. UMI-1465188

Vogt A and Bared JG (1988) Accident prediction models for two-lane rural roads: Segments and intersections. Transport. Res. Rec.1635: 18-29. DOI: http://dx.doi.org/10.3141/1635-03

Harwood DW, Council FM, Hauer E, Hughes WE, Vogt A (2000) Prediction of the expected safety performance of rural two-lane highways. FHWA-RD-99-207, Federal Highway Administration, U.S. department of transportation. http://www.fhwa.dot.gov/p...rch/safety/99207/99207.pdf

Brimley BK, Saito M, Schultz GG (2012) Calibration of highway safety manual safety performance function development of new models for rural two-lane two-way highways. Transport. Res. Rec. 2279: 82–89. DOI: 10.3141/2279-10

Mayora J, Rubio R (2003) Relevant variables for crash-rate prediction on spain's two-lane rural roads. Presented at 82nd Annual Meeting of the Transportation Research Board http://www.ltrc.lsu.edu/TRB_82/TRB2003-000769.pdf.

Moreno AT, Llorca C, Sayed T, Garcia A (2014) Field evaluation of traffic performance measures for two-lane highways in Spain. Innovative Applications of the Highway Capacity Manual 2010.

Garber N, Ehrhart A (2000) Effect of speed, flow, and geometric characteristics on crash frequency for two-lane highways. Transport. Res. Rec. 1717: 76–83. DOI: 10.3141/1717-10

Griffin L, Pendleton O and D Morris (1998) An Evaluation of the Safety Consequences of Raising the Speed Limit on Texas Highways to 70 Miles per Hour. Texas Transportation Institute, Texas A&M University System, 1998.

Matírnez A, Mántaras DA , Luque P (2013) Reducing posted speed and perceptual countermeasures to improve safety in road stretches with a high concentration of accidents. Safety Sci.60: 160–168. DOI: http://dx.doi.org/10.1016/j.ssci.2013.07.003

Elvik R (2009) The power model of the relationship between speed and road safety.

Allpress J, Leland L (2009) Reducing traffic speed within roadwork sites using obtrusive perceptual countermeasures. Accident Anal. Prev. 42: 377–383. DOI:http://dx.doi.org/10.1016/j.aap.2009.08.014

Harwood DW, Bauer KM, Richard KR, Gilmore DK, Graham JL, Potts IB, Torbic DJ, Hauer E (2007) Phases I and II: methodology to predict the safety performance of urban and suburban arterials. NCHRP Web-Only Document 129, Transportation Research Board of the National Academies, Washington, DC.

Kobelo D, Patrangenaru V, Mussa R (2008) Safety analysis of florida urban limited access highways with special focus on the influence of truck lane restriction policy J. Transp. Eng. 134: 297-306. DOI: 10.1061/(ASCE)0733-947X(2008)134:7(297),

Reddy GV, Thakkar J, Vargas F (2008) The effect of lane use restriction for trucks on traffic operations and safety. Enhancing Transportation Safety in the 21st Century ITE International Conference, 2008.

Middleton D, Clayton A, Quiroga C, Jasek D (2003) Truck accommodation design guidance: final report. Texas Transportation Institute. http://d2dtl5nnlpfr0r.cloudfront.net/tti.tamu.edu/documents/0-4364-1.pdf

Khoury J, Hobeika A (2007) Assessing the risk in the design of passing sight distances. J. Transp. Eng. 133 (6): 370-377. DOI: 10.1061/(ASCE)0733-947X(2007)133:6(370).

Llorca C, Moreno AT, García A, Pérez-Zuriaga AM (2013) Daytime and nighttime passing maneuvers on a two-lane rural road in spain. Transport. Res. Rec. 2358: 3–11. DOI:10.3141/2358-01

Cheng-cheng T, Tie-jun Z, Ling-tao W (2007) Safety impact on accesses in two-lane highways. J. Highway Transp. Res. Dev. 104(2): 103-107.

Persaud B, Retting RA, Lyon C (2000) Guidelines for Identification of Hazardous Highway Curves. Transport. Res. Rec. 1717: 14–18.

Schneider WH, Zimmerman KH, Boxel DV, Vavilikolanu S (2009) Bayesian analysis of the effect of horizontal curvature on truck crashes using training and validation data sets. Transport. Res. Rec. 2096: 41–46.

