KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY, KUMASI

STATISTICAL MODELS FOR COUNT DATA WITH APPLICATIONS TO ROAD ACCIDENTS IN GHANA

BY

ADJEI MENSAH ISAAC

A THESIS SUBMITTED TO THE DEPARTMENT OF MATHEMATICS, KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY IN PARTIAL FUFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF M.PHIL MATHEMATICAL STATISTICS

JUNE, 2015

DECLARATION

I hereby declare that this submission is my own work towards the award of the M. Phil degree and that, to the best of my knowledge, it contains no material previously published by another person nor material which had been accepted for the award of any other degree of the university, except where due acknowledgment had been made in the text.

Adjei Mensah Isaac ....

Student Signature Date

Certified by:

Dr.A.Y.Omari-Sasu ....

Supervisor Signature Date

Certified by:

Prof. S. K. Amponsah ....

Head of Department Signature Date

ABSTRACT

Road accidents in Ghana seems to be on ascendency and the root causes of these accidents have been attributed to issues such as human errors and superstitions. Since the occurrence of accidents are discrete, they are often modeled using count regression models. It is therefore the purpose of this study to determine an appropriate count regression model that adequately fits road accidents on urban roads in Ghana and to determine the key predictors of road accidents using the appropriate count model with respect to the expected number of person killed in an accident. Several models fitted using count data (occurrences of road accidents) in the field of transportation were compared.These models includePoisson,NegativeBinomialandConway-Maxwell-Poissoncountregression models . To compare the performance of these models, the various model selectionmethodssuchasDeviancegoodnessoffit, Akaike'sInformationCriterion (AIC)and Bayesian Information Criterion (BIC) were employed. Because the values of the Deviance goodness of fit, AIC and BIC respectively of the Negative Binomial was the smallest as compared to that of Conway-Maxwell-Poisson and Poisson models, it appeared that the Negative Binomial model performed best as compared to the Poisson and the CMP model. Base on the appropriate count regression model selected (Negative Binomial model) the key predictors that contributed significantly and had a high effect on the expected number of persons to be killed in a road accidents within a particular time were Head-on collision as Collision type, Improper-overtaking and Loss of control as Driver errors, Bus/minibus as Type of vehicle, Fig/midst as Weather condition and Night with street lights off as Light condition.

DEDICATION

I wholeheartedly dedicate this research work to my parents Mr. John Sarfo and Madam Comfort Adjei

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to the Almighty God for seeing me through this research work.

My sincere thanks goes to my supervisor Dr.A.Y.Omari-Sasu for supervising this thesisandofferingmedirections, suggestionsandencouragementduringtheentire period of the research work.

I owe a debt of gratitude to Dr.R.K.Boadi for his assistance during the data collection at Building and Road Research Institute (BRRI) of Council for Scientific and Industrial Research (CSIR).

I dully acknowledge here, the contribution of my sister Eunice Nimakoh Baah and my brother Frank Mintah.

CONTENTS

LIST OF ABBREVIATIONS/ACRONYMS

BRRI ....Building and Road Rsearch Institute

CSIR ....Council for Scientific and Industrial Research

NGO ....Non Governental Organization

NRSC ....National Road Safety Commission

ECE ....Economic Commission of Europe

ZIP ....Zero-Inflated Poisson model

ZINB ....Zero-Inflated Negative Binomial model

GLI M ....Genralized Linear Ineractive Model

GLM .... Gneralized Linear Model

CMP ....Conway-Maxwell-Poisson Model

TRRL ....Transport and Road Research Laboratory

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Itiswellknownthat, accidentsareofhighrelevancetothepublichealthspectrum in the world. Moreover, in a developing country like Ghana, the mortality rate from accidents are high as compared to other countries in this region and even classified as the second major cause of death in the country following malaria, Oppong (2012). Not only the majority of the people killed and seriously injured significantly affect the quality of citizens but it also inhibits the economic and social development of this country.

According to a research by the Building and Road Research Institute (BRRI) of Ghana in 2009, there were 12,299 road accidents. There were a total of 18,496 casualties with 2,237 of them losing their lives, whilst 6,242 sustained serious injuries. This therefore gives an indication that there was an average of six (6) deaths every day in Ghana as a result of road accidents. The most dangerous part of it all is that, most of the people who are killed by road accidents are those in the age group that constitute the labor force of Ghana.

Itisthereforeinthisregardthatmoreattentionneedstobeplacedontheresearch into road accidents and its impact on the lives of the citizens and properties in Ghana.

1.1 Background of the study

Road accidents in Ghana have been a serious concern to most Ghanaians in recent times. As a result, of the tremendous effect of accidents on human lives, properties and the environment as a whole. A lot researchers have come out with the causes, effects and the recommendations to road accidents. These causes include drunk driving, over speeding and machine failure Sagberg and Saetermo(1997) and National Road Safety Commission (2009). Despite all these being done, every year the Ghana statistical service, Road Safety Commission and other organizations would report that there is an increase in road accidents in the country; Annual Report, National Road Safety Commission, Ghana (2009). Researchers in recent times have been modeling road accidents with crash prevention models in various parts of the world.

However, it is extremely tedious to just apply model which have worked in other areas 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 in road accidents in Ghana and this might have been a result of the inadequate information available on road accidents and it impact on human lives and the environment as a whole in the country.

Salifu (2004) has developed a forecasting model for traffic crashes for urban junction with no road signs, Afukaar and Debrah (2007) have also model traffic crashes for signalized urban junction in Ghana and Ackaah (2011) has modeled traffic crashes on rural highways in the Ashanti region. It is however surprising that in spite of the numerous factors identified by researches as the causes of road accidents in Ghana and its consequences on human lives and properties; nobody has modeled the causes of road accidents and its contributions to the death and survival of casualties in Ghana to authenticate the contributions of each of these factors to the casualties. death. Furthermore, there has not been any work on the likelihood of a casualty surviving in road accident or the possibility of a casualty surviving in a particular type of vehicle getting involved in a road accident. Road accident is defined as any activity which distracts the normal trajectory of a moving vehicle(s), in a manner that causes instability in the free flow of the vehicle. Vehicular accident has always been attributed to human errors such as high alcoholic content in the blood stream of the driver, over speeding, wrong overtaken among others. It has also been linked to poor road network, poor surfacing of the roads, witchcraft and the death-dying nature of some of the vehicles which ply the roads. There are numerous suggested solutions, various interventions by government, nongovernmental organizations and other road stakeholders to curtail road accident and its replica effects on human lives and properties, it could be possible that these factors such as type and nature of the road, age and sex of casualty contribute casualty survival in road accidents in the country and have still not been considered.

It is therefore in light of this that this research work is conducted in order to determine the appropriate count model that fits accidents data with respect the expected number of persons who will be killed via road accidents and additionally usetheappropriatecountmodeltoinvestigatethekeypredictorsofaccidentsthat results in the death of people.

1.2 Problem Statement

Road accidents in Ghana seems to be on ascendency and these accidents have been categorized by the National Road Safety Commission as fatal, seriuos and minor. This classification is based on the extent of damage to human lives and properties in the country.

Therootcausesoftheseaccidentsasaresulthavebeenattributedtohumanerrors and superstition. It is therefore the purpose of this research work to perform a comparative studies in order to determine an appropriate count model that adequately fits road accident data in Ghana and to determine the key predictors of road accidents using the appropriate model

1.3 Objectives of the Research

This research work focuses on examining the efficiency of different statistical models for count data with application to road accidents in Ghana.

1.3.1 Specific Objectives

This research work seeks to achieve the following specific objectives;

i. To determine the appropriate count regression model that adequately fits accident data in Ghana

ii. To investigate the key predictors of road accidents in Ghana using an appropriate count data model.

1.4 Methodology

This research work made use of categorical data for the analysis. Secondary data was taken from the Building and Road Research Institute (BRRI) of the Council for Scientific and Industrial Research (CSIR) . Conway-Maxwell-Poisson, Poisson and Negative Binomial regression models will be the main statistical tools for the data analysis. Data analysis in this research work would be made possible with the help of SPSS and R statistical packages.

1.5 Significance of the Research work

It is very important to research into accidents in Ghana in order to come out with the reality on the ground so as to help policy makers to deduce strategies to reduce the numerous deaths caused by car accidents to the barest minimum in the country. Also this research work will help to formulate a predictive statistical model for road accidents in Ghana. Lastly, this research work will help to investigate the key predictors of road accidents in Ghana in order to plan for future occurrences.

1.6 Organization of the Thesis

This research work consist of five main chapter with the Abstract which summarizes the whole research work, the Table of content, List of figures, List of abbreviations, Dedications and Acknowledgment being the preludes.

Chapter one of this research work contains the Introduction, Background information, objectives of the research work, significance of the research work and the organization of the research work. Chapter two on the other hand of this research work contains the literature review, chapter three also discusses the various methods used for the research work and the chapter four gives information on the analysis, modeling of the data and the discussions of findings whilst the chapter five which is the last chapter deals with the summary of findings, conclusions and recommendations

The references will also be found after the last chapter that is chapter five.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Thissectionoftheresearchworkdealswiththevariousliteraturesthatarerelated to the topic under discussion in order to reveal significant facts and findings which already have been identified by previous researchers and numerous research works in and around the effects, causes and economic implications of road accidents with respect to causes of death via road accidents. This sections additionally will focus on the existing statistical models for modeling count data example accident frequency data using Poisson, Negative Binomial and Conway-Maxwell-Poisson count regression models and other applications in real life situation.

2.2 State of Road accidents

Road accidents in Ghana has become one of the increasing concerns to most citizens in recent times. This is as a result of the tremendous effect of road accidents on human lives, properties and the environment. Media reports indicates that there is a high road accident in Ghana, when compared with other developing countries. Ghana was ranked as the second highest road accidents prone nation in 2001 among six West African countries with 73 deaths per 1000 accidents., (Akongbota, 2011).

On the 16th day of November, 2011 the Ghanaian times news paper revealed that a total of 1,986 lives were lost in the country through road accidents from January to October 2011 (The Ghanaian Times, 2011).

Globally road accidents claim the lives of about a million of people per year with an estimated 23 to 34 million injures (Jacobs and Cutting, 2000). Accidents or crashes have been a rapidly growing concern in developing countries such as Ghana and they have now reached the point where about 85 percent of all the world’s road deaths occur in these countries,and the trend is of a continuing increase. Accidents on a total national road network cost most countries, whatever their stage of development , on average between 1 and 3 per cent of their individual gross national product (GNP), a substantial figure that developing countries in particular can ill afford to lose.

A study conducted by Salim and Salimah (2005) revealed that road accidents was the ninth major cause of death in low-middle income countries and predicted that road accident was going to the third major cause of death in 2020 if the trend of vehicular accident was to be allowed to continue.

InMarch, 2006, astudyconductedbyHeidyWorleyrevealedroadtrafficaccidents as the leading cause of death by injury and the tenth leading cause of all deaths globally which as a result now make up a surprisingly significant portion of the worldwide burden of ill-health. An estimated 1.2 million people are killed in road crashes each year and as many as 50 million are injured, occupying 30 percent to 70 percent of orthopedic beds in developing countries hospitals. And if present trend continues, road traffic injuries are predicted to be the third -leading contributor of global burden of disease and injury by 2020 (Heidi, 2006). Developing countries bear a l=huge share of burden, accounting for 85 percent of annual deaths and 90 percent of the disability-adjusted life years lost because of road traffic injury. And since road traffic injuries affect mainly males (73 percent of deaths) and those between 15-44 years, this burden is creating enormous economic hardship due to the loss of family breadwinners (Heidi, 2006)

According to Afukaar et al (2009) in their research report presented to the National Road Safety Commission, there was a total of 11320 road accidents which killed 1779 people in 2005. The number of road accidents increased to 12038 in 2007 and killed 2024. At the end of 2009, there were 122299 road accidents in the country and 2237 lives were lost.

Road accidents cost the world an amount of 518 billion dollars annually. It is estimated that if nothing is done globally to minimize the rampant nature of road accidents and most especially the causes of deaths of casualties before they are sent to hospitals then by the year 2020, 1.9 million people will be killed by road accidents in the world (World Health Organization, 2011).

2.3 Causes of Road Accidents

Road accidents is most unwanted thing to happen to a road user, though they happen quite often. The most unfortunate thing is that, we don't learn from our mistakes on road. Most of the road users are quite well aware of the general rules and safety measures while using roads but it is only the laxity on part of road users which cause accidents and crashes.The main cause of accidents and crashes are sue to human errors. Some of the common behavior which results in road accidents include;

a. Over Speeding

b. Drunken Driving

c. Distraction to Driver

d. Red Light Jumping

e. Avoiding Safety Gears like Seat belts and Helmets

Various national and international researches have found these as most common behavior of road drivers which leads to or cause accidents.

a. *Over Speeding*

Most of fatal accidents occur due to over speeding. It is a natural psyche of humanstoexcel. Ifgivenachancemanissuretoachieveinfinityspeed. Butwhen we are sharing the road with other users we will always remain behind some or othervehicle. Increaseinspeedmultiplestheriskofaccidentandseverityofinjury during accident. Faster vehicles are more prone to accidents than the slower ones and the severity of accident will also be more in case of faster vehicles. The higher the speed, the greater the risk. At high speed the vehicle needs greater distance to stop that is breaking distance. A slower vehicle comes to halt immediately while faster one takes long way to stop and also skids a long distance due to law of motion. A vehicle moving on high speed will have greater impact during the crash and thus will cause more injuries and deaths. The ability to judge the forthcoming events also gets reduced while driving at faster speed which causes error in judgment and finally a crash.

b. *Drunken Driving*

Consumption of alcohol to celebrate any occasion is common. But when mixed withdrivingitturnscelebrationintoamisfortune. Alcoholreducesconcentration. It decreases the reaction time of human body. Limbs take more to react to the instructions of brain. It hampers vision due to dizziness. Alcohol dampens fear and incite humans to take risk. All these factors while driving cause accidents and many a times it proves fatal.For every increase of 0.05 of blood alcohol concentration, the risk of accident doubles (Official Website of Jharkhand Police). Apart from alcohol, many drugs, medicines also affect the skills and concentration necessary for driving.

c. *Distraction to Driver*

Though distraction while driving could be a minor cause but it can cause major accidents. Distraction could be outside or inside the vehicle. The major distraction now a days is talking on a mobile phone whist driving. Act of talking on the phone occupies major portion of brain and the smaller part handles the driving skills. The division of the brain hampers reaction time and ability of judgment scientifically. This becomes one of the reasons of accidents.

d. *Red Light Jumping*

The red light jumping is a common sight at road intersections that vehicles cross without caring for the night. The main motive behind Red light jumping is saving time. The common conception is that stopping at red signal is wastage of time and fuel. Researches have shown that, traffic signals followed properly by all drivers saves time and commuters reach destination safely and timely. A red light jumper not only jeopardizes his or her life but also the safety of other road users. This act by one driver incites other driver to attempt it and finally causes chaos at crossing. This chaos at intersection is the main cause of traffi c jams. Eventually everybody gets late to their respective destinations. It has also been seen that the red light jumper crosses the intersection with greater speed to avoid crash and challan but it hampers his ability to judge the ongoing traffic and quite often crashes.

e. *Avoiding Safety Gears like seat belts and helmets*

The use of seat belts in four-wheeler is now necessary and not wearing seat belts invites penalty, same in the case of helmets for two wheeler drivers. Wearing seat belts and helmets doubles the chances of survival in a serious accident. Safety Gears keep you intact and safe in case of accidents. Two wheeler deaths have been drastically reduced after use of helmet has been mandatory.

