6835

KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY
KUMASI, GHANA
COLLEGE OF SCIENCE
DEPARTMENT OF MATHEMATICS

APPLICATION OF ARIMA MODELS TO THE INCIDENCE OF ROAD TRAFFIC ACCIDENT, CASUALTY AND FATALITY CASES IN GHANA

THESIS SUBMITTED TO THE DEPARTMENT OF MATHEMATICS
KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTERS OF PHILOSOPHY IN APPLIED MATHEMATICS

BY

ADU-POKU KOFI ASAMOAH (PG 6202911)
BSC. MECHANICAL ENGINEERING

SUPERVISED BY

DR. R. K AVUGLAH

MAY 2013

DECLARATION

I hereby declare that this submission is my own work towards the Mphil, and that to the best of my knowledge, it contains no material previously published by another person nor material, which has been accepted for the award of any other degree of the University, except where due acknowledgement has been made.

Adu-Poku Kofi Asamoah	....	13/05/2013
Student's Name	Signature	Date
Certified by		
Dr. R. K. Avuglah	....	13TH MAY, 2013
Supervisor's Name	Signature	Date
Mr. Emmanuel Harris	....	13/05/2013
Co-Supervisor's Name	Signature	Date
Certified by		
Mr. F. K Darkwa	....	....
Head of Department	Signature	Date

DEDICATION

This work is dedicated to my lovely mum and dad, Mr. and Mrs. Adu-Poku for their immense support and contribution towards my academic pursuit. Mum and Dad, I am very grateful.

ACKNOWLEDGEMENT

First of all, my deepest gratitude goes to the Almighty God who has seen me throughout my academic life till now and my sincere thanks to all KNUST (mathematics lecturers and students whose contribution have made this work a success).

My profound gratitude also goes to my supervisor's, Dr. Avuglah and Mr. Harris; every contribution towards my academic life is recognized and appreciated. Many thanks also go all the staff and workers of National Road Safety Commission and Building and Road Research Institute Kumasi for their support.

I would like to thank Desmond Appiah, Frank Osei, Mr Anarfi and Mr Henry Afriyie for their support and encouragement towards my academic pursuit, May God richly blesses you. I am equally grateful to my parents, my siblings and my study group members especially Mr. Adu Gyamfi, Richard Tawiah, Felix and Yakubu. Please accept my sincere appreciation.

ABSTRACT

Road traffic accident in Ghana is increasing at a fast rate and has raised major concerns. In this thesis, time series with Box – Jenkins method was applied to 20 years of annual road accident, casualties and fatalities data from 1991 – 2010 to determine patterns of road traffic accident cases, casualties and fatalities in Ghana.

In this thesis, we developed models for accident cases, casualties and fatalities. ARIMA (0,2,1) was used to model accident and casualty cases in Ghana whilst ARIMA (0,1,1) was used to model fatality cases. A five year forecast was made using the models developed and it showed that, road traffic accident cases, casualties and fatalities would continue to increase in Ghana.

Relationship between accident, casualties and fatalities were also examined using the Engle granger approach.

TABLE OF CONTENTS

4.0.0 CHAPTER 4: RESULTS AND ANALYSIS

LIST OF TABLES

LIST OF FIGURES

GLOSSARY

ROAD TRAFFIC ACCIDENT: When a vehicle collides with another vehicle, pedestrian, animal or geographical obstacle.

INJURY: All types of damage to the body such as cuts, wounds, fractures caused by road accident.

ACCIDENT DEATH: Loss of human life immediately after road accident.

ACCIDENT CASES: Number of road accident within a given period.

AUTOCORRELATION: Refers to the correlation of a time series with its own past and future values.

AUTOCORRELATION FUNCTION: Is a set of autocorrelation coefficients arranged as a function of separation in time.

TIME SERIES: Is a system of observations ordered in a time.

STATIONARITY TIME SERIES: When a time series values fluctuates with a constant mean and a constant variance.

DIFFERENCED DATA: If a time series is not stationary, it is differenced. The differenced data contains one point less than the original data.

AUTOREGRESSIVE MODEL: A model in which future values are forecasted purely on the basis of past values of a time.

MOVING AVERAGE MODEL: A model in which future values are forecasted purely on the basis of linear combinations of past forecast errors.

CONFIDENCE INTERVAL: The degree of certainty of obtaining the same results if the study is to be repeated.

