

# Statistical Models for Count Data with Applications to Road Accidents in Ghana

A. Y. Omari-Sasu, Adjei Mensah Isaac\*, R. K. Boadi

Department of Mathematics, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

**Abstract** Road accidents in Ghana seems to be on ascendency and the root causes have been attributed to issues such as human errors and superstitions. Since the occurrences of these accidents are discrete, they are often modelled using count regression models. It is therefore the purpose of this study to determine an appropriate count regression model that adequately fits road accidents in Ghana and determine the key predictors using the appropriate model with respect to the expected number of persons killed in road accidents. Several models were compared to fit count data that encounter the field of transportation. These models include Poisson, Negative Binomial (NB) and Conway-Maxwell-Poisson (CMP) models. In order to compare the performance of these models, the various model selection methods such as Deviance goodness of fit test, Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) were employed. Because the values of Deviance goodness of fit test, AIC and BIC for the NB model was the smallest as compared to that of the Poisson and CMP models, it appeared that, the NB model performed best than the Poisson and CMP models. Base on the appropriate model selected (NB model), the key predictors that contributed significantly and also had a high effect on the expected or mean number of persons killed in road accidents within a particular period were Head-on collision as Collision type, Improper overtaking and Loss of control as Driver errors, Bus/Minibus as Type of vehicle, Fog/Midst as Weather condition and Night with street lights off as Light condition.

**Keywords** AIC, BIC, Goodness of fit, Poisson model, Negative binomial model, CMP model, Count Data

## 1. Introduction

Road accident is defined as any activity that distracts the normal trajectory of moving vehicles in a manner that causes instability in the free flow of the vehicle. Road accidents in Ghana has been a serious concern to many Ghanaians in recent times. As a result of the tremendous effect of road accidents on human lives, properties and the environment as a whole, numerous researchers have come out with the causes, effects and recommendations to road accidents. These causes include drunk driving, over speeding and machine failure (Sagberg, Fosser and Saetermo (1997), Adams (1982) and National Road Safety Commission (2009)). Despite all these being done, yet every year the Ghana Statistical Service, Road Safety Commission and other organizations would report there is an increase in road accidents in the country (Annual Report, National Road Safety Commission. 2009).

Researchers in recent times have been modelling road accidents with crash prevention models in various parts of the world. However, it is extremely tedious to just apply

models which have worked in other area to data obtained from different countries due to the variations in the various factors pertaining in different countries (Fletcher et al., 2006). There has not been much statistical research work conducted with respect to accidents on urban roads in Ghana and this might have been as a result of inadequate information available on accidents on urban roads and its impact on human lives and the environment as a whole in the country.

Salifu (2004) developed a forecasting model for traffic crashes for urban junctions with no road signs, Afukaar and Debrah (2007) on the other hand modelled traffic crashes for signalized urban junctions in Ghana and Akaah (2011) additionally modelled traffic crashes on rural highways in Ashanti region. Road accidents have always been attributed to human errors such as high alcoholic content in blood stream of some drivers, over speeding, wrong-overtaking among others. It has also been linked to poor road networks, poor surfacing of roads, witchcraft and death dying nature of some vehicles which ply the roads. It is however surprising that in spite of the numerous factors identified by researchers as causes of road accidents in Ghana and its consequences on human lives and properties; nobody has modelled the causes of accidents on urban roads and their contributions to the death of casualties in Ghana to authenticate the contributions of each of these factors to the casualties. It is therefore in light of this that this research is conducted in order to

\* Corresponding author:

kwameatoapoma@gmail.com (Adjei Mensah Isaac)

Published online at <http://journal.sapub.org/statistics>

Copyright © 2016 Scientific & Academic Publishing. All Rights Reserved

determine the appropriate count regression model that fits accident data with respect to the expected number of persons who will be killed via road accidents on urban roads in Ghana and additionally use the appropriate count regression model to investigate the key predictors of accidents that results in the death of casualties.

## 2. Related Works

Numerous road accident models have been developed to estimate the expected number of accident frequencies on roads as well as to identify the various factors associated with the occurrence of accidents. It is not possible for regression models to account for each and every factor that affects accident occurrences (Persaud and Dzibik, 1993). Previous researchers have focused on the no-behavioural factors such as traffic flow characteristics, road geometry and environmental conditions. Persaud and Dzibik (1993), as the first accident modellers who worked on multilane road, investigated the relationship between freeway crash data and accident volumes. By using the hourly traffic, the model indicated that, a higher accident risk is associated with congestion and afternoon rush hour. Shankar et al., (1995) modelled the monthly accident frequency of rural motorway as a function of geometrics, weather conditions and their interaction and found it dangerous for the areas with large rainfall and snowfall to have steep grades and tight horizontal curves.

Abdel and Radwan (2000) attempted to use Poisson regression model and then rejected it with the reason that, different mean and variance value of the dependent variable showed over-dispersion in the accident data. Consequently, the Negative binomial model also called the Poisson-Gamma mixture model was then adopted as a superior alternative model to accommodate the vehicle accident analysis for rural highways, arterial roadways, urban roads and rural motorways (Shankar et al., 1995; Lord, 2005; Montella, 2008; Pemmanaboina, 2005). Recently, a number of approaches have been proposed in the domain of accident models seeking improvement from traditional methods earlier discussed. For instance, El-Basyouny and Sayed (2006) proposed a modified Negative binomial regression technique which improved goodness of fit. Caliendo et al, (2007) on the other hand applied Negative Multinomial distribution to model multiple observations in the same road section at different years that may not be mutually independent. The generalized estimating equation was also used by researchers such as Abdel and Abdalla (2004) and Lord and Persaud (2000).

In New Zealand, the generalized linear models (GLMs) have been used in accident research of different intersection types as a well as for two-lane roads. The earlier work usually emphasized on the effect of traffic volume. So far the effort to model accident likelihood on motorway in New Zealand have been minimized except the flow-only models presented by Turner (2011) and the Economic Evaluation

Manual (2010).

In order to obtain a valid and very accurate model that can fit accident data comparison and prediction, other researchers upon identifying the flaws in previous models tried different means by including more factors in the accident data analysis. Livneh and Hakkert (1972) researched into road accidents in Israel using employment and population data. On the other hand Susan and Partyka (1984) also modelled road accidents with the help of employment and population data. Research conducted by Andreassen (1985) raised serious objection to the use of death per vehicles licensed in order to make international comparison of road accident fatalities with the reason that it was found out that the two parameters were not linearly related overtime. As a result, he came out with a general formula;

$$D = \text{const} \tan t \times N^m \times P^{m_2} \quad (1)$$

which could be used to predict the number of deaths in road accidents. Where D is the number of deaths in road accidents, N is the number of vehicles in use, P is the population of the country by Andreassen was how to determine the constant term and the indices which might vary from one country to the other.

Additionally, a research by Minter (1987), discussed the application of two accident models which were developed by Wright (1998) for road safety problems and finally came out with new model for estimating road accidents in the United Kingdom. Pramada and Sarkar (1997) also in their research used road length as an additional parameter and established a model for road accidents. A general model was also presented by Jamal and Jamil (2001) to predict road accidents fatalities.

Jovanis and Chang (1986), Joshua and Garber (1990) and Miaou and Lum (1993) demonstrated that conventional linear regression models are not appropriate for modelling vehicle accidents events on roadways and that, test statistics from these models are often erroneous. They therefore concluded that, Poisson and Negative binomial regression models are more appropriate tool in accident modelling. The inadequacy of linear regression models uncovering the relationship between vehicular accidents and roadway characteristics have therefore led to numerous applications of Poisson and Negative binomial regression models (Shankar, Mannering and Barfield, 1995, Poch and Mannering, 1998).