Many other researchers on the other hand have come out with the causes of road accidents similar to the once listed already. For instance, in 2009, Ayeboo conductedaresearchonroadaccidentsandidentifiedthatthenumerousaccidents on our road networks have been linked to various cause which include over- speeding, drink driving, worn over taking, poor road network and the rickety vehicles which ply on our roads.

Also, the National Road Safety Commission (NRSC) in the year 2007 identified over twenty causes of road accidents in Ghana which include wrong-overtaking, recklessness, lack of proper judgment of drivers, intoxication, machine failure, over-loading, unwillingness to alight from motion objects, dazzling and defective lights, skid and road surface defect, obstruction and level of crossing. Other factors are use of mobile phones whilst driving, failure to wear the seat belt in a four wheeler or a helmet on a two- wheeler drive, corruption and inadequateenforcementofroadlawsandtra ffi cregulations(NationalRoadSafety Commission, 2007).

Upon the factors identified by the National Road Safety Commission, Ocansey in the 2011 conducted a research also on road accidents and observed that, poor vision of drivers could also be a major contributory factor to road accidents. It was obvious that the actual factors which may be contributing to road accidents inGhanahavenotbeenidentifiedsincemostofthefactorsidentifiedbytheNRSC have not been tested with any statistical tool to know their respective significant level when it comes to road accidents.

With reference to the country report on road safety in Cambodia, it was revealed that road accident is caused by factors such as road defects and vehicle defects. It was also observed from the report that road accidents in Cambodia was increased by 50 percent in five years whilst fatality was doubled. In order to help reduce the rate of road accidents a suggestion was made to establish a Road Safety Committee, accident data system, accident evaluation policy and drivers training measures, Ung(2007).Thecausesofroadaccidentselsewherehavebeenconnected to on or combination of the following factors; road design, driver's behavior, equipment failure and poor road network. Studies thus have shown that over 95 percent of all road crashes or accidents are caused by the behavior of the driver and the combination of on or more of three factors (equipment failure, poor road network and road design).

Detailed investigations by Starks (1967), Mackay (1969), Kemp and Ledru (1972) have all revealed that road accidents are multi-factor events and that there are usually a number of factors, such as the condition and the design of the road, the road worthiness of the vehicle and the behaviour of the road user, contributing to road accidents. Hence preliminary results from on-the-spot accident investigation team at Transport and Road Research Laboratory (TRRL) in 1972 showed that roadfactorsincludingenvironmentalconditionswerecontributoryin27percentof the accidents studied, vehicle design or maintenance in 20 percent of the accidents and the road user behaviour in almost 90 percent of all accidents researched.

It was shown in Great Britain that, rainfall can have a significant effect on road accident rates in the research conducted by Codling (1974). In his research of accidents and rainfall he showed that there are significantly more accidents in Great Britain during wet weather, partly as a result of obscured visibility and partly as a result of the increased likelihood of a skid. Investigations were made to determine the effect of rainfall on accidents in Nairobi, Srabaya and Kuala Lumpur.

Research conducted by Salt and Szatowski (1973) showed that the skid resistance of the road surface is also an important factor. In this study, recommendations were made for the minimum desirable levels of skid resistance on sections of the road ranging from 'very difficult' that is approaching traffic signals to 'easy' that is residential roads away from junctions.

The research by Mackay (1969) and Kemp (1972) indicated that the condition of vehicle tyres is also an important factor of road accidents. Consequently, surveys were carried out in Nairobi of the condition of vehicle tyres in that city. Whitlock (1971) made a study on social violence and showed that road accident fatality rates could be related to other aspects of social behavior such as murder and suicide rates in a given country. He attempted a similar research in various developing countries but was unable to isolate other major factors such as the level of vehicle ownership and medical facilities available. Other workers in the filed of human behaviour such as Tillman (1949) have also suggested that drivers involved in road accidents may be of a certain type (may be possibly a criminal) but Drew (1963) on other hand disagreed to this idea.

The effect of alcohol on road accidents rates has long been noted and an article by Beaumont and Newby (1972) ws used to illustrate this fact. In Great

Britain legislation ws introduced making it an offence to drive with more than 80 milligrams of alcohol per 100 millilitres of blood. Newby points out the difficulties of demonstrating the effectiveness of this legislation alone since other factors were involved at the same time.

Upon all these facts about thecauses of road accidents fromresearches conducted, many Ghanaians still link some of the road accidents in Ghana to evil forces, witchcraftandsuperstitions. Thus, itisbelievedthat, asaresultofthesespiritual activities, most people die in road accidents so that more blood would be obtained by the witches evil forces and wizards for their activities Okyere (2006).

2.4 Accident Rates in Developing Countries

In 1948, Smeed studied the relationships between road accident fatality (per licensed vehicle) and vehicle ownership in 20 developed countries, using data for the year 1938. An equation was derived as follows:-

$$\frac{F}{V} = 0.0003\left(\frac{V}{P}\right)^{-0.67} \qquad (2.1)$$

where F = fatalities from the accidents V= number of vehicles in use P= population.

Further researches by Smeed showed that the above relationship derived using the data for the year 1948 was still remarkably good fit for the data from many countries as late as 1968. Smeed used this relationship to show that the future number of road accidents in a country can be predicted from the knowledge for the future number of people and vehicles in that country. Thus:-

$$F = 0.0003(VP^{2})^{0.33} \qquad (2.2)$$

where F,V and P are as defined previously

In 1973, Jacobs derived an equation similar to that obtained by Smeed by taking fatalityandvehicleownershiplevelin32differentdevelopingcountriesfortheyear

1968. It was not Jacobs intention to use the equation to predict future number of accident fatality rates in developed countries and developing countries. The equation derived by the author:-

$$\frac{F}{V} = 0.0007 \left(\frac{V}{P}\right)^{-0.4} \quad (2.3)$$

showed that, fatality rates were greater in developing countries.

Smeed (1968) also found out that definitions of road accident fatalities varied in some countries from that recommended by the United Nations Economic Commission for Europe (ECE). In addition he calculated adjustment factors for fatalities for those countries not using the ECE definition.

A detailed analysis of fatality and vehicle ownership rates was made by Jeffcoate (1958) in Great Britain over the period 1909-1938. Similarly, Dondaville (1970) conducted a research of similar relationships in the United states over the period 1912-1967 and compared resulting equations with fatality and vehicle ownership rates existing in many different countries throughout the world in the year 1967. Dondaville revealed from his research that the relationship between fatality rate and vehicle ownership in the USA from the year 1912 to 1967 was similar to that derived for different countries for 1967.

Reasons for the existence of the relationship derived by Smeed (1949) were put forward by Garwood (1968). Some explanation is required of the fact that as a country or a group of countries, experience vehicle ownership, there is a decrease in the fatality rate. As suggested by Garwood, more vehicles should mean a great possibility of vehicle to vehicle conflict and therefore the likelihood of more accidents.

Garwood (1968) made a suggestion that the reasons for the fail in the total number of casualties per vehicle mile in Great Britain were:-

a. the decreasing proportion of two-wheeled motor traffic

b. pedal cyclist and pedestrian travel is not included in the assessment of vehicle

miles but their casualties are not increasing as fast a rate as in the total of other casualties.

c. the number of pedestrian casualties per motor vehicle mile for the different classes of vehicle is falling.

These and other points were investigated by Garwood and Johnson (1971) and

Smeed (1968), who then confirmed the validity of the above points made by Garwood. Smeed (1968) conducted a research on the changes of fatality and casualty rates and vehicle ownership over 9 year period in 15 mainly developed countries. He observed from his research that, percentage changes in fatality and casualty rates over this period were not related to changes in vehicle ownership. He also found out that, in all but one of the countries, the fatality rate per licensed vehicle reduced whilst in all the countries vehicle ownership increased. The analysis was repeated by Jacobs, 1973 for a number of developing countries and countries showing abnormal trends in fatality were identified. Using the United Nations' Yearbooks (1969), similar relationships were obtained by Jacobs (1973) in a number of developing countries. In this analysis, further insight in the reasons for high indices in developing countries were obtained by using information on medical facilities available in these countries.

Ina detailed review of procedures of rural traffic census used by developing countries, Howe (1972) observed that few developing countries carried out comprehensive national and trend censuses. Consequently Howe was not able to convert the numbers of fatalities into fatality rates per million vehicle kilometers travelled.

2.5 Comparative Accident Costs in Developing Countries

In developing countries, the very problem of valuing road accident costs us a contentious economic issue. Mishan (1971) and Adams (1974) have argued that, thetechniquescommonlyusedhavelittlevalidityfromatheoreticalstandpoint. It is the inability to ask a person what value he or she places on his or her life. There is clearly a difference between costing a certain number of accidents that will probably take place tomorrow and asking people how much they would pay not be killed in a road accident. On the valuation of human life, Godwin(1973) argues that objections may be avoided, if not overcome, only if it clearly specified that the monetary value arrived at is a minimum that society would find it worthwhile spending in order to avoid a fatal accident.

Whatever difficulties there may be in costing road accidents, attempts have been made in Great Britain for instance to derive the overall annual cost of road accidents since before the war. The first estimate at TRRL was derived by

Reynolds (1956) and was based on procedures set down in a report in 1947 to the Minister of transport. This could be defined as a national income approach and the costs which are included could be derived in two groups: those that cause a diversion of current resources, and the losses of future output because of death or injury. Among the former are, repairing of damage to the vehicle nd other property, medical treatment and the administration of insurance, police and the law.

In the year 1967, Dawson produced a comprehensive report on road accident costs using the methodology outlined by Reynolds (1956). In this report Dawson made a suggestion that loss of output should be estimated using the 'net' approach. In this method, the measure required of the loss of future output is its net present value, which is the difference between the future loss of output of those killed and the injured, assuming an otherwise normal expectation of working life, and the future consumption of those killed, assuming a normal expectation of life , both having been discounted to give present day values.

In addition, a further report by Dawson in 1971 showed two significant differences from that produced earlier. In the later report, the 'net' approach described from the previous paragraph was changed to a 'gross' method, where, in the case of road death, future consumption was not subtracted from future loss output. In the later work by Dawson, an attempt is made to assess the cost of suffering and bereavement experienced by the friends and relatives of road accident casualties.

In an attempt to find out how various developing countries cost road accidents, the author reviewed the studies that had been made on this subject in a number of countries. Within the limitations of the different definitions employed, the different factors taken into account, the dates of the researches, the methods of measuring costs and the currency of measurement, the various researches were compared and general conclusions reached.

Most of the researches conducted were concerned with establishing the number of deaths and injuries with road traffic accidents, valuing the resulting loss in output and vehicle damage, and summing the valuation of all injuries and damage to give a total accident cost. In 1972, a research conducted in Ivory Coast by the Ministry of Public Works employed a slightly different data format which meant that the total accident costs could not be derived. A research conducted by Burton (1968)in South Africa on the other hand derived unit costs of injuries and vehicle damage as well as total accident costs: the method of deriving the costs was by sampling the insurance payment s made to the accident claimants.

Furthermore, another study conducted in Kenya by the East African Transport Planning Research Unit in 1965 incorporated both accident and casualty information. This was very useful because it provided a means of specifying the probable number of persons with different types of injury who will be involved in different types of accidents. In this way, the costs of different types of injuries and accidents can be assessed. In an additional research also conducted in Turkey, although the total number of persons killed and injured are documented, the numberofseriousandslightinjuries werebasedonassumptionsabouttheseverity index. Alder(1965)inhisstudysuggestedthatforlessdevelopedcountries, thevaluation of life is not important since the reduction in fatalities is likely to be of minor significance compared with other benefits to be derived from the transport investment. This approach may be true of road investment in rural areas but hardly true in urban areas where traffic management schemes are designed to obviate accident black spots. The above researches together with those from Ghana in 1973, Thailand in 1966 and S. Rhodesia in 1963-1964 have all shown that accident costs in developing countries are a significant cost item.

2.6 Casualty's Death and their Causes in Road Accidents

The cause of casualty's death in road accidents have been linked to many factors such as failure of drivers and vehicle passenger or occupants to put on their seat belt and riders of two-wheelers failing to put on their helmet, Afukaar et al.

(2009).

Researches have revealed that sleeping related accidents tend to be more dangerous and as such many people are killed. This as a result is the driver's inability to prevent and stop certain actions such as applying the breaks before collision and steering onto the main road if the vehicle veers off the road.Research by Strhl et al. (1998) identified that in order to minimize the risk of drowsy driving and its related crashes, drivers are cautioned to have enough sleep and a sufficient one of course, drivers are to avoid the habit of drinking most especially when feeling sleepy and also reduce driving between midnight and 6:00 am.

Zomer in 1990 on other hand from his research, identified that the number of casualties in sleep related road accidents is 50 percent more than all accidents. It wasalsoidentifiedfromhisresearchthatthesleeprelatedroadaccidentshadthree times fatalities and doubles the seriously injured as compared to non-sleeping related accidents. Normally the sleep related accidents are more dangerous and kills a lot of people because there is no control on the part of those involved in the accident, especiallythedriver. Insuchasituation, therearecertaincircumstances which might have been avoided to reduce the number of casualties but because of the driver's inability to control the vehicle the people suffer the consequences. A research by Allan et al (1995) in North Carolina also gives an indication that, sleep related accident was the most severe accidents among all other types of road accidents.Suggestion by Homes and Reyner (1995) additionally revealed that,due to the inactive nature of the sleeping driver to control the vehicle before the accident, sleep related accidents have high risk of death as compared to the other forms of accidents.

Research conducted by Broughton (2007) showed that the driver and occupants of an older vehicle are usually at more risk of being killed than those in the new vehicle in case two vehicles collide. Broughton (2007) additionally estimated from his research that, the mean risk of death of drivers of vehicles which were registeredfrom2000to2003werelessthanhalfoftheriskforthedriverofvehicles which registered in from 1998 to 1999. This as a result gives the notion that, the age of a vehicle involved in an accident cannot be ruled out of the factors with respect to assessing the cause of death of casualties in road accidents. This fact may be due to the weaker nature of the parts of the older vehicles and probably the improvement and modernization in the manufacturing of new cars.

Other researchers have also found the size of a vehicle to be a contributing factor when it comes to the death of road users (casualties) in traffic crashes. For instance,from the research findings of Broughton (2007) into road accidents with the help of the data from 2001 to 2005, it was identified that the driver casualty rate increases with the size of the other vehicles in collision. The problem now is , the weight and size of vehicles have been improved by 30 percent yet the number of casualties' deaths keeps on increasing in accordance with that for the past 30 years, so the question is, what then is the cause of casualties' death. The fact still remains that people rely so much on the strength of the vehicles and take undue risk especially the youth, Broughton (2007).

Other research by Broughton (2007) has also shown that young drivers and young passengers die more in road accidents than their older counterparts.

In addition, Kumar et al (2008) in their research conducted in South Delhi, found out that, among all the people who were killed in road accidents, 88.2 percent of them were males. Earlier researchers such as Salgado and Colombaje (1998),

Shadev (1994) and Herrikson (2001) actually confirmed this result and all of them proposed and sustained their researches respectively that more males are killed in road accidents than females.