Torbic DJ, Harwood DW, Gilmore DK, Pfefer R, Neuman TR, Slack, KL, Hardy KK (2004) NCHRP Report 500: Guidance for Implementation of the AASHTO Strategic Highway Safety Plan. Volume 7: A Guide for Reducing Collisions on Horizontal Curves. Transportation Research Board of the National Academies, Washington, D.C.

Khan G, Bill AR, Chitturi VM, Noyce AD (2013) Safety evaluation of horizontal curves on rural undivided roads. Transport. Res. Rec. 2386: 147–157.

Mclean (2000) Speed prediction of two-lane highway. Research, Development, and Technology Turner-Fairbank Highway Research Center 6300 Georgetown Pike. VA 22101-2296 .Federal highway administration, U.S. department of transportation.

Oña JD, Garach L, Calvo F, Muñoz TG (2013) Relationship between predicted speed reduction on horizontal curves and safety on two-lane rural roads in Spain. J. Transp. Eng. 140(30): DOI:10.1061/(ASCE)TE.1943-5436.0000624, 04013015.

Sivakumar T, Krishnaraj R (2012) Road traffic accidents (rtas) due to drunken driving in india – challenges in prevention. International Journal of Research in Management & Technology. 2:4. http://www.iracst.org/ijrmt/papers/vol2no42012/7vol2no4.pdf

Road Accidents in India (2013) Transport Research Wing, Ministry of Road Transport and Highways, Government of India. New Delhi.

Washington SP, Karlaftis MG, Mannering FL (2003) Statistical and econometric methods for transportation data analysis. Chapman Hall/CRC, Boca Raton, FL.

Lord D, Washington SP, Ivan JN (2004) Poisson, poisson-gamma and zero inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accident Anal. Prev. 37 (1): 35–46. DOI:http://dx.doi.org/10.1016/j.aap.2004.02.004

Maycock G, Hall RD (1984) Accidents at 4-arm roundabouts. TRRL Laboratory Report 1120, Transportation and Road Research Laboratory, Crowthorne, UK.

Miaou SP (1994) The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. Accident Anal. Prev. 26 (4): 471–482. DOI:http://dx.doi.org/10.1016/0001-4575(94)90038-8

Lord D, Miranda-Moreno LF (2008) Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: a Bayesian perspective. Safety Sci. 46 (5): 751–770. DOI:http://dx.doi.org/10.1016/j.ssci.2007.03.005

Washington SP, Karlaftis MG, Mannering FL (2010). Statistical and econometric methods for transportation data analysis, second ed. Chapman Hall/ CRC, Boca Raton, FL.

Conway RW, Maxwell WL (1962) A queuing model with state dependent service rates. J. Ind. Engineering. 12: 132–136.

Oh J, Washington SP, Nam D (2006) Accident prediction model for railway-highway interfaces. Accident Anal. Prev. 38 (2): 346–356. DOI:http://dx.doi.org/10.1016/j.aap.2005.10.004

Johansson P (1996) Speed limitation and motorway casualties: a time series count data regression approach. Accident Anal. Prev. 28 (1): 73–87. DOI:http://dx.doi.org/10.1016/0001-4575(95)00043-7

Shankar VN, Albin RB, Milton JC, Mannering FL (1998) Evaluating median cross-over likelihoods with clustered accident counts: an empirical inquiry using random effects negative binomial model. Transport. Res. Rec. 1635: 44–48. DOI: http://dx.doi.org/10.3141/1635-06

Liang KY, Zeger SL (1986) Longitudinal data analysis using generalized linear models. Biometrika 73 (1): 1322. http://www.biostat.jhsph.edu/~fdominic/teaching/bio655/references/extra/liang.bka.1986.pdf

Guo G (1996) Negative multinomial regression models for clustered event counts. Sociol. Methodol. 26 (1): 113–132. URL: http://www.jstor.org/stable/271020

McCullagh P, Nelder JA (1983) Generalized Liner Models. Chapman and Hall,London, UK.

Raghavachari M, Srinivasan A, Sullo P (1997) Poisson mixture yield models for integrated circuits: a critical review. Microelectron. Reliab, 37 (4): 565–580. DOI:doi:10.1016/S0026-2714(96)00064-9

Madanat S, Ibrahim WHW (1995) Poisson regression models of infrastructure transition probabilities. J. Transp. Eng-Asce. 121:1-09. DOI:http://dx.doi.org/10.1061/(ASCE)0733-947X(1995)121:3(267)

Cameron AC, Trivedi PK (1998) Regression analysis of count data. Cambridge University Press, Cambridge, UK.