Shankar, Mannering and Barfield (1995) used both Poisson (when the data were not significantly over-dispersed) and the Negative binomial (when the data were over-dispersed) to explore the frequency of rural freeway accidents with information on roadway geometry and weather-related environmental factors. Separate regressions of specific accident types, as well as overall accident frequency were modelled. The estimation results showed that, the Negative binomial regression model was the appropriate model for all accident types, with the exception

of those involving overturned vehicles. Poch and Mannering (1996) also demonstrated that Negative binomial regression model was the appropriate model isolating the traffic and geometric elements that influence accident frequencies. Milton and Mannering (1998) on the other hand used the Negative binomial regression model as a predictive toll to evaluate the relationship among highway geometry, traffic related elements and motor-vehicle accident frequencies.

Shankar, Milton and Mannering (1997) argued that the traditional application of Poisson and Negative binomial regression models did not address the probability of zero-inflated counting process. They distinguished the truly safe road section (zero accident state) from the unsafe section (non-zero accident state with the probability of having zero observed accidents) to show that zero-inflated model structure is often appropriate for estimating the accident frequency of road section.

Some works have been done by earlier researchers in Ghana with respect to road accidents in the country. For instance, Oppong (2012) used Poisson count regression to model the number of persons who are killed by road accidents in Ghana. In addition, Salifu (2007) used generalized linear models to predict road accidents in unsignable urban junctions. Afukaar et al, (2007) and Akaah and Salifu (2011) additionally applied Poisson, Negative Binomial and Log-linear regression models in modelling road accidents in Ghana.

### 3. Method

This research work mainly utilized secondary data obtained from the Building and Road Research Institute (BRRI) of the Council of Scientific and Industrial Research (CSIR). The data for the research was originally collected with the help of the Police accident report by the Motor Traffic and Transport Unit (MTTU) of the Ghana Police Service. This research work considered a data for five (5) years period from 2009 to 2013. The mean or expected number of persons killed in individual road accidents for the five years period was used as the dependent or response variable and other categorical variables such as collision type, weather conditions, light condition, type of vehicle and driver error as, the explanatory or independent variables.

#### 3.1. Regression Models

Count data, such as accident fatalities are better modelled using Poisson, Negative Binomial and Conway- Maxwell-Poisson regression models. Regardless of whether the assumed model is Poisson, Negative Binomial or Conway-Maxwell-Poisson, it will be assumed that the occurrences will be independent of each other. The three (3) types of count regression models are briefly explained as follows:

##### 3.1.1. Poisson Regression

The most basic model for event counts is the Poisson

regression model. If the variance of the counts approximately equals the mean counts, then the Poisson regression model is expressed as:

$$P(y_i / x_i) = \frac{\exp(-\theta_i) \theta_i^{y_i}}{y_i!} \quad \text{for } y_i = 0, 1, 2, 3, \dots \quad (2)$$

where  $y_i$  is the number of counts (persons killed in accidents) for a particular period of time  $i$ ,  $\theta_i$  is the expected or mean number of counts (persons killed in road accidents) per period, which can be modelled as;

$$\theta_i = \exp(x_i' \beta) \quad (3)$$

where  $x_i$  is the vector of the explanatory or independent variables and  $\beta$  is the vector of unknown regression parameters.

Equation 3 above as a result gives the indication that a unit increase in an  $x_i$  increases  $\theta_i$  by a multiplicative factor of  $\exp(\beta_i)$ . The main constraint of the Poisson regression model is that, the mean and the variance are approximately equal, that is;

$$E(y_i / x_i) = \text{Var}(y_i / x_i) = \theta \quad (4)$$

As a result when there is heterogeneity or over-dispersion (when the variance increases faster than what the Poisson regression allows), the Poisson regression model does not work well hence there is the need to fit a parametric model that is more dispersed than the Poisson model and a natural choice is the Negative binomial and the Conway-Maxwell-Poisson models.

The Log-likelihood function of the Poisson model is expressed as;

$$\ell = \sum_{i=1}^n (-\theta_i + y_i \ln \theta_i - \ln y_i!) \quad (5)$$

By substituting equation 2 into equation 5, we further obtain the Log-likelihood function as;

$$\ell = \sum [-e^{x_i \beta} + y_i x_i \beta - \ln y_i!] \quad (6)$$

In order to estimate the regression coefficients by the Maximum Likelihood Estimation (MLE) procedure, the derivative of the Log-likelihood that is relative to the  $\beta_s$  must be set to zero as;

$$\frac{\partial \ell}{\partial \beta} = \sum_{i=1}^n (y_i - e^{x_i \beta}) x_i = 0 \quad (7)$$

Estimating the regression coefficient in the Poisson regression model is not obtained from a direct equation but rather the Newton Raphson method is used for estimating the parameters that are unknown in the model (Anseline, 2002).

### 3.1.2. Negative Binomial Model

The Negative Binomial model can be obtained from the mixture of Poisson and Gamma distributions and is expressed as;

$$P(y_i / x_i) = \frac{\Gamma(y_i + \alpha^{-1})}{y_i! \Gamma(\alpha^{-1})} \left[ \frac{1}{1 + \alpha\theta_i} \right]^{\alpha^{-1}} \left[ \frac{\alpha\theta_i}{1 + \alpha\theta_i} \right]^{y_i} \quad (8)$$

for  $i = 0, 1, 2, \dots$

where  $y_i$  is the number of road accidents for a road segment  $i$  and  $\theta_i$  is the mean or expected number of persons killed in a road accidents per period, which can be expressed as;

$$\theta_i = \exp(x_i' \beta)$$

The conditional mean and variance of the Negative Binomial distribution are  $E(y_i / x_i) = \theta_i$  and  $Var(y_i / x_i) = \theta_i(1 + \alpha\theta_i) > E(y_i / x_i)$  respectively. Hence the NB model is over-dispersed and allows extra variation relative to the traditional Poisson model. It has more desirable properties than the Poisson model (Chin and Quddus, 2003).

The variance of the Negative Binomial is significantly greater than the mean.

The NB model,  $\alpha$  represents the dispersion parameter which allows or indicates the degree of over-dispersion. For instance, if  $\alpha = 0$ , the NB model reduces to the traditional Poisson model.

The Log-likelihood function of the NB model is obtained from the following equation;

$$\ell = \sum_{i=1}^n \left\{ \sum_{j=1}^{y_i-1} [\ln(y_i + \alpha^{-1}) - \ln y_i! + \alpha^{-1} [\ln(\alpha^{-1}) - \ln(\alpha^{-1} + e^{x_i\beta})]] + y_i [x_i\beta - \ln(\alpha^{-1} + e^{x_i\beta})] \right\} \quad (9)$$

In order to estimate  $\beta$  and  $\alpha$  as in the Poisson model, the iteration procedure or method of Newton Raphson is applied (Lee and Mannering, 2002).

### 3.1.3. Conway-Maxwell-Poisson Model

The Conway-Maxwell-Poisson model (CMP) is a discrete probability distribution that generalizes the Poisson distribution by adding a parameter to model over-dispersion and also a member of the exponential family. The Conway-Maxwell-Poisson has a probability mass function of the form:

$$P(Y = y_i) = \frac{\theta_i^{y_i}}{(y_i!)^\varphi} \frac{1}{Z(\theta_i, \varphi)} \quad (10)$$

Where

$$Z(\theta_i, \varphi) = \frac{\theta^k}{(k!)^\varphi}$$

For  $\theta > 0$  and  $\varphi \geq 0$ . The function  $Z(\theta_i, \varphi)$  is an infinite

series that converges for  $\theta > 0$  and  $\varphi \geq 0$ .