ABBREVIATIONS / ACRONYMS

ACF	Autocorrelation Function
AIC	Akaike Information Criterion
AIC_C	Corrected Akaike Information Criterion
AR(p)	Autoregressive model of order p
ARIMA	Auto Regressive Integrated Moving Average
ARIMA(p,d,,q)	A model with AR of order p, integrated(differenced) d and MA(q)
ARMA	Autoregressive and Moving Average
BIC	Schwarz"s Bayesian Criterion
MA(q)	Moving Average model of order q
MTTU	Motor Traffic and Transport Unit
NRSC	National Road Safety Commission
PACF	Partial Autocorrelation Function
Q – Q Plot	Quantile - Quantile plot
RTA	Road Traffic Accident
BRRI	Building and road research institute
s.e.	Standard error
WHO	World Health Organisation

CHAPTER 1

INTRODUCTION

1.10 BACKGROUND OF STUDY

Road traffic accidents are the most frequent causes of injury-related deaths world wide (Astrom, et al. 2006). According to the World Report on Road Traffic Injury Prevention (Peden et al., 2004) traffic accidents account for about 3000 daily fatalities worldwide. Statistical projections show that during the period between 2000 and 2020, fatalities related to traffic accidents will decrease with about 30% in high income countries. The opposite pattern is expected in developing countries, where traffic accidents are expected to increase at a fast rate in the years to come.

World Health Organization (WHO) strategy of 2001 reports that currently road traffic injuries are the leading cause of deaths and injuries, the 10th leading cause of all deaths and 9th leading contributor to the burden of disease worldwide based on disability adjusted life years. The number of deaths resulting from road traffic crashes have been projected to reach 8.4 million in the year 2020.

Worldwide reports reveal the problem of accidents being equally serious. According to research carried out by Pierce and Maunder (1998), under the auspices of Road Research Laboratory in UK, they found out that, road accidents world wide are estimated to a total of 20,000,000 victims for a time period which 70% of the accidents occurred in developing countries. The number of accidents per registered vehicles were 10% to 20% higher in developing countries than in the developed world.

Transportation is a major generator of employment and plays a vital role in the distribution of essential goods and services from place to place (Herbert, 1979; Meket, 1997). Road transport plays a key role in the national traffic flow of developing countries and accounts for more than 95 percent of inter-urban transport of goods and passengers in different African countries including Ethiopia (UNTACDA, 2000).

Clearly road transport has an important role in economic, social and cultural functioning of cities. But in many cities today it is also generating significant social and economic costs (Shefer,1997). These costs arise from the external effects of traffic system, particularly accidents, congestion, consumption of public space, air pollution, noise, and disruption of social and economic interaction (Alshuler,1965; Reyonds,1966; Creighton, 1970; Wough, 1990; Rienstra,1996; piet,1997; WHO,2000). These externalities of traffic are especially pertinent in urban areas because here spatial densities are high and the infrastructure networks are most intensively used. For example Barrett (1989) indicated that about 70-80 percent of road accidents of every nation occur in urban areas.

According to the World Health Organization (WHO) about 1.2 million people are killed in accident and 50 million injured on the roads around the world each year and is the most leading cause of death among children 10-19 years of age. This report also noted that the problem was most severe in developing countries and that simple prevention measures could halve the number of deaths. (Peden et al. 2004). Moreover, it is stated that over 90% of the world fatalities on roads occur in low-income and middle-income countries, which have only 48% of the world's registered vehicles and predicted that road traffic injuries will rise to become the fifth leading cause of death by 2030.(WHO 2009).

World Health Organization (2000) statistics also reveal that in developing countries road accident is the major factor that brings about death next to those caused by natural factors;

one death per annum is recorded for every 50 to 500 motor vehicles, whereas the corresponding range in developed countries is 2000 to 5000 motor vehicles.

The total economic costs are also the highest when measured the productivity loss and expenses incurred because of road traffic accidents. TRL (2000) estimated that the social cost of road accident in 1999/2000 was in excess of 500 billion US dollars and the cost in the developing world was estimated to be about 65 million US dollars.

Road traffic accidents occur as a result of several factors associated with the traffic system, namely: road users, road environment, and vehicles conditions.

In Ethiopia, in 2004/5, 93% of all accidents involved human factors, 5% accounted for vehicle factors, and 2% were associated with road environments according to the Federal police report. Growth in urbanization and in the number of vehicle in many developing countries has led to the increase in traffic accidents on road networks which were never designed for the volumes and types of traffic which they are now required to carry. In addition, unplanned urban growth has led to incompatible land uses, with high levels of pedestrians/vehicle conflicts. The drift from rural areas to urban centre often result in large numbers of new urban resident unused to such high traffic levels.

Many developing countries continue to repeat the mistakes of the industrialized countries, many still permit linear development with direct access from frontage properties along major roads even though this is known to lead to safety problems. It is possible to identify hazardous sections of the road network so that appropriate remedial measures can be undertaken to reduce the likelihood and severity of accidents at those locations. This has proven to be one of the most cost-effective ways of improving road safety in industrialized countries.