According to the a research conducted in Wales and Great Britain by Department of Transport,2006 to assess the pattern of death of various age groups of casualties and their genders respectively from the period of 2000 to 2002, it identified that,

40 percent of males and 30 percent of females' who died in road accidents were in the age group 16-19 years. This number had a risen in 44 percent for males and 38 percent for females by the end of 2005. It is however interesting to observe that this pattern changes with age, as the road users grow then the number of females who die through road accidents become more then that of the males. Another factor which identified by Clarke et al (2007) in their research as a contributory factor to death of casualties in road accidents is termed as 'drink driving'. The reason for this could be associated to the inability of the drunk

driver to control the vehicle as a result of sleeping, Zomer et al (1990). In the year 2005 at Great Britain and Wales, there were 1106 vehicle drivers who were killed in road accidents and a research conducted by Clark et al (2007) into this data revealed that 40 percent of those who died worn no seat belts and most of them were people within the age bracket 17-29 years. This as a result gives the notion that not wearing seat belts can also be a contributor to the death of casualties in road accidents. It was further revealed by Clarke et al (2007) that, the desire for waring the seat belt increases as one grows beyond 30 years. Broughton (2007) also in their research found out that drivers and vehicle passengers tend to avoid the wearing of seat belts in the night and as a result casualties' death in road accidents is higher in the night more than in the day.

According to the British Red Cross, 1997, it was indicated that one of the most commonest thing identified as the cause of death in road traffic crashes is anoxia- loss of oxygen supply, which cause a blockage in the air ways of the casualties and if immediate assistance is not given to the casualty, he or she dies after a short while due to inadequate supply of oxygen. Redmond (1994) in his research stated that, although there are certain forms of accidents which cannot be prevented, it is very true with evidence that pre-hospital death of road traffic crashes victims can be avoided if and only if timely and proper first aid measures are put in place. Otherresearcheshavefoundoutthemedicalassertionthat, foranyaccident, there is a golden hour which exist for casualties after the accident. It is therefore very important that immediate first aid is provided to road accident victims before they are rushed to the hospital.

2.7 Road Accident models

Numerous road accident models have been developed to estimate the expected number of accident frequencies on roads as well as to identify 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 occurrence (PersaudandDzbik, 1993). Previousresearchershavefocusedonnon-behavioural factors such as traffic flow characteristics, road geometry and environmental conditions. Persaud and Dzbik (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 higher accident risk is associated with congestion and afternoon rush hour. Shankar et al, intheirresearchworkin1995modelledthemonthlyaccidentfrequencyofrural 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.

Regression models have been most commonly used to relate road accident frequency with explanatory variables. The result of model strongly relies on the choice of regression technique. The earlier traffic accidents used ordinary linear regression models, which follow the assumption of a normal distribution for the response and predictor variables. However, the conventional linear regression method should be used with caution because of the problems associated with non-negative and error terms (Jovanis and Chang, 1986; Abdel-Aty and Radwan, 2000). Generalized linear model using Poisson distribution with non-negative error structure as a mean to describe the random, discrete and non-negative accident was recommended by Jovanis and Chang (1986). Poisson regression assumes exponential relationship between response and explanatory variables. Eenink el al. (2007) in their research advised a basic form of accident prediction model:

$$E(\theta) = \alpha Q e^{\beta \Sigma r x_i i} \quad (2.4)$$

where the expected number of the road accidents E(θ) is a function of the traffic volume, *Q* and a set of risk factors x_i.

Abdel and Radwan (2000) attempted to use the Poisson regression methodology and then rejected it with the reason that, different mean and variance value of the dependentvariableshowedoverdispersionintheaccidentdata. Consequently, the Negative Binomial model also called the Poisson-gamma model was then adopted as a superior alternative model to accommodate the vehicle accident analysis for rural highways, arterial roadways, urban roadways and rural motorways (Shankar et al.

1995;Lord,2005;Montella,2008; Pemmanaboina,2005). In the cases where considerable zeros and extremely low mean value are observed in accident data, negative binomial model is insignificantly reliable to fit the data and the dispersion parameter can be mis-estimated as a result. In order to overcomethedifficultiesarisingwithzeroaccidentsamples, someresearchersused extended Poisson and negative binomial models which can account for example the zero-inflated Poisson and the zero-inflated negative binomial (Miaou, 1994; Shankar et al.,1997;Lord et al., 2005). Zero-inflated models deal with the dual- state system: the zero-accident state in which there is no accident observed, and the non-zero accident saate where accident frequency follw some known distributions such as Poisson or negative binomial distribution.

Recently, a number of approaches have been proposed in the domain of accident models seeking improvement from the traditional methods discusses earlier. For instance, El-Basyouny and Sayed in 2006 proposed a modified negative binomial regression technique which improved goodness of fit . Caliendo et al. (2007) on the other hand applied the negative multinomial distributions 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-Aty and Abdalla in 2004 and Lord and Persaud in 2000. In New Zealand, the generalized linear models have been used in accident research of different intersections types as 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 (2001).

In order to obtain true and very accurate model that can fit into accident data for 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 in 1984 also modeled 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 over time. As a result, he came out with a general formula;

$$D = \text{constant} * N^{m_1} * P^{m_2} \tag{2.5}$$

which could also be used to predict te 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 and m_1 and m_2 are variables of interest. The difficultyinapplyingtheequationbyAndreassenishowtodeterminetheconstant term and the indices which might very from one country to the other.

Additionally, a research by Minter (1987) discussed the applications two accident models which were developed by Wright and Towel 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 use road length as an additional parameter and established a model for road accidents. A general model was also presented by Jamal and Jamil in 2001 to predict road accident facilities.

2.8 Review of Methodological Applications of Statistical models to Accident frequency Analysis

In terms of methodological perspective, many applications of accident frequency statistical modeling have been conducted.

To demonstrate the application of a Poisson regression in accident frequency analysis, consider a set of *i* road sections. Let y_{ij}, a random variable, be the number of road accidents during a given time period,*j*,

$$P(y_{ij}) = \frac{\exp(-\theta_{ij})\theta_{ij}^{y_{ij}}}{y_{ij}!} \tag{2.6}$$

where P (y_{ij}) is the probability of *n* road accidents occurring on a highway section *i*in a given period of time *j*, and θ_{ij} is the expected number of y_{ij},

$$E(y_{ij}) = \theta_{ij} = \exp(\beta X_{ij}) \tag{2.7}$$

for a roadway section *i* in month *j*, where β also is vector of unknown regression coeffi cients and can be estimated by the standard maximum likelihood estimation (Greene, 1997). X_{ij} describes the roadway section geometric characteristics and other relevant roadside feature conditions for highway section *i* in a given period of time *j*.

A limitation of the Poisson distribution is that the variance and the mean must be approximately equal (Maddala, 1977, Cox, 1983 and Agresti, 1996). The possibility of overdispersion that is when the variance exceeds the mean rather the equaling the mean as the Poisson requires, is always a concern in modeling accident frequency and may result in biased and inefficient coefficient estimate. The simple regression based test by Cameron and Trivedi in 1986 and 1990 respectively can bee performed to detect overdispersion in Poisson distribution or process. This regression based test involves simple least square regression to test the significance of the overdispersion coefficients.

In order to relax the constraints of overdispersion introduced by the Poisson regression model, a Negative Binomial distribution with a Gamma distributed error is commonly used (Miaou, 1994, Shankar et al., 1995, Milton and Mannering 1998, Carson, 1998). The Negative Binomial model is derived by rewriting equation 2.5 such that we obtain,

$$\theta_{ij} = \exp(\beta X_{ij} + \varepsilon_{ij}) \tag{2.8}$$

where $\exp(\varepsilon_{ij})$ is the Gamma distributed error term, and this in addition allows the variance to be greater than the mean as shown below:

$$Var[y_{ij}] = E[y_{ij}][1+ \alpha E[y_{ij}]] = E[y_{ij}] + \alpha E[y_{ij}]^2 \tag{2.9}$$

The Poisson regression model is regarded as a limited model of the Negative Binomial regression model as α approaches zero (0) which indicates that the selection between theses two models is dependent upon the value of α. The Negative Binomial distribution has the formulation:

$$P(y_{ij}) = \frac{\Gamma(y_{ij} + \alpha^{-1})}{y_{ij}!\,\Gamma(\alpha^{-1})} \left(\frac{1}{1+\alpha\theta_{ij}}\right)^{\alpha^{-1}} \left(\frac{\alpha\theta_{ij}}{1+\alpha\theta_{ij}}\right)^{y_{ij}} \tag{2.10}$$

The standard maximum likelihood methods can be used to estimate θ_i (Greene, 1997). By using equation 2.8 the likelihood function for the Negative Binomial regression model is,

$$L(\theta_{ij}) = \prod_{i=1}^{N} \frac{\Gamma(y_{ij} + \alpha^{-1})}{y_{ij}!\,\Gamma(\alpha^{-1})} \left(\frac{1}{1+\alpha\theta_{ij}}\right)^{\alpha^{-1}} \left(\frac{\alpha\theta_{ij}}{1+\alpha\theta_{ij}}\right)^{y_{ij}} \tag{2.11}$$

where N is the total number of road accidents. This maximum likelihood function is used to estimate the unknown parameters, β and α^{-1}.

In order to address the zero-inflated accident counting processes on roadway sections, the Zero-Inflated models (Zero-Inflated Poisson (ZIP) and Zero-Inflated Negative Binomial (ZINB)) regression models have been employed for handling zero-inflated count data. The Zero-Inflated Poisson and the Zero-Inflated Negative Binomial models both assume that two different processes are at work for some zero accident count data. The Zero-Inflated Poisson (ZIP) assumes that 26 the events, $Y = (Y_1, Y_2, \ldots, Y_N)$, are independent and

$$Y_i = 0 \text{ with probability } \Psi_i + (1-\Psi_i)\exp(-\Theta_i) \tag{2.12}$$

$$Y_i = y \text{ with the probability } (1-\Psi_i)\frac{\exp(-\theta_i)\theta_i^{y}}{y_i!} \tag{2.13}$$

where y is the number of road accidents, and the mean and the variance of Y_i can be written as

$$E(Y_i) = (1-\Psi_i)\theta_i^{y} \tag{2.14}$$

$$Var(Y_i) = E[Y_i] + \frac{\Psi_i}{1-\Psi_i}E[Y_i]^2 \tag{2.15}$$

The maximum likelihood estimates (MLE) is used to estimate the coefficients of the ZIP regression model and the confidence intervals can be constructed by the likelihood ratio tests.

The Zero-Inflated Negative Binomial (ZINB) regression model on the other hand follows similar formulation as the ZIP and assumes that the events $Y = (Y_1, Y_2, \ldots, Y_N)$ are again independent and

$$Y_i = 0 \text{ with probability } \Psi_i + (1-\Psi_i)\left(\frac{1}{1+\alpha\theta_i}\right)^{\alpha^{-1}} \tag{2.16}$$

$$Y_i = y \text{ with probabilty } (1-\Psi_i)\frac{\Gamma(y_i+\alpha^{-1})}{\Gamma(\alpha^{-1})\, y_i!}\left(\frac{\alpha\theta_i}{1+\alpha\theta_i}\right)\left(\frac{1}{1+\alpha\theta_i}\right)^{\alpha^{-1}} \tag{2.17}$$

Maximum likelihood results are again used to estimate the coefficients of a zero- inflated negative binomial regression model (ZINB).

The choice of an appropriate accidents frequency model for road section with zero- accident involvement is critical. Hence, one cannot test directly as to whether a zero-accident state and non-zero accident state are totally different or not since the traditional Poisson and Negative binomial and the Zero-inflated models are non-nested. In oder to test the appropriateness of using a zero-inflated model rather than the traditional model, a Vuong in 1989 proposed a test statistic called

the Vuong statistic for non-nested models that is well suited for this setting when the distribution can be specified.

The statistic for testing the non-nested hypothsis of Zero-inflated model versus traditional models is (Greene, 1997, Shankar et al., 1997),

$$V = \frac{\sqrt{n}\left(\frac{1}{n}\right)\sum_{i=1}^{n} m_i}{\sqrt{\left(\frac{1}{n}\right)\sum_{i=1}^{n}\left(m_i - \bar{m}\right)^2}} = \frac{\sqrt{n}\left(\bar{m}\right)}{S_m} \qquad (2.18)$$

where $\bar{m}$ is the mean, S_m is the standard deviation and *n* is the sample size.

Vuong's value is asymptotically standard normally distributed and if the absolute value of *v* is less than 1.96 (the 95 percent confidence interval for the t-test), the test does not indicate any other model. Thus the zero-inflated regression model is only favored if the *v* value is greater than 1.96, whilst a *v*value less than -1.96 favors the Poisson or negative binomial regression model (Greene, 1997).

2.9 Applications of Statistical models for count data (Accident data) in some Research Works

With respect to the applications of statistical modls for over-dispersed count data (accident data), the researcher considered how statistical models for over- dispersed count data has been applied in the field of road accidents by other researchers.

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

and Negative Binomial regression model applications (Shankar,Mannering and Barfield, 1995, Poch and Mannering , 1996).

Shankar , Mannering and Barfield (1995) used both the Poisson(when the data were not significantly over-dispersed) and the Negative Binomial regression models (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 modeled. 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 tool 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 models did not address the probability of zero-inflated counting processes. They distinguished the truly safe road section (zero accident state) from the unsafe section (non-zero accident state but with the possibility of having zero observed accidents) to show that a zero-inflated model structure is often appropriate for estimating the accident frequency of road sections (Mullahey 1986, Lambert 1992, Greene 1997). Zero-inflated probabilityprocesses, suchasthezero-inflatedPoisson(ZIP)andthezero-inflated Negative Binomial (ZINB), helps one to better isolate independent variables that determine the relative accident likelihoods of safe versus unsafe roadways. Miaou recommended that the Poisson regression model is an appropriate model for developing the relationship when the mean and the variance of the accident frequencies are approximately equal. If the over-dispersion is found to be moderate or high, the use of both the Negative Binomial and the zero-inflated Poisson regression models were found to be more appropriate. On the whole, the zero-inflated Poisson regression model seems a justified model when the accident data exhibit a high zero-frequency state.

2.10 Review on Poisson Regression and Negative Binomial Regression models

The Poisson regression model is a technique to describe count dependent variables (Cameron et al, 1998).Usually the Poisson model is applied to study the occurrence of small number of counts as a function of a set of predictor variables, in experimental and observational study in many disciplines such as Economics, Demography,Psychology, biology and medicine (Gardener et al, 1995).

Poisson regression model has often been applied to estimate standardized mortality and incidence ratios in cohort studies and in ecological investigations (Breslow et al, 1987). Some variants of the Poisson regression model have been proposed to take into account the extra-variability or over-dispersion observed in actual data, mainly due to the presence of spatial clusters or other sources of autocorrelation (Trivedi et al, 1998).

Apart from medical studies, the Poisson regression model has been applied in different fields of research, ranging from herd management assessment to animal health in domestic and wild animals and control of infectious diseases in different animal species. The Poisson regression model has also been applied to data analysis in multidisciplinary study on cancer incidence in veterinary and other workers of veterinary industry compared to that of other part of active population in Sweden (Travier et al, 2003).