The CMP distribution shows that, there is a non-linear relationship between the ration of successive probabilities as displayed by;

$$\frac{P(Y = y)_i}{P(Y = y_i - 1)} = \frac{\theta_i}{y_i^\varphi} \quad (11)$$

The mean and the variance of the CMP distribution can be represented respectively as;

$$E[Y] = \theta_i \frac{\partial \log(Z(\theta_i, \varphi))}{\partial \theta_i} \quad (12)$$

$$Var(Y) = \theta_i \frac{\partial E[Y]}{\partial \theta_i} \quad (13)$$

Since the moments of the distribution do not have a closed form, Schmueli et al., (2005) showed that (26) can be approximated as;

$$E[Y] \approx \theta^\frac{1}{\varphi} - \frac{\varphi - 1}{2\varphi} \quad (14)$$

and Sellers and Schumeli (2008) also showed that (27) can also be approximated as;

$$Var(Y) \approx \frac{E[Y]}{\varphi} \quad (15)$$

### 3.1.4. Conway-Maxwell-Poisson Generalized Linear Model

The Conway-Maxwell-Poisson distribution was extended into the Classical General Linear Model framework by Sellers and Schumeli (2008). Hence the CMP distribution is a member of the linear exponential family as displayed by;

$$P(Y = y_i) = \exp(y_i \log(\theta_i) - \log(Z(\theta_i, \varphi)) - \varphi \log(y_i!)) \quad (16)$$

As compared to the Poisson case, the nuisance function is assumed to be normalized and has the logarithm link function. The log-likelihood function is therefore given as;

$$\ell(\beta / \theta_i, \varphi) = \sum_{i=1}^n [y_i \log(\theta_i) - \log Z(\theta_i, \varphi) - \varphi \log(y_i!)] \quad (17)$$

The log-likelihood function of the CMP model can be computed in few ways that are different. In the case where Newton-Ralphson algorithm is employed, the log-likelihood function must be maximized under the constraint (Sellers and Schmueli, 2008).

## 3.2. Selecting Over-Dispersed Models

In order to test whether the data is over-dispersed or not, we test the following hypothesis;

$$H_o : \alpha = 0$$

$$H_a : \alpha \neq 0$$

where  $\alpha$  is a dispersion parameter.

The corresponding test statistic is;

$$z = \frac{\hat{\alpha}}{SE(\hat{\alpha})} \quad (18)$$

Under  $H_0$  and for large sample size, the  $z$  has an approximate standard normal distribution. We therefore reject  $H_0$  at alpha level of significance if  $|z| > Z_{\frac{\alpha}{2}}$ .

If we reject the null hypothesis as a result then it gives an indication that the data is over-dispersed hence the Negative Binomial and the CMP models will be more appropriate compared to the Poisson model.

### 3.3. Model Specification

Models considered in this research work has the mean or expected number of persons killed in road accidents within a particular period of time as a function of the categorical variable; collision type, driver error, type of vehicle, weather condition and light condition. Each model parameterizes;

$$\theta_i = \exp(\beta_0 + \beta_1 CT + \beta_2 DE + \beta_3 TV + \beta_4 LC + \beta_5 WC) \quad (19)$$

Where  $i = 1, \dots, n$  individuals, CT denotes the Collision Type, DE represents the Driver Errors, TV is the Type of Vehicle, LC as the Light Condition and WC representing the Weather Condition.

### 3.4. Parameter Estimation

The Maximum Likelihood Estimation (MLE) method has been considered due to the fact that all the models involved in the research work can be estimated using the MLE procedure. To evaluate the models involved in the research work, it is necessary to examine the significance of the variables included in the model. For a better model, the estimated regression coefficients have to be statistically significant.

### 3.5. Model Selection Methods

#### 3.5.1. Goodness of Fit

After fitting the various models involve in the research work to the data obtained, it very necessary or important to check the overall fit as well as the quality of the fit of the respective models. The quality of the fit between the observed values ( $y$ ) and the predicted values ( $\theta$ ) can be measured by the various test statistics, but one useful statistics is called the deviance goodness of fit test which is defined as;

$$D(y: \hat{\theta}) = -2 \sum_{i=1}^n y_i \ln \left( \frac{y_i}{\hat{\theta}_i} \right) - (y_i - \hat{\theta}_i) \quad (20)$$

Where  $y_i$  is the number of events (observed values or counts),  $n$  is the number of observations and  $\hat{\theta}$  represents

the fitted means (predicted values) of the models.

For a better model, one must expect a smaller value of the Deviance. Hence the smaller the value of the deviance of a specific model the better the model or the more statistically significant the model becomes.

#### 3.5.2. Akaike's Information Criterion (AIC)

Akaike's Information Criterion (AIC, Akaike, 1973) is a measure of the relative quality of a statistical model for a given data. That is, given a collection of models for a data, AIC estimates the quality of each model, relative to other models. Hence, AIC provides a means for model selection. For any statistical model, the AIC value is computed using the relation;

$$AIC = -2L + 2k \quad (21)$$

Where  $L$  is the maximized value of the likelihood function and  $k$  is the number of parameters in the model. For the best model with respect to the AIC one must expect the lowest or the smallest AIC value.

#### 3.5.3. Bayesian Information Criterion (BIC)

Bayesian Information Criterion (BIC) is a criterion for model selection among a finite set of models. It is based in the part on the likelihood function and it is closely related to the Akaike's Information Criterion (AIC). Mathematically the BIC is an asymptotic result derived under the assumption that, the data distribution is in the exponential family.

Let;

$x$  = the observed data

$n$  = the number of data points in  $x$ , the number of observation or equivalently the sample size

$k$  = the number of free parameters to be estimated. For instance if the model under consideration is linear regression, then  $k$  is the number of regressors or independent variables, including the intercept.

$p(x/M)$  = the marginal likelihood of the observed data given the model  $M$ .

$\hat{L}$  = the maximized value of the likelihood function of the model  $M$  that is  $\hat{L} = p(x/\hat{\theta}, M)$  where  $\hat{\theta}$  is the parameter value that maximizes the likelihood function.

Then the formula for the Bayesian Information Criterion (BIC) is given by;

$$-2 \ln p(x/M) \approx BIC = -2 \ln \hat{L} + k(\ln(n) - 2 \ln(2\pi)) \quad (22)$$

For a best model using BIC one must expect the lowest BIC value.

## 4. Results and Discussions

A review of the road accidents data with respect to the number of persons killed, revealed that there were 79,114 road accidents in Ghana from 2009-2013 which killed

10,836 people. This as a result gives indication that on the average, 15,823 road accidents occur every year and 2,167 lives were lost through road accidents.

The table 1 below shows the time in years for which road accidents that resulted in death of people occurred and additionally presents the total number of persons killed in road accidents in road accidents yearly from 2009-2013.