In developing countries the proportion of serious injured and killed casualties are higher than in the developed countries. An analysis of cross sectional data on road traffic related deaths has shown that the poorest countries have highest road traffic related mortality rates (Soderlund et al 1995). In this analysis, many industrialised countries appear to have introduced interventions that reduced the incidence of road traffic injuries and improve survival of those injured (Soderlund et al 1995). In developing countries there are some peculiarities regarding the accident profiles. A study done in Calcutta India, reported that there are some host (human) factors (such as the behaviour of drivers, pedestrians and cyclist behaviours) and seasonal factors (weather and time) that contribute to fatal road traffic accidents (Zhang et al 1998). Overall, most traffic accidents occurred on main roads (highways) and in the majority of cases pedestrians were found to be at fault during crossing the roads (Majumder et al 1996).
Studies done world wide have shown that In Ghana, the escalating incidence of road accident is no news to a reasonable Ghanaian of ordinary intelligence. Despite increased road safety campaigns, the rate at which accidents occur on our roads is very alarming. For example, a comparison between the accident statistics of 2002 and 2008 shows that the number of road accidents increased from 10,718 in 2002 to 11,209 in 2008 (National Road Safety Commission (NRSC) Ghana, 2005).

These figures show that the accident situation in Ghana requires attention.

In developing countries, including Tanzania, the scenario is different to developed countries, road traffic accidents are increasing with time and mortality due to road traffic accidents is also on the rise (Asogwa ,1992). When taking the population figures into account, developing countries in Sub-Saharan Africa have the highest frequency of various accidents worldwide Peden et al,(2004). Although an implication of this is that the risk environments in countries need further empirical attention, few studies have investigated how people in those societies perceive risk. This scenario calls developing countries to put more effort toward control and prevention of road traffic accident and their outcome. This can be achieved through multidisciplinary approach and research.

Traffic safety in Ghana is in its infantry. Lacking necessary financial, technical and data resources, the studies of road traffic accidents, its causes and impacts on human, social and economic conditions are sporadic and limited. Some traffic safety programmes have been implemented to control and reduce traffic accidents, not necessarily based on evidence of scientific knowledge of the underlying causes of the problem. Few of these implemented programs have been fully evaluated and two of recent evaluations have been reported in the literature.

Studies done world wide have shown that road traffic accidents are the leading causes of death of many adolescents and young adults (Odero et al 1997;Balogun et al 1992). There

is evidence that using minimum safety standards, crash worthiness improvement in vehicles, seatbelts use laws and reduced alcohol use can substantially reduce deaths on the road (Leon,1996).

However, to promote road traffic safety it is very important to keep records of accidents that do occur and its consequent effects. Many jurisdictions require the collection and reporting of road traffic accident statistics. Such data enables figures for deaths, personal injuries and possible property damage to be produced, and correlated against a range of circumstances.

1.2.0 PROBLEM STATEMENT

Road traffic accidents are a major public health problem worldwide, accounting for almost 1.2 million deaths per year representing 2.2% of all deaths (WHO,2004). Continual media reports reveal that Ghana's road accident is oddly high among developing countries. Road accidents are common in the country to the extent that in 1995, Ghana ranked second to Mexico in terms of road fatalities worldwide. In 1997, it ranked second to Nigeria in West Africa and has remained at this position to date (NRSC, 2009).

Road traffic accidents in Ghana have been increasing since independence in 1957. There has also been a simultaneous increase in the import of vehicles. Motor related accident occur in major up country highway much more than many other urban centres due to high traffic volumes. In 1980's, the Government started to take some measures to control this problem. The existing regulation where enforced, including annual road safety campaigns. For instance the busses were prohibited to travel during the night and recently the Government has introduced vehicles speed limiters installed in the engine in all public busses.

Men, women and children are injured on our roads, some of whom become permanently disabled. Precious lives are lost thereby dwindling down our scarce labour force. Indeed, this conquerable foe is devouring our human and economic resources. Hence, traffic accident posses health, economic and developmental challenges to rural communities, towns, districts and the nation at large.

In Ghana, the number of vehicles on the roads has greatly increased and unfortunately road maintenance, driver's education, vehicle upkeeps and traffic regulations have not grown accordingly. A high accident level increases the dependency burden of the country. Working parents are killed or injured in traffic accidents leaving behind children who relied solely on these deceased persons for sustenance.