Miaou SP, Song JJ (2005) Bayesian ranking of sites for engineering safety improvements: decision parameter, treatability concept, statistical criterion and spatial dependence. Accident Anal. Prev 37 (4): 699–720. DOI:10.1016/j.aap.2005.03.012.

Mitra S, Washington S (2007) On the nature of over-dispersion in motor vehicle crash prediction models. Accident Anal. Prev. 39 (3): 459–468. DOI:doi:10.1016/j.aap.2006.08.002

Miaou SP, Song JJ, Mallick BK (2003) Roadway traffic crash mapping: a space-time modeling approach. Journal of Transportation and Statistics 6 (1); 33–57.

Shankar VN, Ulfarsson GF, Pendyala RM, Nebergall MB (2003). Modeling crashes involving pedestrians and motorized traffic. Safety Sci. 41 (7): 627–640. DOI:http://dx.doi.org/10.1016/S0925-7535(02)00017-6

Shmueli G, Minka TP, Kadane JB, Borle S, Boatwright P (2005) A useful distribution for fitting discrete data: revival of the Conway–Maxwell–Poisson distribution. Journal of the Royal Statistical Society: Series C. 54 (1): 127–142.

Kadane JB, Shmueli G, Minka TP, Borle S, Boatwright P (2006) Conjugate analysis of the Conway–Maxwell–Poisson distribution. Bayesian Analysis. 1(2), 363–374. http://projecteuclid.org/download/pdf_1/euclid.ba/1340371067

Lord D, Washington SP, Ivan JN (2007) Further notes on the application of zero inflated models in highway safety. Accident Anal. Prev. 39 (1): 53–57. DOI: http://dx.doi.org/10.1016/j.aap.2006.06.004

Guikema SD, Coffelt JP (2007) A flexible count data regression model for risk analysis. Risk analysis. 28 (1): 213–223. DOI: 10.1111/j.1539-6924.2008.01014.x

Shankar V, Milton J, Mannering FL (1997) Modeling accident frequency as zero-altered probability processes: an empirical inquiry. Accident Anal. Prev. 29 (6): 829–837. DOI:10.11149/jkaoh.2015.39.1.9

Carson J, Mannering F (2001) The effect of ice warning signs on accident frequencies and severities. Accident Anal. Prev. 33 (1): 99–109. DOI: http://dx.doi.org/10.1016/S0001-4575(00)00020-8

Park ES, Lord D (2007) Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. Transport. Res. Rec. 2019: 1–6. DOI: http://dx.doi.org/10.3141/2019-01

Lord D, Bonneson JA (2007) Development of accident modification factors for rural frontage road segments in Texas. Transport. Res. Rec. 2023: 20–27. DOI: http://dx.doi.org/10.3141/2023-03

Lee J, Mannering F (2002) Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accident Anal. Prev. 34 (2): 149–161. DOI: http://dx.doi.org/10.1016/S0001-4575(01)00009-4

Yamamoto T, Hashiji J, Shankar V (2008) Underreporting in traffic accident data, bias in parameters and the structure of injury severity models. Accident Analysis and Prevention 40 (4); 1320–1329.

Bonneson JA, Pratt MP (2008) Procedure for developing accident modification factors from cross-sectional data. Transport. Res. Rec. 2083: 40–48. DOI: http://dx.doi.org/10.3141/2083-05

El-Basyouny K, Sayed T (2009) Accident prediction models with random corridor parameters. Accident Anal. Prev. 41 (5): 1118–1123. DOI: http://dx.doi.org/10.1016/j.aap.2009.06.025

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Trans. Res. Part-A, 44, 291–305. DOI:10.1016/j.tra.2010.02.001.

Gardner W, Mulvey EP, Shaw EC (1995) Regression analyses of counts and rates: poisson, overdispersed poisson, and negative binomial models. Psychological Bulletin, 118, 392–404. DOI: http://dx.doi.org/10.1037/0033-2909.118.3.392

Hilbe JM (2007) Negative binomial regression. New York: Cambridge.

Kononov J, Allery B (2003) Level of service of safety-a conceptual blueprint and the analytical framework. Transport. Res. Rec. 1840: 57-66. DOI: http://dx.doi.org/10.3141/1840-07

https://victorfang.wordpress.com/2011/05/10/advantages-and-disadvantages-of-logistic-regression/ (Access date: March 31, 2016)