**Table 1.** Distribution of Persons Killed by Road Accidents Annually from 2009-2013 in Ghana

Years	Number of Persons Killed
2009	2,237
2010	1,986
2011	2,199
2012	2,165
2013	2,249
Total	10,836

**Table 2.** Parameter estimates of Poisson Regression Model for the Number of Persons Killed in Road Accidents from 2009-2013

Variable	Parameter Estimates	Standard Error	Rates $Exp(\beta)$	P-value
Intercept	1.50417	0.47481	4.50	0.00000786
COLLISION TYPE				
Overturn	0.00000			
Rear-end	-0.98795	0.05492	0.37	0.09702
Ran-off	0.28387	0.04365	1.33	0.03514
Side swipe	-1.60780	0.06392	0.20	0.91740
Right angle	-1.76591	0.07485	0.17	0.20404
Head on	0.93386	0.03380	2.54	0.00158
Hit object on road	0.92703	0.12729	2.53	0.85036
Hit object off road	-2.21040	0.09100	0.11	0.29680
DRIVER ERROR				
Inexperience	0.00000			
Inattentive	-2.80611	0.12676	0.45	0.00874
Too fast	0.26966	0.04019	1.31	0.13142
Too close	-3.35818	0.16502	0.03	0.07412
No signal	0.07581	0.04201	1.08	0.25692
Improper overtaking	0.22388	0.04059	1.25	0.04291
Fatigue/Asleep	-1.87180	0.08287	0.15	0.99726
Loss of control	-0.84303	0.05517	0.43	0.02658
TYPE OF VEHICLE				
Car	0.00000			
Goods vehicle	-0.55358	0.03504	0.57	0.00127
Bus/Minibus	0.30358	0.02992	1.36	0.02015
Motor cycle	-1.29629	0.04569	0.27	0.11630
Pickup	-1.59607	0.05158	0.20	0.22513
Cycle	-1.86621	0.05785	0.15	0.22513
WEATHER CONDITION				
Clear/Day	0.00000			
Fog/Midst	-0.30015	0.06037	0.74	0.03427
Rain	-0.42441	0.08407	0.65	0.00155
Other	1.02406	0.03835	2.78	0.90316
LIGHT CONDITION				
Day	0.00000			
Night no street light	1.30287	0.11882	3.68	0.02316
Night street light off	0.61502	0.21874	1.84	0.00442
Night street light on	-1.81709	0.04952	0.16	0.68584

The most important feature of the table 1 above is that, the number of persons killed as a result of road accidents in Ghana seems to be increasing as the years go by. In the year 2009 from the table, there were 2,237 people who died in road accidents, this was fortunately reduced to 1,986 in 2010 and in the year 2011 the figure rose to 2,199 people dying as a result of road accidents. By the year 2013, the number of people killed by road accidents had risen to 2,249. This statistics for the five years period therefore supports what Odoom (2010) stated in his research work on road accidents that, as the years go by, the number of motor vehicles in Ghana will be increasing with increasing number of road accidents and as such the number of persons who are likely to be killed in road accidents will also increase.

#### 4.1. Poisson Regression Model for the Number of Persons Killed in Road Accidents from 2009-2013

The table 2 below presents the results on the estimate of the Poisson count regression model for the association of collision type resulting in death, type of vehicle resulting in death, weather condition of occurrence of accident resulting in death, light condition prevailing during accident, and driver error resulting in death. The table contains the parameter estimates, standard errors, death rates and the respective p-values of the various categories of variables used in the model.

The parameter estimate from the table 2 above were interpreted in terms of the rate of the number of persons killed or died in road accidents. These rates reflects the multiplicative effect of the various collision types, driver errors, type of vehicles, weather conditions, and light conditions on the number of persons killed in road accidents in Ghana. The parameter estimates of the intercept of 1.5047 with p-value=0.00000786 with an estimated death rate of  $\exp(1.5047)=4.50$  was statistically significant and different from zero at 0.05 level of significance.

In this model, overturn, inexperience, car, clear and day were used as reference levels for the categorical variables Collision type, Driver error, Type of vehicle, Weather condition and Light condition due to the R statistical software used for the analysis.

With respect to the categorical variable Collision type from table 2 above, rear-end collision, side swipe collision, right angle collision, hit object on road, and hit object off road with p-values 0.09702, 0.20404, 0.85036, 0.29680 and 0.92492 respectively were not significantly associated with the number of persons killed in road accidents in Ghana. However, only two collision types including Ran off collision and Head-on collision with p-values 0.03514 and 0.00158 respectively were significantly associated with the number of persons killed in road accidents. The parameter estimates of these significant collision types that is Ran off collision and Head-on collision were 0.28387 and 0.93386 with death rates of 1.33 and 2.54 respectively.

The death rate value for Ran off collision suggests that, the rate of death in road accidents is 1.33 times higher among

deaths caused by ran-off collision compared to overturn collision whilst that of Head-on collision on the other hand indicates that, the rate of death in road accidents is 2.54 times higher among deaths caused by Head-on collision compared to deaths caused by overturn collision.

Also, inattentiveness on part of drivers, improper overtaking and loss of control with their respective p-values 0.00874, 0.04291 and 0.02658 were the only categories of the variable Driver error which were significantly associated with the number of persons killed by road accidents in the country. The estimated rate of death for inattentiveness from the table above was 0.45 indicating that, the rate of death in road accidents is 0.45 times higher with deaths caused by drivers inattentiveness compared to drivers inexperience, estimated death rate for Improper overtaking was also 1.25 which means that the rate of death in road accidents in Ghana is 1.25 times higher with improper overtaking compared to inexperience and that of loss of control is 0.43 indicating that the rate of death or number of persons killed in road accidents is 0.43 times higher with Loss of control compared to inexperience. The remaining Driver errors including driver driving too fast, too close and Fatigue/asleep were significantly associated with the rate of death in road accidents in Ghana.

In addition, only two types of vehicles were significantly associated with the number of persons killed in road accidents (Goods vehicle and Bus/Minibus) with p-values of 0.00127 and 0.0215 respectively. From the table 2, Goods vehicle as a type of vehicle contributing significantly to number of persons killed in road accidents had a rate value of 0.57 indicating that the rate of number of persons killed in road accidents was 0.57 times higher with Goods vehicle compared to cars whilst the rate value for Bus/Minibus was obtained as 1.36 meaning the rate of death in road accidents was 1.36 times higher with Bus/Minibus compared cars in Ghana.

Furthermore, other weather conditions was the only category of the variable weather condition which was not significantly associated with the number of persons killed in road accidents since it had a p-value (0.90341) greater than the 5 percent level of significance. Thus, Fog/midst and Rain their respective p-values, 0.03427 and 0.00155 proofed to be significantly associated with the number of persons killed in road accidents in Ghana. Fog/midst and Rain as types of weather conditions from the table 2 above had their rate values as 0.74 and 0.65 respectively indicating that, the rate of death in road accidents in the country is 0.74 times higher with Fog/midst compared to when the weather is clear and rate of death in road accidents is 0.65 times higher with Rain compared to clear weather condition.

Lastly from the table 2 above, the last variable Light condition was classified into four categories of which only two were significantly associated with the number of persons killed in road accidents in Ghana. These categories include "Night no street light" and "Night street light off" with p-values 0.02316 and 0.00442 respectively being less the 5 percent level of significance. The rate of death in road

accidents on the other hand of these categories contributing significantly to number of persons killed were 3.68 and 1.84 respectively. This as a result gives the indication that, the death rate in road accidents in Ghana is 3.68 times higher with accidents that occur in the night with no street lights compared to the Day whilst on the other side the rate of death in road accidents is 1.84 times higher with accidents that occur during the Night with street lights off compared to the accidents that occur in the Day.

Hence the mean of the Poisson count regression model with the significant variables for estimating the mean or the average number of persons killed in road accidents in Ghana is:

$$\theta_i = \exp(1.50417 + 0.28387RO + 0.93386HO - 2.80611IN + 0.22388IO - 0.84303LC - 0.55358GV + 0.30358BM - 0.30015F - 0.42441R + 1.30287NS + 0.61502NO) \quad (23)$$

Where RO is Ran off collision, HO IS Head-on collision, IN is Inattentiveness, IO is Improper overtaking, LC represents Loss of Control, GV is Goods vehicle, BM is Bus/Minibus, F is Fog/midst, R is rain, NS is Night with no street lights and NO represents Nights with street lights off.

#### 4.1.1. Goodness of Fit Test of the Poisson Regression

Table 3 below displays the goodness of fit test of the fitted mean Poisson regression model for the number of persons killed in road accidents in Ghana from 2009-2013. This goodness of fit test helps one to determine how quality fit or appropriate the fitted model is as compared to other models.