Casualties from traffic accidents impose a heavy burden on the specialized health care facilities. In addition, the cost of repair and replacement of damaged vehicles demand resources that otherwise could be devoted to other high priority human development sectors such as education, food production and health. Thus, many countries pay dearly for the cost of the modernized transport system and increased mobility in the absence of compensatory mechanisms for ensuring safety. The total cost of road accidents have been estimated to be more than one percent of GDP in developing countries (Down 1997).

The curiosity of this study under its objective seeks to fill the gap of knowledge which exists by identifying the trend and pattern of road traffic accidents so as the findings may be useful for further implementation of the road safety measures or a baseline of similar studies.

1.3.0 OBJECTIVES

The general objective of the study are as follows:

- ❖ To identify patterns of road traffic accident cases, casualties and fatalities in Ghana over the period of 1991-2010.
- ❖ To develop a suitable time series forecasting model for road traffic accident cases, casualties and fatalities in Ghana over the period 1991 - 2010 and use it to estimate 5 years forecast.
- ❖ To determine the long run relationship between accident, casualties and fatalities in Ghana.

1.4.0 SIGNIFICANCE OF THE STUDY

Time series techniques are used in many fields and road safety is no exception. The results of the thesis would add to the many research works carried out in road safety.

This study is mainly concerned with road traffic accident in Ghana and emphasis is given to studying and measuring the trend and pattern of road traffic accident behavior therefore, the significance of the study can be stated as follows:

1. Even though the study is carried out for academic purposes and it is confined to only Ghana, it could be helpful to have a deeper knowledge about the complex problem of urban road transport in general and accidents in particular.

2. Another significance of the work of this thesis was to provide a better opportunity for the national Motor Transport Traffic Unit to use better and more reliable statistical technique such as time series in analyzing their accident data as this would help them in making accident forecasts.

3. The findings obtained from the study would be helpful to gain information and knowledge about the patterns of road accidents in Ghana.

4. It is important by the government and municipal authorities to determine the need for road improvements, vehicle inspections and to initiate programmes for educational and propaganda purposes.

5. The findings would serve as a source of information for those institutions concerned with road safety management and tend to improve the quality of decision-making in urban road transport safety planning.

6. Finally, the findings would help carry out further research to refine the conceptual and methodology of the present study.

1.5.0 STRUCTURE OF THE THESIS

The thesis is organized in five chapters that are linked to the issues in relation to the Study. It also includes information from various sources relating to the study.

Chapter one gives the background of the study, problem statement, significance, scope and limitations and states the objective of the study.

Chapter two reviews the relevant literature. The review covers several road accident related subjects such as previous analyses on highway work zone crashes, statistical methods and applications in fatal and injury data analysis, the current situation of the road accident in the country, the present accident situations and reasons for high number of road traffic accidents are reviewed and discussed in this chapter.

Chapter three outlines the methodology used in the study. It elaborates the qualitative and quantitative methods used.

The fourth chapter discusses the major findings from the data analyses, characterizations and identification of accidents. It also addresses the data analysis procedure and results.

Finally, based on the analyses, discussions and recommendations, the research comes up with a general conclusion which is in chapter five.

1.6.0 METHODOLOGY

Data for the study was secondary. A historical annual traffic crash data for the years 1991 through 2010 was compiled from the National Road Safety Commission (NRSC) and Building and Road Research Institute (BRRI). These two stations are responsible for compiling road traffic accident data in Ghana. The data was classified into number of accident cases, fatalities and injuries. R software was used for the analysis.

Time series analysis was the main statistical tool used for the analysis with greater emphasize on Box – Jenkins method.

The methodology consist of these steps namely

- Model identification
- Model estimation
- Diagnostic check on model adequacy
- Forecasting

1.7.0 SCOPE AND LIMITATION

The scope of the thesis was limited to the following:

Twenty one (20) years of annual number of accident cases, casualties and fatalities in Ghana for the period 1991-2010.

We used Box-Jenkins methodology to develop a time series model for both descriptive and forecasting purposes.

CHAPTER 2

LITERATURE REVIEW

2.1.0 STATISTICAL MODELS OF ROAD TRAFFIC ACCIDENT

Considering the effect of speed limit, Nataliya (2006) researched on the influence of posted speed limit on roadway safety in Indiana. He used binary and multinomial logit model for the research and based on his findings, he concluded that speed limits do not have any significant effect on road accident.

He also considered the relationship between speed limits and roadway safety but focused on the influence of the posted speed limit on the causation and the severity of road accidents. Duncan et al.(1998) found that the probability of one being injured in causation of accident severity is being a female. They also found that injury severity is increased by high speed differential, high speed limit and drunk driving. In a related research, he used the ordered probit model to injury severity outcomes in truck-passenger car rear-end collisions.