A research conducted by Miaou and Lum (1993) assessed the statistical nature of the two conventional linear regression models which have been used by most researchers to develop road accident relationships and revealed that, these two linear regression models fail to consider and describe the distributional characteristics of road accidents as in its randomness, non-negative nature of road accidents, the discrete and count properties of the event of road accidents. Miaou in 1994 additionally compared Poisson and Negative Binomial regressions since they all takes care of the distributional properties of accident data. Numerous researchersontheotherhandhavealsoassessedtheuseoflinearregressionmodels for road accidents models for road accident models and confirmed the limitations in such models, Kim et al (2005) used generalized log-linear models and Garber and Wu (2001) applied stochastic models in fitting models to road accident data. Simon (1961), following his seminal work differentiating the Poisson and negative binomial models (1960), was the first to publish a maximum likelihood algorithm for fitting the negative binomial. He was one of the many actuarial scientists at the time who were engaged in fitting the Poisson and negative binomial distributions to insurance data. His work stands out as being the most sophisticated, and he was perhaps cited more often for his efforts in the area than anyone else in the 1960s. Birch (1963) is noted as well for being the first to develop a single predictor maximum likelihood Poisson regression model which he used to analyze tables of counts. It was not until 1981 that Plackett first developed a single predictor maximum likelihood negative binomial while working with categorical data which he could not fit using the Poisson approach. Until the mid 1970's, parameterizing a non-linear distribution such as logit, Poisson, or negative binomial, so that the distributional response variable was conditioned on the basis of one of more explanatory predictors, was not generally conceived to be as important as understanding the nature of the underlying distribution itself,

i.e. determining the relationships that obtain between the various distributions. When considering the negative binomial distribution, for example, the major concern was to determine how it related to other distributions-the chi-square, geometric, binomial and Bernoulli, Poisson, gamma, beta, incomplete beta, and so forth.

Regression model development was primarily thought to be a function of the normal model, and the transformations that could be made to both the least squares response and predictors. It was not until 1981 that the first IBM personal computer became available to the general public, an event that changed forever the manner in which statistical modeling could be performed. Before that event, most complex statistical analyses were done using mainframe computers, which were usually at a remote site. Interactive analyses simply did not occur. Computer time was both time-consuming and expensive. The emphasis on distributional properties and relationships between distributions began to change following the development of generalized linear models (GLM) by Nelder and R.W.M. Wedderburn (1972).

The new emphasis was on the construction of non-linear models that incorporated explanatory predictors. In 1974 Nelder headed a team of statisticians, including Wedderburn and members of the statistical computing working party of the Royal Statistical Society, to develop GLIM (Generalized Linear Interactive Modeling), a software application aimed to implement GLM theory. GLIM software allowed users to estimate GLM models for a limited set of exponential family members, including, among others, the binomial, Poisson, and, for a constant value of its heterogeneity parameter, the negative binomial. Although GLIM did not have a specific option for negative binomial models, one could use the open option to craft such a model.

In Ghana much work has not been done in this regard but researchers such as Oppong (2012) used Poisson regression model to model road accidents, Salifu (2004) used the generalized linear model to predict road accidents in unsignable urban junctions, Afukaar et al, (2007) and Ackaah and Salifu (2011) applied Poisson, Negative Binomial and Log-linear regression models to road accidents in Ghana.

2.11 General Overview of Conway-Maxwell- Poisson Regression

The Poisson regression is one of the widely use distributions in statistical applications The distribution does have it s limitation since its mean and variance are equal. In many conditions this assumption is not realistic. Thus researchers have used di ff erent specifications of the Poisson to deal with this problem. Normally, this is done by introducing a mixing distribution or estimating the Poisson with extra parameters that account for over-dispersion. A recent specification of the Poisson distribution is that of the Conway-Maxwell-Poisson regression. The Conway-Maxwell-Poisson regression is a two parameter Poisson regression that was first introduced by Conway and Maxwell (1962). Even though Conway and Maxwell introduced this regression long time ago, it was relatively ignored in the literature and its properties were not fully developed.

Shmueli et al.(2005) reintroduced this regression and into the literature and developmanypropertiesforit. Theriseinthepopularityofthedistribution, other than it being relatively new, is that the Conway-Maxwell-Poisson (CMP) family belongs to the exponential family. Many useful properties of the exponential family have been developed and make the distribution favorable for the use of Bayesian analysis and other statistical inference. Since the reintroduction of the CMP distribution, there has been a recent increase in its applications including analyzing motor vehicle crashes, electric power reliability, retail predictions of households and prediction cancer recurrence.

CHAPTER 3

METHODOLOGY

3.1 Introduction

This chapter focusses on the detailed description of the methods used for this research work and explained theories behind the various distributions and models for the analysis. This chapter also addressed the possible probability distributions for accident data (count data) and their likely regression models which may include Poisson, Negative Binomial and the Conway-Maxwell-Poisson regression models.

3.2 Regression Models

With regards to count data, for example accident , crashes that are non-negative and discrete in nature, makes more sense to model these count data using Poisson, Negative Binomial and Conway-Maxwell-Poisson regression models. Regardless of whether the assumed model is Poisson, Negative Binomial or Conway-Maxwell- Poisson, itwillbeassumedthattheoccurrenceswillbeindependentofeachother.

But all these models can be presented and understood using the Generalized Linear model framework that emerged in the statistical literature in the early 1970’s (Nelder and Wedderburn, 1972)

3.3 Generalized Linear Models (GLM)

The Generalized Linear Models (GLM) was first introduced by Nelder and Wedderburn in 1972. Generalized Linear Models provides a unified framework to study various regression models, rather than a separate study for individual regression model. It is an extension of the classical linear models.

ThemainpurposeoftheGeneralizedLinearModelistospecifytherelationshipof the observed dependent variable with the covariates or independent variables. A particular interesting feature of this class of models is they allow for the violation of the Gauss-Markov assumption. This allows for the linearization of a non-linear relationship between the observed dependent variable and the predictor variables.

3.3.1 Components of Generalized Linear Models

A generalized linear model consist of three (3) components:

i. Random component: The random component refers to the dependent or responsevariablethathasorfollowsacertaindistributionoritisacomponent which specifies the conditional distribution of the response variable Y_i, given explanatory variables, x_{ij}.

ii. Systematic component: The systematic component on the other hand refers to the covariates (independent variables) that are linearly combined with the coefficients or it refers to a linear function of the regression variables called linear predictors,

$$\eta_i = \alpha + \beta_1 X_{i1} + \ldots + \beta_k X_{ik} = X \beta'_i \quad (3.1)$$

iii. Link function: The link function links the random component to the systematic component or it transforms the expectation of the response variable to the linear predictor,

$$g(E(y)) = X \beta_i = \eta_i \quad (3.2)$$

The inverse of the link function is sometimes called the mean function and is given as;

$$E(y) = g^{-1}(\eta_i) \qquad (3.3)$$

For additional linear models in which the random component includes the assumption that the dependent variable (response variable) follows the Normal distribution, the canonical link function is the identity link. The identity link gives the indication that the expected mean of the response variable is identical to the linear predictor. For example for the normal linear model the link function,

$$g(\mu_i) = \mu_i \qquad (3.4)$$

3.4 Exponential Family (Canonical form)

General Linear Models (GLM) can be used to model variables following distributions in the exponential family in their canonical forms with the probability density function;

$$f(y; \theta,\varphi) = \exp\left(\frac{y \cdot \theta - b(\theta)}{\varphi} + c(y,\varphi)\right) \qquad (3.5)$$

or

$$\log f(y; \theta,\varphi) = \frac{y \cdot \theta - b(\theta)}{\varphi} + c(y,\varphi) \qquad (3.6)$$

where θ is the canonical parameter that depends on the regressors via linear predictor and φ is the dispersion parameter that is often known. The functions b(·) and c(·) are also known and determine which of the family is used example the Normal, Binomial or Poisson distributions.

For distributions in the exponential family in their canonical form, the conditional mean and variance of y_i are respectively given by;

$$E(y_i/x_i) = b^0(\theta_i) \qquad (3.7)$$

and

$$VAR(y_i/x_i) = \varphi \cdot b''(\theta_i) \quad (3.8)$$

where $b'(\theta_i)$ and $b''(\theta_i)$ are the first and the second derivatives of $b(\theta)$. The dispersion parameter is normally fixed to one for some distributions for example the Poisson distribution.

3.5 The Poisson Distribution

The Poisson distribution is a discrete probability function that expresses the probability of number of event occurring in a particular period of time if these events with a known average rate and each count occur independently of the time of the last event. The Poisson distribution can also be used for the number of events i other specified intervals such as distance, area or volume. Poisson regression analysis has been used in the analysis of accident data for modelling traffic crashes in most parts of the world. Upon the numerous the applications of the Poisson regression, it has been mainly applied to evaluate the causes of road accidents.

The distribution was first introduced by Simeon-Dennis Poisson (1781-1840) and published in 1838 in his theory. The work of Simeon was focused on certain random variables *N* that count, among other things, the number of discrete occurrences that takes place during a time arrival given length, Poisson (1838). If the expected number of occurrences in this interval is θ then the probability that there are exactly *y* occurrences (*y* being the non-negative integer,y = 0, 1, 2, 3... is equal to;

$$P(y/\theta_i) = \frac{\exp(-\theta_i)\theta_i^{y_i}}{y_i!} \quad (3.9)$$

where

. *y* is the number of occurrences of an event- the probability of which is given by the function $f(y/\theta)$.

. θ is a positive real number, equal to the expected number of occurrences that takes place in a given period of time.

For example, if the events occur on an average rate of 10 times per minute, and one is interested in the probability for times of events occurring in a 15 minute interval, then one would use as the model a Poisson distribution with $\theta = 15 * 10 = 150$.

The parameter θ is not only the mean number of occurrences but also it is the

variance since $\zeta^2_y = E(y^2) - [E(y)] =^2 \theta$.

Hence, the number of observed occurrences fluctuates about the mean with a

standard deviation $\sigma_y = \sqrt{\theta}$.

Thefunctionof y isadiscreteprobabilitymassfunction. ThePoissondistribution can be derived from a limiting case of the Binomial distribution. The Poisson distribution can therefore be applied to systems with a larger number of possible events, each of which is rare. An example is cancer cases.

3.5.1 Derivation of the Poisson Distribution

The Poisson distribution is the most basic distribution for event counts. There are two ways of coming up with the Poisson distribution function as a good distribution of event counts. Approach 1 The first approach is to start with an abstract of an event that assumes;

i. events occur overtime

ii. there is a constant rate (θ) at which events occur-this rate is the expected number of events in a period of length *h*or time

iii. events are independent in that the occurrence of one event does not affect the probability of another and iv. asthelengthoftheinterval *h*goestozerotheprobabilityofaneventoccurring

in the interval (t,t + h) = θh and the probability of two events occurring in the interval is zero.

The dependent variable is the number of events that a occur in interval *t* of length *h*. The probability that th e number of events in (t,t + h) is equal to some value $y \in 0, 1, 2, \ldots$ is:

$$P(Y_t = y_i) = f(y_i) = \frac{e^{-\theta h_i} (\theta h)^{y_i}}{y_i!} \qquad (3.10)$$

InthisPoissonprocess, eventoccurindependentlywithaconstantprobability equal to θ times the length of the interval i.e θh. Typically, it is assumed that all intervals are of the same length equal to one and so we have;

$$P(Y_t = y_i) = f(y_i) = \frac{e^{-\theta_i} \theta_i^{y_i}}{y_i!} \qquad (3.11)$$

Approach 2

The second approach of deriving the Poisson distribution is to think of event counts as a count of rare events. Specifically ,a Poisson random variable approximates the binomial random variable when the binomial parameter *n* (number of trails) is large and *p* (probability of success) is small. The binomial distribution represents the probability of *y* successes (events) in *N* independent trials(possible events), where the probability of an event in each trial is *p*, has

the following distribution:

$$P(Y = y) = \binom{N}{y} p^y (1 - p)^{N-y} \qquad (3.12)$$

N N!

where y = $_{(N-y)!y!}$ We can therefore think of the binomial distribution with time that is thinking of the number of trials as a length of time. For instance , the span of time of length 1 can be divided into N ’sub-intervals’ of length $_N{}^1$. The probability of an event (p) equals $_N{}^\theta$ where θ is simply the number of events that occurred in all of the N periods. Thus the binomial distribution can be rewritten as

$$P(Y = y) = \binom{N}{y}\left(\frac{\theta}{N}\right)^{y}\left(1 - \frac{\theta}{N}\right)^{N-y} \qquad (3.13)$$

By dividing the span of time into more and more intervals and making the probability of an event (p) proportionally smaller, the time span becomes a continuum. Hence the probability of an *y* events as $N \to \infty$ is;

$$P(Y = y) = \lim_{N \to \infty} \binom{N}{y}\left(\frac{\theta}{N}\right)^{y}\left(1 - \frac{\theta}{N}\right)^{N-y} = \frac{e^{-\theta}\theta^{y}}{y!} \qquad (3.14)$$

This is the same as equation 3.11. The bottom line is that, for a larger Bernoulli trials, where the probability of an event in anyone trial is small, the total number of events observed will follow a Poisson distribution.

3.5.2 Properties of the Poisson Distribution

The Poisson distribution has certain characteristics or properties which includes:

i. As θ (mean count) increases, the mass of the distribution moves or shifts to the right. The reason being that E(y) = θ and θ can be taught as a mean count or the expected number of times an event occurs per unit time (fits with the binomial derivation) or as the expected or mean count (fit with more abstract model of event counts).

ii. Var(y) = E(y) = θ. This means that the model is restrictive (as any model with only one parameter).

iii. As θ increases the probability of 0s decreases. iv. As θ , the Poisson distribution approximates a Normal Distribution.

3.5.3 Verification of the Poisson distribution as exponential family member

By taking logarithm of the Poisson distribution function from equation 3.9, we obtain

$$\log f_i(y_i) = \log \frac{\theta^{y_i} e^{\theta}}{y_i!}$$

$$\log f_i(y_i) = y_i \log \theta - \theta - \log(y_i) \qquad (3.15)$$

where

. y_i is the number of occurrences of the counts or events (accidents)

. θ_i is th expected number of occurrences that occur during a given time interval

By comparing equation 3.10 to 3.6 (probability density function for exponential family in canonical form), we observe quickly that the coefficient of y_i which represents the link function is log θ_i and the canonical parameter (θ) on the other hand is given by;

$$\theta = \log(\theta_i) \qquad (3.16)$$

By solving for θ_i from equation 3.11. we obtain the inverse link function as;

$$\theta_i = e^{\theta} \qquad (3.17)$$

Also the second term in from the equation 3.6 by comparing it again to equation 3.10 can be written as;

$$b(\theta) = \theta_i \qquad (3.18)$$

By substituting the equation 3.12 into equation 3.13 we obtain;

$$b(\theta) = e^{\theta} \qquad (3.19)$$

The last term is a function of y_i only, so it can be observed or identified from equation 3.6 that; $c(y,\varphi) = \log(y_i!)$

Lastly it should be noted that the dispersion parameter ($\theta = 1$) can be taken, just as it is in the binomial case and verify that the Poisson distribution belongs to the exponential family.