**Table 3.** Goodness of Fit test of the Poisson Regression

Criterion	Value	Degree of Freedom
Null deviance	2493.63	65
Residual deviance	470.85	54
AIC	1059.31	
BIC	1070.65	
Dispersion parameter	8.71944	

The above table reveals that the Akaike's Information Criterion (AIC) is 1059.31, Bayesian Information Criterion is also 1070.65 and the Residual deviance with degree of freedom 54 on the other hand is 470.85. This value of the residual deviance divided by its degree of freedom gives 8.71944. The resulting figure is approximately 9 which is greater than 1.

This gives the indication that there is over dispersion. As a result, the fitted model as a whole is not appropriate since the major assumption of Poisson regression which is equidispersion (when the dispersion parameter=1) is violated hence the Negative Binomial and the Conway-Maxwell Poisson is preferred.

## 4.2. Negative Binomial Regression Model for the Number of Persons Killed in Road Accidents from 2009-2013

The table 4 below shows the results of the results of the Negative Binomial regression model with collision type, driver error, and type of vehicle, light condition and weather condition as the independent categorical variables and the number of persons killed as the response variable. The parameter estimates associated with the categorical independent variables were the same as those of the Poisson regression model. As a result the estimated rate of the number of persons killed in road accidents associated with collision type, driver error, type of vehicle, light condition and weather condition remained unaltered. However, the standard errors and the respective probability values p-values changed from one variable to the other.

From the table 4 above, the intercept which depicts the effects of the other variables which were not considered in the model was significant with a parameter estimate 1.50417, a death rate of 4.50 and a p-value of 0.00153. With respect to the variable "collision type", head on collision was the only type of collision that was significantly associated with the number of persons killed in road accidents. The parameter estimate for this collision type was 0.92703 with an estimated death rate of  $\exp(0.92703)=2.53$ . This figure as a result suggest that, the rate of death in road accidents is 2.53 times higher among accidents caused as a result of head-on collision compared to overturn collision.

However, Improper-overtaking and loss of control under the categorical variable "Driver error" with respective p-values 0.02638 and 0.00303 remained significantly associated with the number of persons killed in road accidents in Ghana at 5 percent level of significance with death rate of 1.25 and 0.43 respectively. This therefore implies that, the rate of death in road accidents is 1.25 times higher among accidents caused by improper overtaking compared to inexperience nature of some drivers and also the rate of death in road accidents is 0.43 times higher among accidents caused by a drivers losing control compared to inexperience. On the other hand, a driver driving too fast and No signal" from the table 4 had a high effect on the rate of death in road accidents with rate values of 1.31 and 1.08 respectively but were not significantly associated with the number of persons killed at the 5 percent level of significance.

Also among the various classifications of the type of vehicle as a categorical independent variable, only Bus/Minibus was significantly associated with the number of persons killed in road accidents with a p-value of 0.02107 and a death rate value of 1.36. The death rate value of the Bus/Minibus as a result gives an indication that in Ghana, the rate of death when it comes to road accidents is 1.36 higher among accidents that involves Bus/Minibus as compared to accidents caused by cars.



**Table 4.** Parameter estimates of Negative Binomial Regression Model for the Number of Persons Killed in Road Accidents from 2009-2013

Variable	Parameter Estimates	Standard Error	Rates $Exp(\beta)$	P-value
Intercept	1.50417	0.33361	4.50	0.00153
COLLISION TYPE				
Overtake	0.00000			
Rear-end	-0.98795	0.57719	0.37	0.75397
Ran-off	0.28387	0.20232	1.33	0.64892
Side swipe	-1.60780	0.03504	0.20	0.27370
Right angle	-1.76591	0.22843	0.17	0.09011
Head on	0.93386	0.37271	2.54	0.03095
Hit object on road	0.92703	0.06450	2.53	0.23864
Hit object off road	-2.21040	0.42633	0.11	0.12839
DRIVER ERROR				
Inexperience	0.00000			
Inattentive	-2.80611	0.57719	0.45	0.24798
Too fast	0.26966	0.56526	1.31	0.17305
Too close	-3.35818	0.62470	0.03	0.07204
No signal	0.07581	0.59167	1.08	0.29318
Improper overtaking	0.22388	0.56653	1.25	0.02638
Fatigue/Asleep	-1.87180	0.58143	0.15	0.43640
Loss of control	-0.84303	0.89334	0.43	0.00303
TYPE OF VEHICLE				
Car	0.00000			
Goods vehicle	-0.55358	0.25324	0.57	0.07204
Bus/Minibus	0.30358	0.34636	1.36	0.02107
Motor cycle	-1.29629	0.13662	0.27	0.29904
Pickup	-1.59607	0.21640	0.20	0.17673
Cycle	-1.86621	0.27435	0.15	0.08471
WEATHER CONDITION				
Clear/Day	0.00000			
Fog/Midst	-0.30015	0.35381	0.74	0.01425
Rain	-0.42441	0.25853	0.65	0.28574
Other	1.02406	0.32640	2.78	0.46261
LIGHT CONDITION				
Day	0.00000			
Night no street light	1.30287	0.27094	3.68	0.28091
Night street light off	0.61502	0.34691	1.84	0.00215
Night street light on	-1.81709	0.34095	0.16	0.21509

Not forgetting the categorical variable "light condition", Night with street light off remained statistically significant with a probability of 0.00125 and an estimated death rate of 1.84. This rate value therefore indicates that, the rate of death in road accidents is 1.84 times higher with accidents that occur on streets with lights off during the night compared to accidents that occur during the day.

The mean regression of the Negative Binomial model for estimating the expected number of persons killed in road accidents within a specific year with significant variables is therefore formulated as;

$$\theta_i = \exp(1.50417 + 0.93386 HO + 0.22388 IO - 0.84303 LC + 0.30358 BM - 0.30015 F + 0.61502 NO) \quad (24)$$

where HO is Head-on collision, IO is Improper overtaking, LC represents Loss of control, BM is Bus/Minibus, F is Fog/Midst, NO represents Nights with street lights off.

#### 4.2.1. Goodness of Fit Test of the Negative Binomial Regression

Table 5 below depicts the goodness of fit test of the fitted Negative Binomial regression model for the number of

persons killed as a result of road accidents in Ghana from 2009-2013. This table gives the details of null deviance, residual deviance, AIC, BIC and the dispersion parameter of the model.

From the table, the residual deviance of the model decreased substantially that is 65.666, which is much smaller compared to 470.85 of the Poisson model. The corresponding value by dividing the residual deviance by its degree of freedom is 1.21603 (dispersion parameter) which is approximately equal to 1, therefore suggesting the elimination of over dispersion. These results suggest that the

Negative binomial model is reasonable or appropriate.