Analyzing safety issues, Shankar et al. (1996) used rural highways in Washington state to analyze accident severity and suggested that environmental conditions, accident type, vehicle characteristics, highway design, and driver characteristics significantly influence accident severity using the nested logit model.

They also analyzed several factors causing accident severity such as overturn accidents, rear end accidents on wet pavement fixed object accident and failure to use the seat belt contribute to higher probabilities of property damage and finally, he concluded that icy pavement and single vehicle collision lead to higher probability of property damage.

Considering the effects of trucks and vehicles involvement in an accident, Chang and Mannering (1998) did a research on that and based on their findings, they found that accident injury severity is worsened if the accident has a truck involved using the nested logit model, and also the effects of trucks are more significant for multi-occupant vehicle than for a single occupant vehicle. In a related research, Kockelman and Kweon (2002) focused on pickups, sporty utility vehicles and passenger cars and were comparing the accident severity rate between them.

They applied the ordered probit models in modeling of a driver injury severity outcomes and concluded that pickups and sport utility vehicles are less safer than passenger cars in a single vehicle collision. In a year later, Kweon and Kockeleman (2003) again used poisson and ordered probit model in another research and focused on the probabilities of accident severity outcomes for a given fixed driver exposure. They found that young drivers are far more crash prone than older drivers and also sporty utility vehicles and pickups are more likely to be involved in a rollover accident. Khattack (2001) used the ordered probit models for severity outcomes of multi vehicle rear end accident in North Carolina. In his analyses, he found that being in a newer vehicle protects the driver in the rear end collisions. The results of his research indicated that in a three vehicle collisions, the driver in the middle is more likely to be severly injured and in a two vehicle collisions, the leading driver in the middle is more likely to be severly injured.

The effect of ice warning signs was also analysed by Carson and Mannering (2001) and used two statistical models in the research. They used the inflated negative binomial and logit models respectively to study the effects of ice warning signs on ice accident frequencies and severities in Washington state and based on their findings, they concluded that the presence of ice-warning signs were not a significant factor in reducing ice-accident frequencies and severities.

Using the concept of genders, Ulfarsson (2001) and Mannering (2004) used multinomial logit model to establish the behavioral and physiological differences between genders. Their analysis focused on males and females and based on their analysis, they found that the probability of being involved in a disabling injury and fatal accident is higher for females as compared to males. Abdel-Atty (2003) analyzed driver injury severity outcomes at different road locations in Central Florida using the ordered probit model and based on his findings, he concluded that male drivers, those not wearing seatbelt, over speeding, those who drive in rural areas and lastly older drivers contribute to higher probability of injury severity.

Abdel-Aty and Radwan (1999) carried out a similar research in central Florida where they used binomial technique to model the frequency of accidents on a principal arterial in Central Florida. In ~~their research~~, they found that female drivers experience more accidents than male drivers in heavy traffic volume, reduced mean width, narrow lane width and larger number of lanes.

Male drivers also have a greater tendency to be involved in traffic accidents while speeding and also the research indicated that young older drivers experience more accidents than middle aged drivers in heavy traffic volume.

Finally they concluded that narrow shoulder width, reduced median width, urban roadway intersections, heavy traffic volumes and speeding increase the likelihood of accident involvement. Yamanutu and Shankar (2004) proposed a research in the analysis of driver and passenger injury severity collisions with fixed object using bivariate ordered probit model and concluded that collisions with sign posts, faces of guardrail, concrete barrier and fences tend to cause less severe injuries whilst collisions with leading ends of guardrail and trees tend to cause more severe injuries. Savolamen (2006) and Mannering (2006a) used the multinomial and nested logit model to focus on motorcycle safety in Indiana roads.

Based on their analysis, they found that head-on collisions, collisions with fixed object, right angle, not wearing helmet, alcohol use, unsafe speed and lastly poor visibility contribute significantly to motorcycle accident in Indiana. Multinomial logit model was also used by Khurashadi et al (2005) to explore the differences of driver injury severities in rural and urban accidents involving large trucks and found that in accident where alcohol, drug use ~~is identified~~, the probability of severe/fatal injury is increased by 250% and 80% in rural and urban areas. They also concluded that the probability of severe/fatal injury increases by 26% in rural areas and 700% in urban areas.

Analysis of traffic accidents at urban intersections in Riyadh was done by Al-Ghamdi. In his study, an attempt was made to investigate traffic accident that occurred at both intersections and non-intersections site. He used the conditional probability and contingency table analysis to make his inferences and based on his inferences, he found that improper driving behavior is the primary cause of accidents at signalized urban intersections in Riyadh. Running a red light and failing to yield are also contributing factors. Abhishek analysed run-off road crashes involving over corrections. Overcorrection occurs when a vehicle begins drifting off the road one way and the driver over steers in the opposite direction leading the vehicle to cross over into incoming lanes of traffic. This research developed models to identify the factors that influence run-off road crashes involving overcorrections.