3.5.4 Poisson Regression

The most basic model for event counts is the Poisson regression model. If the variance of the counts approximately equals the mean count, then the Poisson regression model is expressed as ;

$$P(y/\theta_i) = \frac{\exp(-\theta_i)\theta_i^{y_i}}{y_i!} \quad \text{for} y_i = 0, 1, 2, 3, \ldots \qquad (3.20)$$

where y_i is the number of counts (accidents) for a particular period of time *i*, θ_i is the expected number of counts (accidents) per period, which can be as result modeled as;

$$\theta_i = \exp(x_i)^0 \qquad (3.21)$$

where x_i^0 is the vector of the explanatory or independent variables and β is the vector of unknown regression parameters. The Poisson regression model assumes that the sample of *n* observations , y_i are observations on independent Poisson variables Y_i with mean θ_i. If this model is true, then the equal variance assumption of classic linear regression is violated since Y_i have means equal to their variances.

Hence the generalized linear model needs to be fitted,

$$\log(\theta_i) = x\ \beta^0{}_i \qquad (3.22)$$

We therefore say that Poisson regression model is a generalized linear model with Poisson error and log link, so that

$$\theta_i = \exp(x\ \beta^0{}_i) \qquad (3.23)$$

This as a result indicates that, a unit increase in an x_i increases θ_i by a multiplicative factor of $\exp(\beta_i)$.

The main constraint or assumption of the Poisson regression model is that, the mean and the variance are equal, that is $E(y_i/x_i) = Var(y_i/x_i) = \theta$. As a result, when there is heterogeneity or over-dispersion (when the variance increases faster than what the Poisson model allows), the Poisson regression model does not work well hence it is important to fit a parametric model that is more dispersed than the Poisson model and a natural choice in this case is the Negative Binomial model.

3.5.5 Derivation of the Poisson Mean Regression model

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

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

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

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

by taking the exponent of the above equation we obtain;

$$= e^{-\theta_i}\exp\left(\frac{\theta_i^{y_i}}{y_i!}\right)$$

$$= e^{-\theta_i}\exp\ln\left(\frac{\theta_i^{y_i}}{y_i!}\right)$$

$$= e^{-\theta_i}\cdot\exp(y_i\ln\theta_i - \ln y_i!)$$

$$= \exp\left(y_i \ln \theta_i - \theta_i - \ln y_i\right) \qquad (3.24)$$

By comparing the pdf of the exponential family in canonical form above to equ. 3.24 we have;

θ(canonical Parameter or link function)= ln θ_i, θ (dispersion parameter)=1 and

$c(y,\varphi) = -\ln y_i$

But $\theta = E(g(y)) = x\beta_i^0$ $\therefore$

$\ln \theta_i = x\beta_i^0$

$$\theta_i = \exp(x\beta_i^0) \tag{3.25}$$

as the Mean Poisson regression model.

3.5.6 Estimation of the Poisson Regression Model

With respect to the estimation of the Poisson regression model, there is a need to always parameterize θ_i as $f(x\beta_i)$. But it is known that θ must be positive because it is a count that is $E(Y_i > 0)$.

Hence in order to estimate the Poisson regression model it is usually assume that:

$$E(Y_i) = \theta_i = e^{x\beta_i^0}$$

(Link function)

This as a result yields the following probability model by substituting the above equation into equation 3.9 to obtain:

$$\Pr(Y_i = y) = \frac{e^{-\theta_i}\theta_i^{y_i}}{y_i!}$$

$$\Pr(Y_i = y) = \frac{e^{-e^{x_i\beta}}\left(e^{x_i\beta}\right)^{y_i}}{y_i!} \tag{3.26}$$

Hence the likelihood function for the whole sample is:

$$L = \prod_{i=1}^{N} \frac{e^{-\theta_i}\theta_i^{y_i}}{y_i!}$$

$$L = \prod_{i=1}^{N} \frac{e^{-e^{x_i\beta}}(e^{x_i\beta})^{y_i}}{y_i!} \quad (3.27)$$

The log likelihood thus is:

$$\ell = \sum_{i=1}^{N} [-\theta_i + y_i x_i\beta - \ln(y_i!)]$$

$$\ell = \sum_{i=1}^{N} -e^{x_i\beta} + y_i x_i\beta - \ln(y_i!) \quad (3.28)$$

When maximizing the log-likelihood, the last term can be essentially be ignored since it does not vary with θ_i and with β. The log-likelihood is globally concave and as a result easy to estimate.

The gradient vector then becomes;

$$G = \frac{\partial\ell}{\partial\beta} = \sum_{i=1}^{N} -e^{x_i\beta}x_i + y_i x_i$$

$$= \sum_{i=1}^{N} (y_i - e^{x_i\beta})x_i$$

$$G = \frac{\partial\ell}{\partial\beta} = \sum_{i=1}^{N} (y_i - \theta_i) = 0 \quad (3.29)$$

and

$$\partial\beta\partial\beta \qquad {}_{i=1}$$

Hence the Hessian function and with typical element is:

$$\frac{\partial^2 L}{\partial\beta_j \partial\beta_l} = -\sum_{i=1}^{n} [e^{x_i\beta_0} x_{ij} x_{il}] \; ; \; jl = 1, 2, 3, \ldots, p \tag{3.31}$$

$$\frac{\partial^2 L}{\partial\beta_0^2} = -\sum_{i=1}^{n} [e^{x_i\beta_0} x_i x_i] \tag{3.30}$$

Because equation (3.31) does not involve the response variable *y* then we let

$$k_{jl} = \frac{\partial^2 L}{\partial\beta_j \partial\beta_l} = -\sum_{i=1}^{n} [e^{x_i\beta_0} x_{ij} x_{il}]$$

and the information matrix is;

$$K = \sum_{i=1}^{n} [e^{x_i\beta_0} x_i x_i] \tag{3.32}$$

There is no close form solution to $\frac{\partial L}{\partial \beta} = \sum_{i=1}^{n}(y_i - e^{x_i'\beta})x_i' = 0$ so the Maximum Likelihood Estimate (MLE) for β must be obtained numerically. Thus, as the Hessian is a negative definite for all the x and β, the MLE($\hat{\beta}$) is unique if it exists.

From equation (3.31) ~~and~~ $k_{jl} = \frac{\partial^2 L}{\partial \beta_j \partial \beta_l} = -\sum_{i=1}^{n}[e^{x_i'\beta} x_{ij} x_{il}]$ we obtain;

$$k_{jlr} = E\frac{\partial^3 L}{\partial \beta_j \partial \beta_l \partial \beta_r} = -\sum_{i=1}^{n}[e^{x_i'\beta} x_{ij} x_{il} x_{ir}] \tag{3.33}$$

and

$$k_{jl}^{(r)} = \frac{\partial k_{jl}}{\partial \beta_r} = -\sum_{i=1}^{n}[e^{x_i'\beta} x_{ij} x_{il} x_{ir}],\; j,l,r = 1, 2, 3, \ldots, p \tag{3.34}$$

In order to make matters more understandable and transparent, we consider the case of a single covariate (independent variable) and an intercept. Hence x_i is a scalar observation and

$$L = \sum_{i=1}^{n}[-\theta_i + y_i(\beta_1 + \beta_2 x_i) - \ln(y_i)]$$

where $\theta_i = e^{\beta_1 + \beta_2 x_i}$, for i=1,2,3,...,n.

The first order conditions, $\frac{\partial L}{\partial \beta} = 0$ yield a system of K equations for each β of the form;

$$\sum_{i=1}^{n}(y_i - e^{x_i\beta})x_i = 0$$

Where $\hat{y}_i = e^{x_i\beta_0}$ is the fitted or predicted value of y_i. The fitted or predicted value is usually taken as the estimated value of (y_i/x_i). This first order condition indicates that the vector of residual is *r* orthogonal to the vectors of the explanatory variables.

3.5.7 Model Specification of Poisson regression

The primary equation or the probability density function of he model is;

$$P(Y/y_i) = \frac{\exp(-\theta_i)\theta_i^{y_i}}{y_i!} \qquad (3.35)$$

The most common formulation of this model is the log-linear function specification as in the equation

$$\log(\theta_i) = x_i'\beta \qquad (3.36)$$

The expected number of events per period is given by;

$$E(y_i/x_i) = \theta_i = e^{x_i'\beta} \qquad (3.37)$$

Thus;

$$\frac{dE(y_i/x_i)}{dx_i} = \beta e^{x_i'\beta} = \beta\theta_i \qquad (3.38)$$

3.6 Negative Binomial Distribution(Gamma- Poisson Mixture Distribution)

If a Poisson regression model doesn't fit the data and it appears that the variance of *y* is increasing faster than the Poisson model allows, then one way to handle this situation is to fit a parametric model that is more dispersed than the Poisson. AnaturalchoiceistheNegativeBinomialmodel. TheNegativeBinomial distributioncanbeobtainedfromthemixtureofPoissonandGammadistribution and is expressed as:

$$P(y_i/x_i) = \frac{\Gamma(y_i + \alpha^{-1})}{y_i!\Gamma(\alpha^{-1})}\left(\frac{1}{1+\alpha\theta_i}\right)^{\alpha^{-1}}\left(\frac{\alpha\theta_i}{1+\alpha\theta_i}\right) \qquad (3.39)$$

i i i

where y_i is the number of road accidents for a road segment *i*, θ_i is the expected number of accidents per period, which can be expressed as;

$$\theta_i = \exp(x_i \beta_0)$$

The mean and the variance of the Negative Binomial distribution are $E(y_i/x_i) = \theta_i$ and $Var(y_i/x_i) = \theta_i[1+\theta_i \alpha_i] > E(y_i/x_i)$.

The Negative Binomial distribution is also 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.

From the Negative Binomial distribution, α represents te dispersion parameter which allows or indicates the degree of over-dispersion. For instance, if $\alpha = 0$, the Negative Binomial regression model reduces to the traditional Poisson regression model. A lot of researchers in several different fields have considered both the Poisson and the Negative Binomial models:Miaou(1994), Berthanu (2004), Lord et al.,(2005), Karlaftis and Tarko (1998), Lee et al., (2004), Yau et al. (2004) and Byers et al,. to mention a few.

However when excess zeros occur, both the Poisson and the Negative Binomial regression models are not useful to fit the given data. Hence in this case, the Zero-inflated models and the Hurdle models are the appropriate choice.

3.6.1 Derivation of the Negative Binomial distribution

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

$$P(y_i/\theta_i) = \frac{e^{-\theta_i}\theta_i^{y_i}}{y_i!}$$

where y_i is the observed number of counts (accidents) for i = 0, 1, 2, 3,...,n and θ_i is the mean or the expected number of counts (accidents) of the Poisson distribution.

In the Poisson distribution, the θ_i was fully determined by a linear combination of x_s. The Negative Binomial essentially adds some unobserved heterogeneity, such that θ_i is determined by the x_s and some unobserved specific random effect (Gamma distributed error) ε_i which is assumed uncorrelated with the x_s.

Hence we obtain the following relationships:

$$\bar{\theta}_i = e^{x\beta_i + \varepsilon_i}$$

$$= e^{x\beta_i} e^{\varepsilon_i}$$

$$\bar{\theta}_i = \theta_i \delta_i \qquad (3.40)$$

where $\delta_i = e^{\varepsilon_i}$ which is defined as the unobservable random effect; $\theta_i = e^{x\beta_i}$ is the log-link between the Poisson mean and the covariates or independent variables (x_s); and the β_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, $E(\delta) = 1$ because this gives $E(\theta_i) = \theta_i$. This as a result indicates that we have the same expected count as the Poisson distribution.

Since the distribution of the observations is still Poisson with given x and δ, we obtain the relation;

$$P(y_i / x_i, \delta_i) = \frac{e^{-\bar{\theta}_i} \bar{\theta}_i^{y_i}}{y_i!} = \frac{e^{-\theta_i \delta_i} (\theta_i \delta_i)^{y_i}}{y_i!} \qquad (3.41)$$

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

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

Hence if *g* is the probability density function (pdf) of δ then;

$$P(y_i/x_i) = \int_0^\infty [P(y_i/x_i,\delta_i) \times g(\delta_i)]\, d\delta_i \qquad (3.42)$$

In order to solve equation 3.32, there is the need to specify the form of the pdf for δ. Most researchers assume that δ_i has a Gamma distribution with parameter υ_i.

Hence, we have;

$$g(\delta_i) = \frac{\upsilon_i^{\upsilon_i}}{\Gamma(\upsilon_i)} \delta_i^{\upsilon_i - 1} e^{-\delta_i \upsilon_i} \text{ for } \upsilon_i > 0 \qquad (3.43)$$

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

An interesting thing about this distribution is that, $E(\delta_i) = 1$ which is a convenient assumption indicating that the mean function of the Negative Binomial is the same as the Poisson. Also the variance is $Var(\delta_i) = \frac{1}{\upsilon_i}$. By using equation 3.31 and equation 3.33 to solve 3.31, we obtain the following Negative Binomial probability distribution:

$$P(y_i/x_i,\upsilon_i) = \frac{\Gamma(y_i + \upsilon_i)}{y_i!\,\Gamma(\upsilon_i)} \left(\frac{\upsilon_i}{\upsilon_i + \theta_i}\right)^{\upsilon_i} \left(\frac{\theta_i}{\upsilon_i + \theta_i}\right)^{y_i} \qquad (3.44)$$

The expected value od the above relation (NB) 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:

$$Var(y_i/x_i) = \theta_i(1+\theta_i) = e^{x_i\beta}(1+e^{x_i\beta_i}) \qquad (3.45)$$

υ_i υ_i

Because θ and υ are both positive, the variance will automatically exceed the conditional mean. This as a result increases the relative frequency of low and high counts.