**Table 5.** Goodness of Fit test of the Negative Binomial Regression

Criterion	Value	Degree of Freedom
Null deviance	985.72	65
Residual deviance	65.66	54
AIC	749.10	
BIC	758.46	
Dispersion parameter	1.21603	

**Table 6.** Parameter Estimates of Conway-Maxwell-Poisson model for the number of persons killed in Road Accidents from 2009-2013

Variable	Parameter Estimates	Standard Error	Rates $Exp(\beta)$	P-value
Intercept	1.50417	0.03026	4.50	0.0000542
COLLISION TYPE				
Overtake	0.00000			
Rear-end	-0.98795	0.35472	0.37	0.321272
Ran-off	0.28387	0.48344	1.33	0.251744
Side swipe	-1.60780	0.41280	0.20	0.344902
Right angle	-1.76591	0.28196	0.17	0.166218
Head on	0.93386	0.82213	2.54	0.001123
Hit object on road	0.92703	0.58774	2.53	0.454128
Hit object off road	-2.21040	0.54452	0.11	0.344902
DRIVER ERROR				
Inexperience	0.00000			
Inattentive	-2.80611	0.37596	0.45	0.036919
Too fast	0.26966	0.37975	1.31	0.478366
Too close	-3.35818	0.77529	0.03	0.559538
No signal	0.07581	0.39298	1.08	0.240611
Improper overtaking	0.22388	0.51608	1.25	0.021518
Fatigue/Asleep	-1.87180	0.32783	0.15	0.848205
Loss of control	-0.84303	0.54376	0.43	0.111903
TYPE OF VEHICLE				
Car	0.00000			
Goods vehicle	-0.55358	0.18784	0.57	0.301883
Bus/Minibus	0.30358	0.16035	1.36	0.007719
Motor cycle	-1.29629	0.24488	0.27	0.461480
Pickup	-1.59607	0.27646	0.20	0.089356
Cycle	-1.86621	0.31005	0.15	0.566059
WEATHER CONDITION				
Clear/Day	0.00000			
Fog/Midst	-0.30015	0.28615	0.74	0.872703
Rain	-0.42441	0.39853	0.65	0.021015
Other	1.02406	0.18178	2.78	0.112926
LIGHT CONDITION				
Day	0.00000			
Night no street light	1.30287	0.68385	3.68	0.069205
Night street light off	0.61502	0.15302	1.84	0.039922
Night street light on	-1.81709	0.23322	0.16	0.591063

### 4.3. Conway-Maxwell-Poisson Regression Model for the Number of Persons Killed in Road Accidents from 2009-2013

The parameter estimates of the Conway-Maxwell-Poisson model from the table 5 below on the other hand as compared to the Negative Binomial model and Poisson regression model were the same. This as a result give the indication that the estimated rate of death values for the respective variables used in the analysis remains the same as well. However the standard errors and the probability values of the variables involved were changed.

The table 6 above depicts that the parameter estimates of the intercept of 1.50417 with a probability value 0.0000542 and an estimated rate of 4.50 was statistically significant. In this model, only head-on collision was significantly associated with the number of persons killed in road accidents under the variable "collision type" with a p-value of 0.001123 and an estimated death rate of 2.53. This estimated rate value of the head-on collision indicates that, the rate of death in road accidents is 2.53 times higher with accidents as result of head-on collision compared to overturn collision.

Also, inattentive and improper overtaking with their respective p-values 0.036919 and 0.021518 were the driver errors in this model which were significantly associated with the number of persons killed by road accidents. The estimated death rate for inattentiveness from the table 6 is 0.45 which suggests that the rate of death in road accidents is 0.45 times higher among accidents caused by drivers inattentiveness compared to drivers' inexperience. The estimated rate of death for improper overtaking on the other hand is 1.25 which as result indicates that, the rate of death in road accidents in this country is 1.25 times higher with accidents caused as a result of improper overtaking compared to inexperience of some drivers.

In addition, from the table 6, only Bus/Minibus remained statistically significant with respect to the number of persons killed in road accidents with p-value of 0.007719 and an estimated death rate of 1.36 indicating that, the rate of death in road accidents in our country is 1.36 times higher with Bus/Minibus compared to cars (taxi, trotro etc.).

Furthermore, under the variable "Weather condition" fog/midst and other weather conditions with p-values 0.872703 and 0.112926 respectively were not significantly associated with the number of persons killed in road accidents. However rain in this case with a p-value of 0.021015 was significantly related to the number of persons killed in road accidents at 5 percent level of significance. The parameter estimate of this variable was -0.42441 and an estimated death rate of 0.65 suggesting that, that rate of death among accidents on the road in Ghana is 0.65 times higher with deaths that resulted from accidents that occurred during the rainy weather condition compared to clear weather condition.

Lastly in this model, the categorical independent variable "Light condition" had only Night with street lights

off having a significant relation with the number of persons killed as a result of road accident. Night with street light off from table 6 had a p-value of 0.039922 and death rate of 1.84 which implies that, the rate of death in road accidents is 1.84 times higher when it comes to the accidents that occur during the night with street lights off compared to accidents that occur during the day.

Therefore, the mean of the Conway-Maxwell-Poisson count regression model for estimating the mean or the average number of persons killed in road accidents yearly for the various significant variables is stated as:

$$\theta_i = \exp(1.50417 + 0.93386HO - 2.80611IN + 0.22388IO + 0.30358BM - 0.42441R + 0.61502NO) \quad (25)$$

where HO is Head collision, IN is Inattentiveness, IO is Improper Overtaking, BM is Bus/Minibus, R is Rain and NO is Night with street lights off.

#### 4.3.1. Goodness of Fit test of the Conway-Maxwell-Poisson Model

The table 7 below indicates the goodness of fit test of the Conway-Maxwell-Poisson regression model for the number of persons killed in road accidents from 2009-2013. This table helps to determine as to whether the fitted model is appropriate or whether the model is quality fit.

**Table 7.** Goodness of Fit test of the Conway-Maxwell-Poisson Model

Criterion	Value	Degree of Freedom
Null deviance	1159.04	65
Residual deviance	76.28	54
AIC	826.46	
BIC	835.79	
Dispersion parameter	1.41249	

The table 7 above reveals that, in this model, the residual deviance, AIC and BIC all have decreased substantially with their respective values 76.28, 826.46 and 835.79, which are much smaller compare to residual deviance of 470.85, AIC of 1059.31 and BIC OF 1070,65 of the Poisson regression model. The corresponding value by dividing the residual deviance by its degrees of freedom is approximately 1, therefore suggesting that the CMP regression model compared to the Poisson model can take care of the presence of over-dispersion.

This results indicates that the CMP model is also an appropriate model.

### 4.4. Model Evaluation and Comparison

With respect to the table 8 presented above, there is a clear evidence that Negative binomial model performs best and is the best model which best fit the accident data well with respect to the expected number of persons killed in road accidents as compared to the Poisson and the CMP model.

This is because in order to select the best model that performs best or best fits with respect to a certain data among other models with the help AIC, BIC or deviance, the

criterion set is that, the smaller the value of the AIC or BIC or Deviance the better that model becomes. Hence by comparing the respective values of the AIC, BIC as well as the residual deviance of the Poisson, Negative Binomial and the CMP model from the table 8, it is seen that the Negative binomial has the smallest AIC value of 749.10, BIC value of 758.46 and a residual deviance value of 65.666 as compared those of Poisson and the CMP model.

**Table 8.** Results of the model evaluation and comparison

Characteristic	Poisson model	NB model	CMP model
Null deviance	2493.63	985.725	1159.04
Df	65	65	65
Residual deviance	470.85	65.66	76.28
Df	54	54	54
Dispersion parameter	8.71944	1.21603	1.41249
AIC	1059.31	749.10	826.46
BIC	1070.65	758.46	835.79

Although both the Negative Binomial and the CMP model takes care of over dispersion, the dispersion parameter value from the table 8 indicates that the Negative binomial regression takes care of over dispersion more than that of the CMP model. The above statistics therefore gives an indication that, the model that best fits accident data with number of persons killed in the Ghana is the Negative Binomial model followed by the CMP model.

Hence in order for one to estimate the expected number of persons killed in road accidents in Ghana per year, the mean of the Negative binomial model with the significant variables head-on collision as a collision type, Improper overtaking and loss of control as driver errors, Bus/Minibus as type of vehicle, Fog/midst as weather condition and Night with street lights off as Light condition must be used and is given as:

$$\theta_i = \exp(1.50417 + 0.93386 HO + 0.22388 IO - 0.84303 LC + 0.30358 BM - 0.30015 F + 0.61502 NO)$$

where  $\theta_i$  represents the expected number of persons killed in road accidents in a particular period of time, HO is Head-on collision, IO is Improper overtaking, LC represents Loss of control, BM is Bus/Minibus, F is Fog/Midst, NO represents Nights with street lights off.