He used binary logistic regression model in his work and concluded that straight moving vehicle is less likely to be involved in an overcorrection crash than a vehicle moving in any other pattern. It was also analyzed that the presence of rumble strips have a negative impact on overcorrection and concluded that all of these factors were significant at minimum of 85% level. In his research, an additional statistical model analyzed the factors that lead the vehicle to leave the paved shoulder prior to overcorrection. These factors are drivers age, sex, vehicle movement, vehicle type, speeding, annual daily traffic and the presence of speed rumble strips. Finally he concluded that female drivers are nearly 1.5 times more likely to overcorrect than males.

Lee and Mannering (2002) also carried out a similar research to determine the frequencies and severities of run-off road accident in Washington state using the zero inflated count data models and nested logit models and based on their results, they concluded that high speeds, the presence of guardrail, alcohol and impaired driving contribute to the high increase of run-off roadway accident severity in the state.

In their research, they also proposed measures to curb these menace such as avoiding cut side slopes, decreasing the number of isolated trees along roadway, decreasing the distance from outside shoulder edge to guardrail. Considering over steering and vehicle dynamics as a contributing factor Melcher(2003) analyzed how over steering leads to loss of directional control based on dynamics principles. George et al. (2006) did a research on median cross over crashes in wisconsin between 2001 and 2003 and based on their findings, they realized that seasons or time or year play a significant role in determining the median cross over crashes. Their results reveal that younger drivers are prone to a higher severe crash than older drivers when the traffic volumes on roadway is relatively high.

Bedard et al. (2002) researched on the independent contribution of vehicle characteristics to the drivers fatality risk. Multinomial logistic regression revealed that speed in excess of 69 mph prior to or impact were related to higher fatality odds compared with speeds of less than 35 mph. Also female drivers and blood alcohol concentration greater than 0.3 were also associated with higher fatality odds and lastly the analysis shows that the odds ratio of fatality increases with age. Aworami et al. proposed a research on analytical study of the causal factors of road traffic crashes in Southwestern Nigeria.

The study examines the causal factors of road traffic crashes in some selected states in Southwestern part of Nigeria. Regression analysis was adopted in analyzing the data obtained in order to establish the relationship between human characteristics, vehicular characteristics, roadway characteristics and environmental characteristics and based on their findings, it was concluded that human, vehicle, roadway and environment have significant contribution of about 79.4% on the road traffic crashes in the study area. Dissanyake researched on the characteristics and contributory causes related to large truck crashes.(fatal crashes)

In his research he used the Bayesian statistical approach to analyze characteristics and factors contributing to truck-involved crashes. Driver, vehicle and crash related contributory causes were identified. He also used multinomial logistic regression to model the type of fatal crash(truck vs non-truck) to compare the relative significance of various factors in truck and non-truck crashes. He identified factors such as cellular phone usage, failure to yield right of way, inattentiveness and failure to obey traffic rules also have a greater probability of resulting in fatal truck crashes. Amoung several other factors, inadequate warning signs and poor shoulder conditions were also found to have greater predominance in contributing to truck crashes than non-truck crashes.

His findings also revealed the pattern and trend of motor traffic accidents in Kibaha district from 2001 to 2004. Based on his analysis it was revealed that young male aged between 25 and 34 ~~that are~~ economically active are highly prone to motor traffic accidents. Also the road users who are always at risk of dying on the road were found to be passengers and pedestrians. The study also revealed that the risk of dying if one is involved in an accident during the night was significantly higher than during the day especially when it is raining. Age, sex, over speeding, reckless driving, being a pedestrian or motor cyclist were identified as risk factors to motor accidents and finally his study shows that buses followed by minibuses which are operated by private companies contribute the highest percentage of injured and killed casualties in kibaha district. Jehle and Cottington(1998) used the chi-squared test to analyze studies on pedestrian injury severity in pittsburg, USA and concluded that the more alcohol consumed, the more severe injuries, also the proportion of alcohol related accident was highest in the 25-34 age group. Holubowcz(1995) also used the chi-squared test on the fatality and severity injury in Adelaide, Australia and the findings was that the highest fatality rate was in those ages greater or equal to 75, large proportion of the serious of fatality injured were males and lastly there was an increase in alcohol consumption amoung fatality injured young and middle aged males.