We will have more parameters than observations since the variance remaine undefined if υ_i varies by observation. The typical assumption is that, υ_i is the same for the observation (counts) that is;

$$\upsilon_i = \alpha^{-1} \text{for} \alpha > 0 \tag{3.46}$$

This formulation as a result help to simplify some equations if we substitute equation 3.36 into 3.35 and 3.34 hence we obtain the Negative Binomial distribution as;

$$P(y_i/x_i,\alpha) = \frac{\Gamma(y_i+\alpha^{-1})}{y_i!\Gamma(\alpha^{-1})}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\theta_i}\right)^{\alpha^{-1}}\left(\frac{\theta_i}{\alpha^{-1}+\theta_i}\right)^{y_i} \quad (3.47)$$

and the variance as;

$$Var(y_i/x_i) = \theta_i\left(1+\frac{\theta_i}{\alpha^{-1}}\right) = \theta_i(1+\alpha\theta_i) = \theta_i+\alpha\theta_i^2 \text{ for } \alpha > 0 \quad (3.48)$$

3.6.2 Estimation of the Negative Binomial Distribution Function

The probability density function (pdf) is defined as:

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

The likelihood function for the NB function is:

$$L = \prod_{i=1}^{N} P(y_i/x_i) = \prod_{i=1}^{N} \frac{\Gamma(y_i+\alpha^{-1})}{y_i!\Gamma(\alpha^{-1})}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\theta_i}\right)^{\alpha^{-1}}\left(\frac{\theta_i}{\alpha^{-1}+\theta_i}\right)^{y_i} \quad (3.49)$$

$$= \prod_{i=1}^{N} \frac{\Gamma(y_i+\alpha^{-1})}{y_i!\Gamma(\alpha^{-1})}\left(\frac{\alpha^{-1}}{\alpha^{-1}+e^{x_i\beta}}\right)^{\alpha^{-1}}\left(\frac{e^{x_i\beta}}{\alpha^{-1}+e^{x_i\beta}}\right)^{y_i} \quad (3.50)$$

The next step consist of defining the log-likelihood of the Negative Binomial distribution:

$$\sum^{n} \quad {}^{-1} \quad {}^{-1} \quad {}^{\alpha^{-1}} \quad x_i\beta \quad {}^{y_i}$$

$$\ell = \sum_{i=1} \ln \frac{\Gamma(y_i + \alpha^{-1})}{y_i!\,\Gamma(\alpha^{-1})} + \ln \frac{\alpha^{-1}}{\alpha^{-1} + e^{x\beta_i}} + \ln \frac{e^{x\beta_i}}{\alpha^{-1} + e^{x\beta_i}} \tag{3.51}$$

But;

$$\ln \frac{\Gamma(y_i + \alpha^{-1})}{y_i!\,\Gamma(\alpha^{-1})} = \sum_{i=1}^{y-1} \ln(y_i + \alpha^{-1}) \tag{3.52}$$

By substituting equation 3.52 into 3.51, the likelihood can be computed using the following equation:

$$\ell = \sum_{i=1}^{n} \left\{ \sum_{i=1}^{y-1} \ln(y_i + \alpha^{-1}) - \ln y_i! + \alpha^{-1} \ln\left(\frac{\alpha^{-1}}{\alpha^{-1} + e^{x_i\beta}} \right) + y_i \ln\left(\frac{e^{x_i\beta}}{\alpha^{-1} + e^{x_i\beta}} \right) \right\} \tag{3.53}$$

$$\ell = \sum_{i=1}^{n} \left\{ \sum_{i=1}^{y-1} \ln(y_i + \alpha^{-1}) - \ln y_i! + \alpha^{-1} \ln(\alpha^{-1}) - \alpha^{-1} \ln(\alpha^{-1} + e^{x_i\beta}) + y_i x_i\beta - y_i \ln(\alpha^{-1} + e^{x_i\beta}) \right\} \tag{3.54}$$

Recall that $\theta_i = e^{x_i\beta}$

The values of α and β that maximize ℓ (log of the likelihood function) will be the maximum likelihood estimates we seek to obtain and the estimated variance- covariance matrix of the estimators is;

$$\Sigma = -H^{-1}$$

, where H is the Hessian matrix of second derivatives of the log-likelihood function. Then the variance-covariance matrix can be used to find the Wald confidence intervals and the p-values of the coefficient estimates.

3.6.3 Derivation of the Negative Binomial Mean Regression Model

For regression purposes, we typically assume that, y_i Negbin(θ ,α_i) and apply a log link, so that

$$\log \theta_i = \eta_i = x\, \beta^0{}_i$$

Hence from the General Linear Model framework, the Negative Binomial functions needs to be expresses as an exponential family in canonical form to obtain the following equations:

!

$$f(y;\, \theta,\varphi) = \exp \frac{y\theta^{\cdot} - b(\theta)}{\varphi} + c(y,\varphi)$$

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

$$= \frac{\Gamma(y_i+\alpha^{-1})}{y_i!\,\Gamma(\alpha^{-1})}\left(\frac{1}{1+\alpha\theta_i}\right)^{\alpha^{-1}}\left(\frac{\alpha\theta_i}{1+\theta_i}\right)^{y_i} \qquad (3.55)$$

by taking exponent of the above equation, we obtain;

$$= \exp\left\{\ln\left[\frac{\Gamma(y_i+\alpha^{-1})}{y_i!\,\Gamma(\alpha^{-1})}\left(\frac{1}{1+\alpha\theta_i}\right)^{\alpha^{-1}}\left(\frac{\alpha\theta_i}{1+\theta_i}\right)^{y_i}\right]\right\}$$

$$= \exp\left\{y_i \ln\left(\frac{\alpha\theta_i}{1+\alpha\theta_i}\right) - \alpha^{-1}\ln(1+\alpha\theta_i) + \ln\frac{\Gamma(y_i+\alpha^{-1})}{y_i!\,\Gamma(\alpha^{-1})}\right\} \qquad (3.56)$$

By comparing equation 3.46 to equation 3.5 we have;

$$\theta = \ln\left(\frac{\alpha\theta_i}{1+\alpha\theta_i}\right) \qquad (3.57)$$

where θ is the canonical parameter or link function and θ_i is the expected counts (accident).

But from the GLM framework, the link function is given by;

$$E(g(y)) = \theta = x\beta^{0}_{i} \tag{3.58}$$

Hence by substituting equation 3.47 into 3.48 we obtain;

$$\ln\left(\frac{\alpha\theta_i}{1+\alpha\theta_i}\right) = x\beta^{0}_{i} \tag{3.59}$$

By simplifying equation 3.49 and making θ_i the subject we have the equation:

$$\theta_i = \frac{e^{x\beta^{0}_{i}}}{\alpha(1-e^{x\beta^{0}_{i}})} \tag{3.60}$$

3.6.4 Assumptions of the Negative Binomial regression model

The assumptions of the Negative Binomial regression includes;

i. The expected value of the exponential error is equal to 1 ($e^{\varepsilon} = 1$) ii. The exponential error is Gamma distributed

3.7 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.

TheConway-Maxwell-Poisson(CMP)hasaprobabilitymassfunctionoftheform;

$$P(Y = y_i) = \frac{\theta_i^{y_i}}{(y_i!)^{\varphi}} \frac{1}{Z(\theta_i, \varphi)} \quad (3.61)$$

where

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

for $\theta > 0$ and $\varphi \geq 0$. The function $Z(\theta_i, \varphi)$ is an infinite series that converges for $\theta > 0$ and $\varphi > 0$.

The Conway-Maxwell-Poisson distribution shows that there is a non-linear relationship between the ratio of successive probabilities as displayed by

$$\frac{P(Y = y_i)}{P(Y = y_i - 1)} = \frac{\theta_i}{y_i^{\varphi}} \quad (3.62)$$

The mean and the variance of the Conway-Maxwell-Poisson distribution can be represented as:

$$E[Y] = \theta_i \frac{\partial \log(Z(\theta, \varphi_i))}{\partial \theta_i} \quad (3.63)$$

$$Var(Y) = \theta_i \frac{\partial E[Y]}{\partial \theta_i} \quad (3.64)$$

Seller and Schumeli (2008) showed that the variance can also be specified as

$$Var(Y) = \frac{\partial E[Y]}{\partial \log \theta_i} \quad (3.65)$$

Since the moments of the distribution do not have a closed form, Schumeli et al. (2005) showed that (3.62) can be approximated by

$$E[Y] \approx \theta^{\frac{1}{\varphi}} - \frac{\varphi - 1}{2\varphi} \quad (3.66)$$

and Sellers and Shumeli show (3.63) can also be approximated by

$$Var(Y) \approx \frac{E[Y]}{\varphi} \quad (3.67)$$

where φ measures the dispersion of the mean.

3.7.1 Conway-Maxwell-Poisson Generalized Linear Model

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

$$P(Y = y_i) = e^{y \log_i (\theta_i) - \log(Z(\theta, \varphi_i)) - \varphi \log(y_i!)} \quad (3.68)$$

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 given as

$$\sum^{n}$$

$$\ell(\beta/\theta, \varphi_i) = \sum_{i=1} \{y \log_i (\theta_i) - \log Z(\theta, \varphi_i) - \varphi \log(y_i!)\} \quad (3.69)$$

The log-likelihood function of the CMP distribution can be computed in a few different ways. In case the standard Newton-Ralphson algorithm is employed, the log of the likelihood function must be maximized under the constraint $\varphi \geq 0$ (Sellers and Schmueli, 2008).

3.7.2 Parameter Estimation of the CMP Model

There are few methods of estimation the parameters of the Conway-Maxwell- Poisson distribution from a given count data. Two of these few methods includes;

i. Weighted Least Square

i. Maximum Likelihood Estimation

The Weighted Least Square approach is simple and precise but on the other hand lacks precision. Maximum Likelihood on the other hand is precise but more complex and computationally intensive.

Weighted Least Square

The weighted least square provides simple, efficient method to derive rough estimates of the parameters of the CMP distribution and determines if the distribution would be an appropriate model.

Following the use of this method, an alternative method should be employed to compute more accurate estimates of the parameters if the model is deeemed appropriate.

This method uses the relationship of successive probabilities from eq (3.62). By taking logarithm of both sides of eq(3.62), the following relationship arises as;

$$\log \frac{P_{y-1}}{P_y} = -\log \theta + \varphi \log y \quad (3.70)$$

where P_y denotes $P(Y = y)$ and P_{y-1} also denotes $p(Y = y-1)$. When estimating the parameters, the probabilities can be replaced by the relative frequencies of y and $y - 1$.

If the data appears to be linear, then the model is likely to be a good fit. Once the appropriateness of the model is determined, the parameters can be estimated

by fitting a regression of $\frac{\hat{P}_y}{\hat{P}_{y-1}}$ on log y.

However, the basic assumption of Homoscedasticity is violated, hence weighted lest square regression must be employed. The inverse weight matrix will have the variance of each ratio on the diagonal with one-step covariance on the first diagonal, both given below;

$$V\left[\log\left(\frac{\hat{P}_{y-1}}{\hat{P}_y}\right)\right] \approx \frac{1}{nP_y} + \frac{1}{nP_{y-1}} \tag{3.71}$$

$$Cov\left(\log\frac{\hat{P}_{xy}}{\hat{P}_y}, \log\frac{\hat{P}_x}{\hat{P}_{x+1}}\right) \approx -\frac{1}{nP_y} \tag{3.72}$$

Maximum Likelihood Estimate of the CMP Distribution

The Conway-Maxwell-Poisson likelihood function is;

$$L(\theta,\varphi / y_1 \ldots y_n) = \theta^{S_1} \exp(-\varphi S_2) Z^{-n}(\theta,\varphi) \tag{3.73}$$

where $S_1 = \sum_{i=1}^{n} y_i$ and $S_2 = \sum_{i=1}^{n} \log y_i!$

Maximizing the likelihood function yields the following two equations

$$E[y] = \bar{x} \tag{3.74}$$

$$E[\log y!] = \overline{\log y!} \tag{3.75}$$

which do not have an analytic solution. Hence, the ML estimates are approximated numerically by the Newton Raphson Iteration Method.

In each iteration, the expectations, variances and covariances of y and log y! are approximated by using the estimates of θ and φ from the previous iteration in the expression;

$$E[f(y)] = \sum_{k=o}^{\infty} f(k) \frac{\theta^k}{(k!)^{\varphi} Z(\theta,\varphi)} \tag{3.76}$$

where $Z(\theta,\varphi) = \sum_{k=0}^{\infty} \frac{\theta^k}{(k!)^{\varphi}}$

This is continued until converges to $\hat{\theta}$ and $\hat{\varphi}$

3.8 Selecting Over-dispersed Models

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

$$H_o : \alpha = 0$$

$$H_a : \alpha = 06$$

The corresponding test statistic is;

$$z = \frac{\hat{\alpha}}{SE(\hat{\alpha})} \tag{3.77}$$

Under H_o and for large sample size, the z has an approximate standard normal distribution. We therefore reject H_o 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 and the quasi-Poisson models are more appropriate compare to the Poisson.

3.9 Model Specification

Models considered in this thesis has the mean number of persons killed in an accident with a particular period as a function of the collision type, driver error,type of vehicle, weather condition and light condition. Each model parameterizes

$$\theta_i = \exp(\beta_o + \beta\, CTYPE_1 + \beta\, DERR_2 + \beta\, TYPV_3 + \beta\, LCOND_4 + \beta\, WCOND_5) \tag{3.78}$$

where i = 1,....,n individuals, CTYPE denotes the collision type, DERR represents the driver error,TYPV is the type of vehicle, LCOND as the light condition and WCOND also representing the weather condition.

3.10 Model Selection Method

Model selection methods like Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Deviance Goodness of fit are commonly used when it comes to the comparison especially among non-nested models. All the models discussed in this research work are categorized as non-nested models.

3.10.1 Goodness of Fit

After fitting the models to the data obtained, it very necessary or important to check the overall fit as well as the quality of the fit. The quality of the fit between the observed values (y) and the predicted values ($\hat{\theta}$) can be measured by various test statistics, but one useful statistics is called deviance and is defined as:

$$D(y : \hat{\theta}) = -2 \sum_{i=1}^{n} \left[y_i \ln \left(\frac{y_i}{\hat{\theta}_i} \right) - (y_i - \hat{\theta}_i) \right] \quad (3.79)$$

where y_i is the number of events (observed values or counts), *n* is the number of observations and $\hat{\theta}_i$ represents the fitted means of the models involved in the research (predicted values). For a better model, one must expect a smaller value of the deviance ($D(y : \hat{\theta})$). Hence the smaller the value of the deviance of a specific model the better the model or the more statistically significant the model becomes.

3.10.2 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 forthedata, AICestimatesthequalityofeachmodel, relativetotheothermodels.

Hence AIC provides a means for model selection. For any statistical model, the AIC value is;

$$AIC = -2L + 2k \tag{3.80}$$

where L is the maximized value of the likelihood function and k is the number of parameters in the model. For the best model one must expect the lowest AIC value.

3.10.3 Bayesian Information Criterion

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 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 observations 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 $\hat{L} = p(x/\theta,\hat{M})$ where $\hat{\theta}$ are the parameter values that maximizes the likelihood function.

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

$$-2\ln p(x/M) \approx BIC = -2\ln \hat{L} + k(\ln(n) - \ln(2\pi)) \qquad (3.81)$$

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

3.11 Parameter Estimation

The maximum likelihood estimation method has been considered to since all the models involved in the study or the research work can be estimated using Maximum Likelihood Estimation (MLE). To evaluate the models involved, it is necessary to examine the significance of the variables included in the models. For a better model, the estimated regression coefficients have to be statistically significant. Normally, the t-test used to determine the significance of the regression coefficients.

3.12 Statistical Package for Analysis

The preliminary analysis of the data set will be done with the help of SPSS 20 and Micro-soft excel.

The R statistical package on the other hand will be of need in the further analysis since it uses the concept of maximum likelihood estimation in analyzing General Linear Models. The R statistical package applies a built in function *glm()* to model Poisson regression (Chambers and Hastie, 1992) from the stats package,*glmnb()* for Negative Binomial from the MASS package,*cmp()* for Conway-Maxwell-Poisson model from the COMPoissonReg package.

CHAPTER 4

DATA ANALYSIS AND RESULTS

4.1 Introduction

This chapter features the preliminary analysis on the number of persons killed as a result of road accidents in Ghana and the further analysis of the data using statistical tools such as Poisson regression, Negative Binomial(Poisson-Gamma mixture) regression and Conway-Maxwell-Poisson regression model.

4.2 Data Source

This research work mainly utilized secondary data obtained from the Building and Road Research Institute (BRRI) of the Council for 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 2009-2013. The number of persons killed in individual road accidents for the five years period was used as the response or dependent variable and the other categorical variables such as collision type, weather condition, light condition, and driver error as the explanatory or independent variables

4.3 Preliminary Analysis

4.3.1 Number of Persons Killed by Road Accidents in Ghana

There were 79,114 road accidents in Ghana from 2009-2013 which killed 10,836 people. This as a result gives an indication that on the average, 15,823 road accidents occurred every year and 2,167 lives were lost through accidents.