## 5. Conclusions

This research work was aimed at examining the efficiency of different statistical models for count data with application to road accidents in Ghana. As a result three (3) statistical models including Poisson, Negative Binomial and Conway-Maxwell-Poisson count regression models were fitted. All the fitted models include significant explanatory variables.

Base on the deviances, AIC and BIC of the respective

fitted models it appeared that only Negative Binomial model performed best as compared to Poisson and the Conway-Maxwell-Poisson model. The predictors in this model were investigated using their respective p-values and was found out that Head-on collision as collision type, Improper overtaking and Loss of control as driver errors, Bus/Minibus as type of vehicle, Fog/midst as weather condition and Night with street lights off as Light condition were the key predictors or independent variables contributing significantly to the expected number of persons to be killed in road accidents in Ghana.

The empirical study of this research work additionally revealed that in the presence of over-dispersion, both the Negative Binomial and Conway-Maxwell-Poisson count regression models are potential alternative to the Poisson count regression model due to its major assumption of equi-dispersion.

Thus the Poisson count regression model serves well under equi-dispersion condition whiles both Negative binomial and CMP count regression models serve better whiles the data is over-dispersed.

## Appendix

### Derivation of the Poisson Regression

From the generalized linear model framework, the exponential family in canonical form is given by the probability function;

$$f(y, \theta, \varphi) = \exp\left(\frac{y \cdot \theta - b(\theta)}{\varphi} + c(y, \varphi)\right) \quad (1)$$

But the Poisson distribution function on the other hand is given by;

$$P(y_i / x_i) = \frac{\exp(-\theta_i) \theta_i^{y_i}}{y_i!} \quad \text{for } y_i = 0, 1, 2, 3, \dots$$

By taking the exponent and natural log (ln) of the Poisson pdf, we obtain the following;

$$\begin{aligned} &= e^{-\theta} \exp\left(\frac{\theta_i^{y_i}}{y_i!}\right) \\ &= e^{-\theta} \exp\left[\ln\left(\frac{\theta_i^{y_i}}{y_i!}\right)\right] \\ &= e^{-\theta} \cdot \exp(y_i \ln \theta_i - \ln y_i!) \\ &= \exp(y_i \ln \theta_i - \theta_i - \ln y_i!) \end{aligned} \quad (2)$$

By comparing equation 9 to equation 8, we have;  $\theta = \ln \theta_i$ ,  $\varphi = 1$ ,  $c(y, \varphi) = -\ln y_i!$ .

But

$$\theta = E(g(y)) = x_i' \beta \quad (3)$$

Therefore

$$\ln \theta_i = x_i' \beta \quad (4)$$

Hence by making  $\theta_i$  the subject from (11) we obtain;

$$\theta_i = \exp(x_i' \beta) \quad (5)$$

as the Poisson mean regression model

#### Derivation of NB Model

Suppose that we have a series of random counts that follows a Poisson distribution;

$$P(Y_i = y_i) = \frac{\exp(-\theta_i) \theta_i^{y_i}}{y_i!}$$

Where  $y_i$  is the observed number of counts (accidents) for  $i = 0, 1, 2, 3, \dots, n$  and  $\theta_i$  is the mean or the expected number of counts (persons killed in road accidents) of the Poisson distribution.

In the Poisson distribution, the  $\theta_i$  was fully determined by a linear combination of  $x_s$ . The Negative Binomial on the other hand adds some unobserved heterogeneity, such that  $\theta_i$  is determined by the  $x_s$  and some unobserved specific random effect called Gamma distributed error ( $\varepsilon_i$ ) which is assumed to be uncorrelated with the  $x_s$ .

Hence we obtain the following relationships;

$$\begin{aligned} \theta_i &= e^{x_i \beta + \varepsilon_i} \\ \theta_i &= e^{x_i \beta} e^{\varepsilon_i} \\ \therefore \bar{\theta} &= \theta_i v_i \end{aligned} \quad (6)$$

where  $v_i = e^{\varepsilon_i}$  is defined as the unobserved random effect;  $\theta_i = e^{x_i \beta}$  is the log-link between the Poisson mean and the covariance or independent variables  $x_s$ ; and the  $\beta_s$  are the regression coefficients.

The Negative Binomial is not defined without the assumption about the mean error term and the most convenient assumption is that,  $E(v_i) = 1$  since this gives  $E(\theta_i) = \theta_i$ . This is as a result indicates that we have the same expected counts as the Poisson distribution.

Since the distribution of the observations is still Poisson with given  $x$  and  $\delta$ , we obtain the relation:

$$P(y_i / x_i, v_i) = \frac{e^{-\theta_i} \theta_i^{y_i}}{y_i!} = \frac{e^{-\theta_i v_i} (\theta_i v_i)^{y_i}}{y_i!} \quad (7)$$

where  $v_i$  is unknown.

Since  $v_i$  is unknown,  $P(y_i / x_i, v_i)$  cannot be computed and instead there is a need to compute the distribution of  $y$  given  $x$  only. Hence, to compute

$P(y_i / x_i)$  without conditioning  $v_i$ , we compute the average of  $P(y_i / x_i)$  by the probability of each value of  $v_i$ .

Hence if  $g$  is the probability density function of  $v_i$  then;

$$P(y_i / x_i) = \int_0^{\infty} [P(y_i / x_i, v_i) \times g(v_i)] dv_i \quad (8)$$

where  $g(v_i)$  is a mixing distribution. In the case of the case of the Poisson-Gamma mixture,  $P(y_i / x_i, v_i)$  is the Poisson distribution and  $g(v_i)$  is the Gamma distribution. This distribution has a closed form and as a result leads to the Negative Binomial distribution.

In order to solve (17), there is the need to specify the form of the pdf for  $v_i$ . Assuming that the variable  $v_i$  follows a two-parameter Gamma distribution then we have;

$$g(v_i; \psi, \delta) = \frac{\delta^\psi}{\Gamma(\psi)} v_i^{\psi-1} e^{-\delta v_i} \quad \text{for } \psi > 0, \delta > 0, v_i > 0 \quad (9)$$

$$\text{where } \Gamma(\psi) = \int_0^{\infty} t^{\psi-1} e^{-t} dt$$

An interesting thing about this distribution is that,  $E(v_i) = 1$  which is a convenient assumption indicating that the mean of the Negative Binomial is the same as the Poisson.

$$\text{Also } \text{Var}(v_i) = \frac{1}{\psi}.$$

By using (7) and (9) to solve (8), we obtain following Negative Binomial model;

$$P(y_i / x_i, v_i) = \frac{\Gamma(y_i + v_i)}{y_i! \Gamma(v_i)} \left[ \frac{v_i}{v_i + \theta_i} \right]^{v_i} \left[ \frac{\theta_i}{v_i + \theta_i} \right]^{y_i} \quad (10)$$

The expected value of the above relation (NB model) is the same as the Poisson distribution that is  $E(y_i / x_i) = \theta_i$ . But the conditional variance differs from that of the Poisson distribution;

$$\text{Var}(y_i / x_i) = \theta_i \left[ 1 + \frac{\theta_i}{v_i} \right] = e^{x_i \beta} \left[ 1 + \frac{e^{x_i \beta}}{v_i} \right] \quad (11)$$

Because  $\theta_i$  and  $v_i$  are both positive from (11), the variance automatically exceeds the conditional mean. This as a result increases the relative frequency of low and high counts. There will be more parameters than observations since the variances remain undefined if  $v_i$  varies by observation.