Pless et al (1989) used the logistic regression on injured children aged 0-14 in Canada and concluded that higher injury risk was related to fewer years of parents education, family members, accident history, unsafe environment and poor parental supervision were weakier risk factors compared to family and neighborhood characteristics. Argan et al.(1989) also applied the conditional logistic regression model on Hispanic children in southwestern united states and the result was that household crowding, family's moving within the past year, poverty and parents inability to read well increased injury risk and finally being in one-parent and non-driving family did not result in injury risk.

Logistic regression model was also used by Ballesteros et al.(2003) on fatality and serious injuries and concluded that being hit by sport utility vehicles and pickups related to fatal injuries compared with conventional passenger cars and vans with increasing speed limits, mortality and injury severity values increased. Plurad et al.(2006) used the logistic regression model in los angeles and based on their analysis, high alcohol level was not associated with injury with severities. Macleod et al.(2010) used the logistic regression model to analyse hit and run accident cases in USA and concluded that the risk of hit and run increased in the early morning, during non day light and on weekends, young male driver, alcohol use and history related to hit and run. Logistic regression model and ordered probit model were used by moudon et al.(2009) to analyse collision with motor vehicle in Washington, USA and concluded that crossing at intersections without signals and being hit by vehicles moving straight on the roadway increased injury severity.

Zajac and Ivan (2003) used the ordered probit model to analyse injury and fatality rate in Connecticut, USA and concluded that individuals greater than 65 years of age had more severe injuries, alcohol and increase in roadway width increased injury severity, crashes in downtown and compact residential areas were less serious compared with low density residential areas, crashes in low and medium density commercial areas were less severe compared with village and downtown fringe areas. Siddiqui et al. (2006) also used the ordered probit model to analyse crashes while crossing streets in Florida, USA and concluded that death were more likely at midblock locations than intersections for any light condition, daylight and street lighting reduced the odds of fatal injuries. Clifton et al (2009) used the generalized ordered probit model on the same research and concluded that children and men are more likely to be injured and older people were more likely to be killed crossing against the traffic signal, not being in a cross walk and crash after dark increased injury severity. Hill et al. (2003) used the logistic regression model in the analysis of the characteristics of workzone fatal crashes in texas. Dissanayake and Lu (2002) also used a set of sequential binary logistic regression model to analyse the alerting factors and predict the crash severity of single vehicle fixed object crashes involving young drivers.

Their research scope focused on young drivers because they have been considered to be vulnerable in severe traffic crashes. Based on their findings, they concluded that restraint device usage and being a male clearly reduced the tendancy of high severity whilst some other variables such as weather conditions, residence locations and physical condition were not important at all. Also they concluded that the influence of alcohol or drugs, ejection in the crash, rural crash locations and speed of the crash vehicle could significantly increase the probability of having a more severe crash. Ouyang et al (2002) developed a methodology which used a simultaneous binary logit model to account for the interrelationship among the injury severity outcomes in multi vehicle collision and concluded that head on collision, alcohol impairment and curve-high speed interaction for truck could cause high severity crashes.

Donnel et al. (2002) developed a negative binomial regression model to statistically analyse cross-median crashes and based on their findings, they concluded that inside-shoulder width, narrow lane width and cross-slope pavement could increase the likelihood of cross median crashes. Yuan et al. (2001) conducted a study to evaluate the safety benefits of ~~angle realignment~~ and curve realignment of intersection approach and based on their study they concluded that the improvement studied reduced the total number of crashes in different levels for different type of crashes, the curve realignment improvement reduced run-off road crashes and head on/rear-end crashes significantly while the run off road crashes increased at some sites with angle realignment and also intersection realignment combined with the addition of left turn lane did not have extra benefits in reducing the total number of crashes.

Turner and Georggi (2001) studied the characteristics of alcohol related motorcycle crashes in Florida using the general data statistics method and considered human related factors and physical aspects. The human related factors included age, gender, alcohol use, licensing status and helmet usage whilst the physical aspects included time of day, day of week, monthly trends, vehicle condition, road condition, environmental conditions and driver factors, based on the study, they found that the largest percentage of fatal motorcycle alcohol related crashes happened in march when an annual motorcycle event was held in Florida.

Human errors caused most of the crashes whilst the impact of environmental factors were negligible, also drivers without helmets likely had very severe injuries when a crash happened. Taylor et al. (2001) conducted a study to determine the safety impact of replacing bi-directional median cross overs with directional median cross overs on urban arterials in Michigan and concluded that an average of over 30% reduction in both total crashes and injury crashes which indicated that changing bi-directional cross overs to directional cross overs was an effective safety enhancement. Kim and his colleagues (2000) used ~~logistic~~ regression to facilitate their analysis on motorcycle crash characteristics and found that motorcycle safety needed to be improved by focusing both educational and enforcement efforts on specific age groups at specific times and at key locations. Xiao et al. (2000) developed two fuzzy logic models to predict the risk of crashes that could occur on wet pavements.