4.3.2 Distribution of Person Killed by Road Accidents Annually

The table 4.1 below indicates the time in years for which accident that resulted in the death of people occurred. This table additionally presents the total number of persons killed in road accident yearly from 2009-2013.

Table 4.1: Total number of persons killed in Road Accidents in Ghana from 2009-2013

YEAR	Number of Persons Killed
2009	2237
2010	1986
2011	2199
2012	2165
2013	2249

The most important feature of table 4.1 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 luckily reduced to 1,986 in 2010 and in the year 2011 the figure rose to 2,199. By the year 2013, the number of people who killed by road accidents had risen to 2,249. This statistics for the five years period support what Odoom (2010) stated in his research work on road accidents that as years go by, the number of motor vehicles in the Ghana will be increasing with increasing number of road accidents and as such the number of people who are likely to die in road accidents will as a result also increase.

4.4 Further Analysis

Further analysis was conducted out of the research work to determine the appropriate count regression model the statistically and adequately fits accident data in Ghana and to use the appropriate model to investigate the key predictors of road accidents with emphasis on the number of persons killed in the accident.

However, the problem of over-dispersion was encountered, as a result the Negative Binomial and the Conway-Maxwell-Poisson were applied. These models were based on a 5% level of statistical significance. The further analysis in this research work was performed with the help of R statistical package. Specifically, the Poisson, Negative Binomial and the Conway-Maxwell-Poisson count regression models were executed computational using the R functions such as, glm(), glm.nb() and cmp() respectively.

4.4.1 Poisson Regression model for the number of persons killed in Road Accidents from 2009-2013 in Ghana

Table 4.2 below reveals results 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 od 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 p-values of different collision types, types of vehicle, weather conditions, light conditions and driver errors.

Table 4.2: Parameter Estimates of Poisson Count Regression model for the number of persons killed by Road Accidents from 2009-2013 in Ghana

VARIABLE	ESTIMATES	STD ERROR	RATE P	-VALUE

The parameter estimates were interpreted in terms of the rate of the number of persons killed or death 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 wit p-value of 7.85e- 06 with an estimated death rate of exp(1.5047)=4.50 was statistically different from zero at 0.05 level of significance.

Inthismodel, overturn, Inexperience, car, clearanddaywereusedasthereference level for the categorical variables collision type,Driver error, Type of vehicle, Weather condition and Light condition. With respect to the variable "Collision type" from table 4.2, rear end collision, side swipe collision, Right angle collision, , hit object on road, Hit object off road and Hit parked vehicle/pedestrian/animal withp-values0.09702, 0.91740, 0.20404, 0.85036, 0.29680and0.92492respectively were not significantly associated with the number of persons killed in road accidents in Ghana.

However, onlytwoCollisiontypesincludingRano ff collisionandHead-oncollision with p-values 0.03514 and 0.00158 respectively were statistically significantly associated with the number of persons killed in road accidents. The parameter estimates for 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 suggest that, the rate of death in road accidents was 1.33 times higher among death caused ran-off collision compare to overturn collision whilst that of Head-on on the other hand indicates that, the rate of death in road accidents was 2.54 times higher among death caused by Head-on collision compare to deaths caused by overturn collision.

Also, inattentive, Improper overtaking and Loss of control with their respective p-values 0.00874, 0.04291 and 0.02658 were the only dimensions of the variable

"Driver error" which were significantly associated with the number of persons killed by road accidents. The estimated rate of death for "Inattentive" table 4.2 was 0.45 which indicates that, the rate of death in road accidents was 0.45 times higher with Inattentiveness compared to 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 was 0.43 indicating that the rate of number of persons killed in road accidents is 0.43 times higher with Loss of control compared to Inexperience. The remaining Driver errors including Too fast,

No signal, Too close and Fatigue/asleep were not statistically and 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 figure, 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 has a p-value (0.90341) greater than the

5% level of significance. Thus, Fog/midst and Rain with 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 from Fig:4.1 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 clear (good weather) and rate of death in road accidents is 0.65 times higher with Rain compared to clear weather condition.

Lastly, from table 4.2 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% 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 higher with "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 'Night with street light off" compared to 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:

$$\theta_i = \exp(1.50417+0.28387RO+0.93386HO-0.80611IN+0.22388IO$$
$$-0.84303LC-0.55358GV+.30358BM-0.30015F-0.42441R+1.30287NS+0.61502NO) \quad (4.1)$$

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 Rains, NS is Night with no street light and NO represents Nights with street lights off.

4.4.2 Goodness of Fit test of the Poisson Regression for the number of persons killed in Road Accidents

The table 4.3 below displays the goodness of fit test of the fitted mean Poisson regression model for the number of person killed in road accidents in Ghana. This goodness of fit table helps one to determine how quality fit or appropriate the fitted model is as compared to other models.

Table 4.3: Goodness of fit test of the Poisson regression for the number of persons killed in Road Accidents in Ghana

CRITERION	VALUE	DF
Null Deviance	2493.63	65
Residual Deviance	470.85	54
AIC	1059.31	
BIC	1070.65	
Dispersion Parameter	8.71944	

The table 4.3 reveals that the Akaike's Information Criterion (AIC) was 1059.31, Bayesian Information Criterion (BIC) was also 1070.65 and the Residual deviance with 54 degree of freedom was 470.85. This value of the 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 is equidispersion (when dispersion parameter=1) and hence the Negative Binomial and the Conway Maxwell Poisson model is preferred.

4.4.3 Negative Binomial regression for the number of Persons killed in Road Accidents from 2009-2013 in Ghana

The table 4.4 below also shows the results of the Negative Binomial regression modelwithcollisiontype, drivererror, typeofvehicle, lightconditionandweather condition as the independent categorical variables and the number of persons killed as the dependent variable.

Table 4.4: Parameter Estimates of Negative Binomial Count Regression model for the number of persons killed by Road Accidents from 2009-2013 in Ghana

VARIABLE	ESTIMATES	STD ERROR	RATE P	-VALUE

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 types, drivererror, typeofvehicle, lightconditionandweatherconditionremained unaltered. However, the standard errors and the respective probability values changed from one variable to the other.

From the table 4.4, 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" the head on collision was the only type of collision that was significantly associated with the number of persons killed in road accidents. The parameter estimates for the 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 was 2.53 times higher among accidents as result of head-on collision compared to overturn collision.

However, Improper-overtaking and loss of control under the categorical variable

"Drivererror"withrespectivep-values0.02638and0.00303remainedsignificantly associated with the number of persons killed in road accidents in Ghana at 5% level of significance with death rate of 1.25 and 0.43 respectively.This therefore implies that, the rate of death in road accidents was 1.25 times higher among accidents caused by improper overtaking compare 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 drive losing control compare inexperience.

On the other hand, a driver driving too fast and 'No signal" from the table 4.4 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% 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 cars. Inadditiontothevariable"weathercondition"Rainandotherweatherconditions

with p-values 0.28574 and 0.4261 respectively were not statistically significantly associated with the number of persons killed in road accidents . However, only

Fog/midst with probability value of 0.01425 remained significantly associated with the number of persons killed in road accidents at 5% level of significance. The parameter estimate for this variable was -0.30015 and death rate value of 0.74. This value as a result reveals that, the rate of death is 0.74 times higher with accidents that occurs in a foggy/midst weather condition compared to a clear weather condition.

Notforgettingthecategoricalvariable"lightcondition", Nightwithstreetlighto ff 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 was 1.84 times higher with accidents that occur on streets with lights off during the night compared to accidents that occur during the day. ThemeanoftheNegativeBinomialmodelforestimatingtheexpectedorthemean number of persons killed in road accidents per year with significant variables is therefore formulated as:

$$\theta_i = \exp(1.50417+0.93386HO+0.22388I0-0.84303LC+0.30358BM-0.30015F+0.61502) \quad (4.2)$$

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.

4.4.4 Goodness of Fit test of the Negative Binomial regression for the Number of Persons killed in Road Accident

The table 4.5 below depicts the goodness of fit test of the Negative binomial count model for the number of persons killed in road accidents in Ghana from 2009-2013. This table gives the details of the null deviance, residual deviance, AIC, BIC and the dispersion parameter of the model.

Table 4.5: Goodness of fit test of the Poisson regression for the number of persons killed in Road Accidents in Ghana

CRITERION	VALUE	DF
Null deviance	985.72	65
Residual devaince	65.666	54
AIC	749.10	
BIC	758.46	
Dispersion parameter	1.21603	

From the table, the residual deviance of the model decreased substantially 65.666, which is much smaller compare 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.

4.4.5 Conway-Maxwell-Poisson regression model for the number of persons killed in Road Accidents in Ghana from 2009-2013

The parameter estimates of the Conway-Maxwell-Poisson model on the other hand as compared to that of the Negative binomial and the Poisson regression models 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 4.6 below depicts that the parameter estimates of the intercept of 1.50417 with a probability value 5.42e-06 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 varible "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 0.036918 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 rate of death for inattentiveness from the table 4.6 was 0.45 which suggests that the rate of death in road accidents was 0.45 higher among accidents caused by drivers inattentiveness compared to drivers inexperience. The estimated rate of deathforimproperovertakingontheotherhandwas1.25whichasresultindicates that, the rate of death in road accidents in this country was 1.25 times higher with accidents caused as a result of improper over taking compared to inexperience of some drivers.

In addition, from the table 4.6, 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 was 1.36 times higher with Bus/Minibus compared to cars (taxis, 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.

Table 4.6: Parameter Estimates of Conway-Maxwell-Poisson Count Regression model for the number of persons killed by Road Accidents from 2009-2013 in Ghana

VARIABLE	ESTIMATES	STD ERROR	RATE	P-VALUE

Thus rain in this case with a p-value of 0.0021015 was significantly related to the number of persons killed in road accidents at 5% 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 was 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 fig:4.3 had a p-value of 0.0039922 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 number of people dying in an accident that occurs 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:

$$\theta_i = \exp(1.50417+0.93386HO-2.80611I+0.22388IO+0.30358BM-0.42441R+0.61502) \tag{4.3}$$

4.4.6 Goodness of Fit test of the Conway-Maxwell-Poisson regression for the Number of Persons killed in Road Accident

The table 4.7 below indicates the goodness of fit test of the Conway-Maxwell- Poisson regression model for the number of persons killed in road accidents from 2009-20131. This table helps to determine as to whether the fitted model is appropriate or whether the model is quality fit

Table 4.7: Goodness of fit test of the Conway-Maxwell-Poisson regression for the number of persons killed in Road Accidents in Ghana

CRITERION	VALUE	DF
Null deviance	1159.04	65
Residual devaince	76.28	54
AIC	826.46	

BIC	835.79	
Disoersion parameter	1.41249	

The table 4.7 above reveals that, in this model, the residual deviance, AIC and BICallhavedecreasedsubstantiallywiththeirrespectivevalues76.28, 826.46and 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 ia also an appropriate model.

4.4.7 Parametric comparison between Poisson, Negative Binomial and Conway-Maxwell-Poisson Regression models for Goodness of Fit Test

Table 4.8: Parametric comparison between Poisson, Negative Binomial and Conway-Maxwell-Poisson Regression models for Goodness of Fit Test

Characteristics	Poisson model	Negative Binomial model	CMP model
Null deviance	2493.63	985.725	1159.04
Degrees of freedom	65	65	65
Residual deviance	470.85	65.666	76.28
Degrees of freedom	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

Looking at the results presented in the table 4.8 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 4.8,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. Although both the Negative Binomial and the CMP model takes care of overdispersion, the dispersion parameter value from the table 4.8 indicates that the Negative binomial regression takes care of overdispersion more than that of the CMP model.

The above statistics therefore gives an indication that, the model that best fits accidentdatawithnumberofpersonskilledintheGhanaistheNegativeBinomial 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:

$$\theta_i = \exp(1.50417+0.93386HO+0.22388I0-0.84303LC+0.30358BM-0.30015F+0.61502)$$

where θ_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.

CHAPTER 5

CONCLUSION AND RECOMMENDATION

5.1 Introduction

This section of the research work focuses on the various conclusions drawn from the findings of the analysis conducted in the previous chapter and suggested alternatives and recommendation to aid in reducing the number of persons killed by road accidents in Ghana.

5.2 Conclusion

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 collisionascollisiontype, ImproperovertakingandLossofcontrolasdrivererrors, Bus/Minibus as type of vehicle, Fog/midst as weather condition and Night with streetlightsoffasLightconditionwerethekeypredictorsorindependentvariables 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 regressionmodelduetoitsmajorassumptionofequi-dispersion. ThusthePoisson count regression model serves well under equi-dispersion condition whiles both Negative binomial and CMP count regression models serve better while the data is over-dispersed.

5.3 Recommendations

With regards to the state of road accidents and the number of persons killed by road accidents in Ghana it is first and foremost recommended that education on road accidents should be intensified especially among drivers since driver errors such as Improper overtaking and loss of control has a high effect on the number of persons killed by road accidents.

Also, since the vehicle type involved in accidents has effect on the number of personsexpectedtobekilled, driversofvehicletypesmostespeciallyBus/Minibus drivers should be given sufficient training to be able to avoid accidents. The study revealed that with respect to the variable weather condition, Rain was contributing significantly to the expected number of person to be killed via road accidents hence the researcher recommends that drivers should drive cautiously when raining in order to prevent accident occurring.

Finally driving at night with street lights off was found to have a high effect on the expected number of persons to be killed via road accidents hence the researcher additionally advices drivers to drive vigilantly at night in areas with street lights not functioning.

REFERENCES

Abdel-Aty, M.A., Radwan, A.E. (2000) Modelling traffic accident occurrence and involvement.

Accident Analysis and Prevention, Vol. 32, pp. 633-642.

Abdel-Aty, M.A., Radwan, A.E. (2000) Modelling traffic accident occurrence and involvement. Accident Analysis and Prevention, Vol. 32, pp. 633-642. Eenink el al. (2007) Accident Analysis and Prevention, Vol. 39, pp. 657-670.

Ackaah A.S., (2011). Modelling traffic crashes on rural highways in Ashanti Region, Journal of Safety Research, 34 (5), 597-603.

Adams J.G.V., (1974)....... and how much for your grandmother? Environment and Planning A, (6), pp 619-626.

Afukaar F.K., Agyemang W, and Most I. (2009) Accident statistics 2007. Building and Road Research Institution council for scientific and industrial Research, Kumasi.

Afukaar F.K., Agyemang W., and Ackaar W. (2006) Estimation of levels of under –reporting of road traffic accident in Ghana. Ministry of transportation: Nation Accident management project, report. Final Report.

Afukaar F.K., and Debrah E.K., (2007). Accident prediction for signalized intersections in Ghana. Ministry of transportation: Nation Accident management project, report. Final report

Agresti A. (1996). An introduction to categorical data analysis. 2 nd Edition, John Wiley and Sons, Inc., New York.

Akongbota J., (2011). Reducing accidents on our roads. Retrieved December 21, 2014,from:http://www.ghana.gov.gh/index.php?option=com_content&view=article&id=5864:reducingaccidents- on-our-roads&catid=24: features&Itemid=167.

Allan T, Weston D., Katala J., Mallya J., and Mwangos N., (1995). Road maintenance management systems-implementation of the Road Mentor 4 system in the central zone. In Proc; Annual Roads Covention (ARC 1995). TANROADS, Dares Salaam, Tanzania.