The typical assumption is that  $v_i$  is the same for the observations (counts) that is;

$$v_i = \alpha^{-1} \quad \text{for } \alpha > 0 \quad (12)$$

This formulation as a result helps to simplify some equations if we substitute (21) into (19). Hence we obtain the

NB model as;

$$P(y_i / x_i, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{y_i! \Gamma(\alpha^{-1})} \left[ \frac{\alpha^{-1}}{\alpha^{-1} + \theta_i} \right]^{\alpha^{-1}} \left[ \frac{\theta_i}{\alpha^{-1} + \theta_i} \right]^{y_i} \quad (13)$$

and the conditional variance as;

$$Var(y_i / x_i) = \theta_i \left[ 1 + \frac{\theta_i}{\alpha^{-1}} \right] = \theta_i (1 + \alpha \theta_i) \quad \text{for } \alpha > 0 \quad (14)$$

## REFERENCES

- [1] Abdel-Aty, M.A., Radwan, A.E. (2000), Modelling traffic accidents occurrence and involvement. *Accident Analysis and Prevention*, Vol. 32, pp. 633-642. [http://dx.doi.org/10.1016-4575\(99\)00094-9](http://dx.doi.org/10.1016-4575(99)00094-9).
- [2] Abdel-Aty and Abdalla (2004), Modelling drivers diversion from normal routes under ATIS using generalized estimating equations and binomial probit link function. <http://dx.doi.org/10.1023/b:port.0000025396.32909.dc>.
- [3] Adams JGV.....and how much for your grandmother? *Environment and Planning*, 1974, (6), pp 619-626. <http://dx.doi.org/10.1068/a060619>.
- [4] Afukaar F.K., Agyemang W, and Most I. (2009). Accident statistics 2007. Building and Road Research Institution, Council for Scientific and Industrial Research, Kumasi.
- [5] Afukaar F.K., and Debrah E.K., (2007). Accident prediction for signalized intersections in Ghana. Ministry of Transport: Nation Accident Management Project, report, Final report.
- [6] Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In the second international symposium on information theory, edited B.V Petrov and B.F Csaki, Academic Kiado.
- [7] Andreassen D. (1985). Linking deaths with vehicles and population. *Traffic Engineering and Control*. 26:547-549.
- [8] Caliendo, C., Guida, M., Parisi, A. (2007). A crash-prediction model for multilane roads. <http://dx.doi.org/10.1016/j.aap.2006.10012>.
- [9] Chin, H.C. and Quddus, M.A. (2003). Modelling count data with excess zeros: An empirical application to traffic accidents. *Sociological Methods and Research*, 32, 90-116. <http://dx.doi.org/10.1177/0049124103253459>.
- [10] El-Basyouny, K., Sayed, T. (2006). Comparison of two Negative Binomial Regression Techniques in Developing Accident Prediction Models. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1950, Transport Research Board of the National Academies, Washington, DC, 2006, pp. 9-16. <http://dx.doi.org/10.3141/1950-02>.
- [11] Fletcher J.P., Baguley C.J., Sexton B., and Done S. (2006). Road accident modeling for highway development countries. Main Report Trials in India and Tanzania. Report No. PPRO95.
- [12] Jamal R.M.A., and Jamil A.N. (2001). Casual models for road accidents in Yemen. *Accident Analysis and Prevention*. 33:547-561. [http://dx.doi.org/10.1016/s0001-4575\(00\)00069-5](http://dx.doi.org/10.1016/s0001-4575(00)00069-5).
- [13] Jovanis, P.P., and Chang, H. (1986). Modelling the relationship of accidents to miles travelled.
- [14] Lee, J., and Mannering, F. (2002). Impact of roadside features on the frequency and severity of run-off-roadway accidents: An empirical analysis. *Accidents Analysis and Prevention*, 34, 149-161. [http://dx.doi.or/10.1016/s001-4575\(01\)00009-4](http://dx.doi.or/10.1016/s001-4575(01)00009-4).
- [15] Lord, D., Persaud, B.N. (2000). Accident prediction models with and without trend: application of the generalized estimating equations procedure. *Transportation Research*. <http://dx.doi.org/10.1016/j.aap.2004.02.004>.
- [16] Lord, D., Washing ton, S.P., and Ivan, J.N. (2005). Poisson, Poisson-gamma, and Zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis and Prevention*, 37(1), 35-46.
- [17] Miaou, S.P., and Lum, H. (1993). Modelling vehicle accidents and highway geometric design relationships. *Accident Analysis and Prevention*. 25:689-1993. [http://dx.doi.org/10.1016/0001-4575\(93\)90034-t](http://dx.doi.org/10.1016/0001-4575(93)90034-t).
- [18] Minter A.L., Road casualties improvement by learning processes. *Traffic Engineering and Control*, 28:74-79.
- [19] Montella (2008). Rates ration relationships for rural and urban freeway segments. *Accidents Analysis and Record*, Vol.1717, pp. 102-108
- [20] National Road Safety Commission (2009). Annual Report. National Road Safety Commission, Ghana
- [21] Oppong Richard Asumadu, (2012), Statistical Analysis of Road Accidents Fatalities in Ghana using Poisson regression, Mphil thesis, Kwame Nkrumah University of Science and Technology.
- [22] Pemmanaboina, R (2005). Assessing accident occurrence on urban freeways using static and dynamic factors b applying a system of interrelated equations, M.E. thesis, University of Central Florida.
- [23] Persaud, B., Dzibik, L., (1993) Accident prediction models for freeways. *Transportation Research Record*, Vol.1401, pp. 55-60.
- [24] Poch, M., and Mannering, F.L. (1996). Negative Binomial analysis of intersection accident frequency. *Journal of Transportation Engineering*, 122, 105-133. [http://dx.doi.org/10.1061/\(asce\)0733-947x\(1996\)122:29105](http://dx.doi.org/10.1061/(asce)0733-947x(1996)122:29105).
- [25] Pramada, V.P., and Sarkar, P.K. (1997). Variation in the pattern of road accidents in different states and union territories in India. *Proceedings of the third national conference on transportation systems studies: Analysis and Policy*. 1X-5 to 1X-9. *Prevention*, Vol. 37, pp. 185-199.
- [26] Sagberg, F., and Saetermo (1997). An investigation of behavioural adaptation to airbags and antilog brakes among taxi drivers (29 ed). *Accident Analysis and Prevention*. 293-302. [http://dx.doi.org/10.1016/s0001-4575\(96\)00083-8](http://dx.doi.org/10.1016/s0001-4575(96)00083-8).
- [27] Shankar, V., Mannering, F., and Barfield, W., (1995). Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. *Accidents Analysis and Prevention*, 27,371-389.

- [28] Shankar, V., Milton, J., and Mannering, F. (1997). Modeling accident frequencies as zero-altered probability processes: An empirical inquiry. *Accidents Analysis and Prevention*, 29, 829-837. [http://dx.doi.org/10.1016/s0001-4575\(97\)00052-3](http://dx.doi.org/10.1016/s0001-4575(97)00052-3).
- [29] Shmueli, G., Minka, T.P., Kadane, J.B., Borle, S., and Boatwright, P. (2005). A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution, *Applied Statistics*, 54, 127-142. <http://dx.doi.org/10.1111/j.1467-9876.2005.00474.x>.
- [30] Susan, C. and Partyka (1984). Simple model of fatality trends using employment and population data. *Accident Analysis and Prevention*. 16:21-222. [http://dx.doi.org/10.1016/0001-4575\(84\)90015-0](http://dx.doi.org/10.1016/0001-4575(84)90015-0).