Andreassen D., (1985). Linking deaths with vehicles and population. Traffic Engineering and Control. 26:547-549.

Beaumont K. and R.F. Newby., (1972), Traffic Law and road safety research in United Kingdom. British countermeasures. National Road Safety Symposium, Canberra.

Birch M., (1963). Some major problems in the field of urban traffic and transport in the city of Surabaya. Workshop on urban traffic and Transportation, ECAFE 1963.

Breslow S., Di Silvestro G., Persaud B., Begum M.A., (2010). Revising the variability of the Dispersion Parameter of Safety Performance Functions using Data for Two-Lane Rural Roads. Paper 10-35722, 89th Annual Meeting of the Transportation Research Board, Washington D.C.

Broughton J. (2007). Casualty rate by type of car. TRL Report No. PPR 203. Crowthrone: TRL Limited.

Caliendo, C., Guida, M., Parisi, A. (2007) A crash-prediction model for multilane roads.

Cameron A.C., and Trivedi P.K. (1998). Regression Analysis of Count Data, Cambridge University Press, Cambridge, U.K.

Carson J. (1998). The effect of ice warning signs on ice accident frequencies and severities; An investigation using advanced econometric modelling, Ph.D. Dissertation Seattle, WA: University of Washington.

Clarke D., Ward P., Truman W., and Bartle C. (2007). Fatal vehicle-occupant collisions: An indepth study. Road Safety Report No. 75. London: Department for Transport.

Codlingp J., (1972), Weather and road accidents. Paper presented to the symposium in climatic resources and economic activity, University College of Wales.

Conway, R.W. and W.L. Maxwell (1962), „A queuing model with state dependent service.

Cox D.R. (1983). Some remarks on Over-dispersion, Biometrika, 70, 269-274.

Dawson F.F., (1967). Cost of road accidents in Great Britain. Ministry of Transport, RRL Report LR 79. Crowthrone, (Road Research Institute)

Drew G.C., (1963). The study of accidents. Bull Brit.Psychol. Soc., 16, pp.1-10.

EI-Basyouny, K., Sayed, T. (2006) Comparison of Two Negative Binomial Regression.

Fletcher J.P., Baguley C.J., sexton B., and Done S. (2006) Road accident modeling for highway development counties. Main Report Trials in India and Tanzania. Report No PPRO95.

Garber C.J. and M. Wu (2001). A method to cope with the random errors of observed accident rates in regression analysis. Accident Analysis and Prevention Vol. 21, No. 4, 317-332.

Gardener B.L., Baguley C.J., and Kirk S.J., (1995). Cost and Safety Efficient Design Study of Rural Roads in Developing countries. Report PR/INT/242/1995. Transport Research Laboratory, Crowthrone.

Garwood Dr. F and J.M Munden, (1968). Variations in the pattern of road accident rates in Great Britain over the last ten years. OTA 9th International Study Week in Traffic and Safety Engineering Documentation.

Ghanaian Times (2011). Road accidents cost in Ghana millions of dollars. Retrieved November 2014 from the website: http://ghanaweb.com/GhanaHomePage/NewsArchive/artikel.php.

Godwin J.O., (1973). Comparison of the pattern of accident rates on roads of different countries, Traffic Engineering and Control, pp.432-34.

Green, William H (1997), Econometric Analysis, 3rd Edition, Prentice Hall.

Heidi W. (2006). Road accidents increase dramatically worldwide. On the web, http://www.prb.org.

Herrikson B.K., (2001), Roadway traffic crash mapping; A space-time modelling approach. Journal of Transportation and Statistics 6(1), 33-57.

Home J.A., Zomer J., and Lavie P. (1990). Sleep related automobile accidents-when and who? Horne J.A. ed. Sleep 90. Buchum: Pontenagel press. 448-51

Homes N. and A.A. Reyner, (1995). Exploring the relationship between development and road traffic injuries; a case study in India, European Journal of Public Health, Vol. 16, No. 5, pp. 487-491.

Howe J.D.G.F., (1972). A review of rural traffic counting methods in developing countries. Department of Environment TRRL Report LR 427, Crowthrone (Transport and Road Research Laboratory)

Jacobs G D and P Hutchinso' N., (1973). A study of accident rates in developing countries. Department of, the Environment-, TRRL Re-port LR 546, Crowthrone (Transport and, Road Research, Laboratory).

Jacobs G.D., and Cutting C.A. (2000). Further research on accident rates in developing countries. Accident Analysis and Prevention. 18:119-127.

Jamal R.M.A., and Jamil A.N. (2001). Casual models for road accident fatalities in Yemen. Accident Analysis and Prevention. 33:547-561.

Joshua Sarath C. and Nicholas J. Garber (1990), Estimating truck accident rate and involvements using linear and Poisson regression models. Transportation Planning Technology, Vol. 15, 41-58.

Jovanis, P.P., Chang, H. (1986) Modelling the relationship of accidents to miles travelled. Transportation Research Record, Vol. 1068, pp. 42-51.

Kemp R.N., (1972). Preliminary report on an on-the –spot survey of accidents. Department of the Environmental TRRL, Report LR 434, Crowthrone (Transport and Road Research Laboratory)

Kim S.H., Chung S.B., and Song K.S. (2005). Development of an accident prevention model using GLM (generalized log -linear model) and EB Method a case of seal. Journal of the Eastern Asia society for Transportation studies 6:3669-36682

Kuma A.M.D., Sanjeau L.M.D., Agawam D.M.C.H., Kava R.M.D., and Dora T.D.M.D, (2008). Fatal road accidents and their relationship with head injuries. An epidemiological survey of five years. Indian Journal of Neurotrauma (IJNT) 5(2):63-67.

Lambert Diane (1992), Zero-inflated Poisson regression with an application to defects in manufacturing. Technometrics, Vol. 34, No. 1, 1-14.

Liveneh J.A. and J.G. Hakkert (1972), Effect of gradient and curvature on accidents on London-Birmingham motorway. Traffic Engineering and Control 1972, 7, (10) 617-21.

Lord, D., Manar, A., Vizioli, A. (2005) Modelling crash-flow-density and crash-flow-V/C ratio relationships for rural and urban freeway segments. Accident Analysis and Prevention, Vol. 37, pp. 185-199.

Lord, D., Persaud, B.N. (2000) Accident prediction models with and without trend: application of the generalized estimating equations procedure. Transportation Research.

Mackay G.M., C.P de Fonseka, I. Blair and A.B Clayton, (1969). Causes and effects of road accidents. Univ. of Brim. Dept. of Transportation and Environmental Planning. Publication No.33.

Maddala T.M., (1977): An investigation into the effect of road layout on accidents in rural bucks. Department of Scientific and Industrial Research, Road Research Laboratory, Research Note No.RN/2340. Hardsworth.

Miaou S.P., and Lum H. (1993). Modeling vehicle accidents and highway geometric design relationships. Accident Analysis and Prevention. 25:689.

Miaou, S. (1994). The relationship between truck accidents and geometric design of road.

Milton J., Mannering F., (1998). Highway accident severities and the mixed logit model; An exploratory empirical analysis. Accident Analysis and Prevention. 40(1), 260-266.

Minter A.L. (1987) Road casualties improvement by learning processes. Traffic Engineering and Control. 28:74-79.

Mishan F J., (1971). Cost benefit analysis. London (Allen & Unwin) Chap. 22 and 23.

Montella, A. (2008) Crash prediction models for rural motorways. Transportation Research Record, Vol. 1717, pp. 102-108.

Mullahey, John (1986), Specification and testing of some modified count data models, Journal of Econometrics, Vol. 33, 314-365.

National Road Safety Commission (2009). Annual Report. National Road Safety Commission, Ghana.

Nelder, J. A. and R.W.M Wedderburn (1972), „Generalized linear models", Journal of Royal Statistical Society, Ser. A, 135, 370-384.

Okyere J.N. (2006). Research into road safety problems in Ghana. Building and Rural Research Institute, Kumasi, Ghana.

Oppong, Richard Asumadu (2012), Statistical Analysis of Road Accidents Fatality in Ghana using Poisson Regression, Mphil. Thesis, Kwame Nkrumah University of Science and Technology.

Pemmanaboina, R. (2005) Assessing occurrence on urban freeways using static and dynamic factors by applying a system of interrelated equations. M.E. thesis, University of Central Florida.

Persaud, B., and Dzbik, L. (1993) Accident prediction models for freeways. Transportation Research Record, Vol. 1401, pp. 55-60.

Poch M., and Mannering F. (1996). Negative Binomial analysis of intersection accident frequency. Journal of Transportation Engineering, 122, 105-113

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.

Redmond A.D. (1994). Letters. British Medical Journal. 309:57 Record, Vol. 2083, pp. 180-189.

Reynolds J., (1956). The cost of road accidents. Journal of Royal Statistical Society, Series A, Vol 119, Part IV.

Salgado M.J., Colombaje Y.C., (1998). A bivariate Negative Binomial model to explain traffic accident migration. Accident Analysis and Prevention 22(5), 487-498.

Salifu M., Mosi I., Addae –Boafah K., and Sarpong K. (2004). Accident holistic 2005. Building and road institute. Council for scientific and industial Research, Kumasi.

Salim T.B., and Salima C. (2005). Rule mining classification of road traffic accidents using adaptive regression trees. International Journal of Simulations Systems. Science and Technology Special Issue on Soft Computing of Modelling and Simulation, 6, 10-11.

Salt G.F. and W.S. Szatkowski., (1973). A guide to levels of skidding resistance for roads. Department of Environment TRRL Report LR 510, Crowthrone, (Transport and Road Research Laboratory)

Sargberg 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.

Sellers, K.F and G. Schumeli (2008), „A Flexible Regression Model for Count Data", University of Maryland School of Business Working Paper.

Shadev C.E. (1994). The relationship between truck accidents and highway geometric designs of road sections; Poisson versus Negative Binomial regressions. Accident Analysis and Prevention.

Shankar, V., Mannering, F., Barfield, W. (1995) Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. Accident Analysis and Prevention, Vol. 27, pp. 371-389.

Shanker, Venkataraman, Fred L. Mannering and W. Barfield (1995). Effect of roadway geometrics and environmental factors on rural accident frequencies, Accident Analysis and Prevention, Vol. 27. No. 3, 371-389.

Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S., and P Boatwright (2005), „A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution", Applied Statistics, 54, 127-142.

Simon G.D., (1961). Pedestrian behaviour before and after the installation of a zebra crossing (1) Yiewstey High Street (A408). Department of Scientific and Industrial Research. Laboratory Note No. LN/654/GDJ. Harmondsworth 1961 (Road Research Laboratory).

Smeed J., (1968). Variations in the pattern of accident rates in different countries and their causes. Traff. Engng & Control, 10(7). 364-371.

Starks H.J.H. and R.G Knight., (1967). Road accident investigation. The large-scale approach. Detailed Studies. Police Journal, 40(3), 115-26.

Susan C., and Partyka (1984). Simple model of fatality trends using employment and population data. Accident Analysis and Prevention. 16:21-222.

Tillman W.A. and G.E. Hobbs., (1949). Accident-prone automobile driver; study of psychiatric and social background. Amer. J. Psychiat., 106, pp.321-31.

Travier J.C., Garber N.J., and Ehrhart A.A., (2003). Effect of Speed, Flow and Geometric Characteristics on Crash frequency for Two-Lane Highways. Transportation Research Record 1717, pp. 76-83, Washington.

Trivedi D.E., Del Mistro R., and Fieldwick R., (1998). The contribution of traffic volumes, speed, congestions, road section block length, and kerbside activity to accidents on urban arterial roads Xth IRF World Meeting, Stockholm, 1-5 June.

Turner, S.A. (2001) Accident prediction models. Transfund New Zealand Research Report.

Ung J. (2007). Improved prediction models for road crash savings –urban. NBS.9807, Contract Report RC 90122-2 for Austroads, ARRB, Australia Road Research Board, Melbourne.

Whitlock F.A., (1971). Death on the road; a study in social violence. Tavistock Publications Ltd.

World Health Organization, (2011). “Global Status Report on Road Safety, time for action”, WHO, Geneva.

Zomer G.D., Aeron-Thomas A., Astrop A. and Kalaga R.R. (1990), Estimating global road fatalities. TRL Report 445. Transport Research Laboratory Ltd, Crowthrone.

Table 4.2: Parameter estimates of Poisson regression model for the number of persons killed by road accidents from 2009-2013 in Ghana

VARIBLES	ESTIMTES	STD. ERROR	RATE	P-VALUE
Intercept	1.50417	0.47481	4.50	**7.85e-06**
COLLISSION TYPE				
Overturn	0.0000			
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
Hit parked vehicle/pedestrian/animal	-2.03934	0.08431	0.13	0.92492
DRIVER ERROR				
Inexperience	0.0000			
Inattentive	-0.80611	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	0.0000			
Good 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	0.07525
Cycle	-1.86621	0.05785	0.15	0.22513
Others	-2.08935	0.06380	0.12	0.90341
WEATHER COND.				
Clear	0.0000			
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 CONDTION				
Day	0.0000			
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	0.04052	0.16	0.68584

Table 4.4: Parameter Estimates of Negative Binomial Count Regression Model for the number of persons killed by Road Accidents from 2009-2013 in Ghana

VARIBLES	ESTIMTES	STD. ERROR	RATE	P-VALUE
Intercept	1.50417	0.33361	4.50	0.00153
COLLISSION TYPE				
Overturn	0.0000			
Rear end	-0.98795	0.57719	1.33	0.75397
Ran off	0.28387	0.20232	0.20	0.64892
Side swipe	-1.60780	0.03504	0.17	0.27370
Right angle	-1.76591	0.22843	2.54	0.09011
Head on	0.93386	0.37271	**2.53**	**0.03095**
Hit object on road	0.92703	0.06450	0.11	0.23864
Hit object off road	-2.21040	0.42633	0.13	0.12839
Hit parked vehicle/pedestrian/animal	-2.03934	0.62446	0.37	0.10792
DRIVER ERROR				
Inexperience	0.0000			
Inattentive	-2.80611	0.57791	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	0.0000			
Good 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
Others	-2.08935	0.25879	0.12	0.85038
WEATHER COND.				
Clear	0.0000			
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 CONDTION				
Day	0.0000			
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

Table 4.6: Parameter Estimates of Conway-Maxwell Poisson count regression Model for the number of persons killed by Road Accidents from 2009-2013 in Ghana

VARIBLES	ESTIMTES	STD. ERROR	RATE	P-VALUE
Intercept	1.50417	0.03026	4.50	5.42e-06
COLLISSION TYPE				
Overturn	0.0000			
Rear end	-0.98795	0.35472	1.33	0.321272
Ran off	0.28387	0.48344	0.20	0.251744
Side swipe	-1.60780	0.41280	0.17	0.344902
Right angle	-1.76591	0.28196	2.54	0.166218
Head on	0.93386	0.82213	**2.53**	**0.001123**
Hit object on road	0.92703	0.58774	0.11	0.454128
Hit object off road	-2.21040	0.54452	0.13	0.344902
Hit parked vehicle/pedestrian/animal	-2.03934	0.21837	0.37	0.891653
DRIVER ERROR				
Inexperience	0.0000			
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			
Good 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
Others	-2.08935	0.34198	0.12	0.982485
WEATHER COND.				
Clear	0.0000			
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 CONDTION				
Day	0.0000			
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