The input variables were posted speed, average daily traffic, skid number and driving difficulty. One of the models was based on mamdani's fuzzy inference method and the other one was sugeno-type fuzzy logic model. In their research, the developed fuzzy logic model was compared with a regression model and a probabilistic model. The researchers found that the regression model is not a good model for evaluating the wet pavement crash risk and the probabilistic model lacked reliability and efficiency in describing the relationship between traffic characteristics. Thus they concluded that their fuzzy logic models were superior in accurately predicting the occurrence of wet pavement crashes.

Jovanis and Chang (1986) found a number of problems with the use of linear regression in their study applying Poisson regression as a means to predict accident. For example they discovered that as vehicle- kilometers travelled increases, so does the variance of the accident frequency. Thus, this analysis violates the homoscedasticity assumption of linear regression. There is also the possibility of predicting accident frequency with negative values. In a well summarized review of models predicting accident frequency, Milton Mannering (1997) stated that ''the use of linear regression models is inappropriate for making probabilistic statements about the occurrences of vehicle accidents on the road''. They showed that the negative binomial regression is a powerful predictive tool and one that should be increasingly applied in future accident frequency studies.

Nassar et al. developed an Integrated Accident Risk Model (ARM) for policy decisions using risk factors affecting both accident occurrences on road sections, and injury severity of occupants involved in the accidents. Using negative binomial regression and a sequential binary logit formulation, the model they developed are practical and easy to use.

Mercier et al. (1997) used logistic regression to determine whether age or gender (or both) was a factor influencing severity of injuries suffered in head-on automobile collisions on rural highways. Logistic regression was also used by Veilahti et al. (1989) in predicting automobile driving accidents of young drivers. James and Kim developed a logistic regression model for describing the use of child safety seats for children involved in crashes in Hawaii from 1986 to 1991. The model reveals that children riding in automobiles are less likely to be restrained; drivers that use seat belts are far more likely to restrain their children; and one -and-two year olds are less likely to be restrained.

Precious researches have also identified the relationship between type of vehicles and accident severity. Sharker and Mannering (1996), Kockelman and Kweon (2002), Viner et al. (1994), Chang and Mannering (1999) did some of these studies. These studies helped the researchers to understand the factors affecting severity and provided an idea about the safety countermeasures to be adopted which would reduce severity of accidents. Again other studies investigated the influence of risk factors related to the driver, the vehicle, the ~~road~~ environment and crash characteristics on severity of accident. These studies mostly examined and emphasized on different types of collision effect, large mass effect as well the effect of age on severity. Amongst the Evans and his associates (1985, 1987, 1992, 1993, and 1994) and Ducan, Khattak and Council (1998) investigated the relationship between vehicle masses and degree of injuries.

2.20 TIME SERIES AND OTHER STATISTICAL METHODS USED IN ROAD TRAFFIC ACCIDENT

In light of problems associated with ordinary (regression) methods because of the assumption that the observations are independent, several researchers have turned to analyzing road traffic accidents data with time series techniques such as ARMA, ARIMA, DRAG and state space models or structural models as a means to better predict accident variables.

Abdel (2005) conducted a road accident research in Kuwait using the time series analysis. His main objective was to establish the differences between ARIMA models and ANN models and concluded that for a long term series without seasonal fluctuations, then Artificial Neural Network (ANN) model is better as compared to ARIMA model.

Road traffic injury procedure in china was established by Wen et al (2005) using RTI data from 1951 to 2003 and proposed a series of predictive equations based on ARIMA models. They concluded that time series models thus established proves to be of significant usefulness in RTI prediction.

Two time series were used by Cejun and Chiou-lin (2004) to predict annual motor vehicle crash fatalities. The models are ARMA and Holt-Winters Algorithm and concluded that the values predicted by Holt-Winters Algorithm are a bit lower than those predicted by ARMA Models.

Ayvalik (2003) also used intervention analysis with univariate Box-Jenkins method to identify whether a change in a particular policy had made an impact on the trends in

fatalities and fatality rates in Illinois. He developed ARIMA forecasting model for future trends in motorway fatalities in an effort to provide assistance to policy development in reducing fatality rates in Illinois.

CHAPTER THREE

METHODOLOGY

This chapter describes in details the theoretical background of time series and the method of time series which we have used in modeling road traffic accident in Ghana during the period of 1991-2010.

3.1.0 TIME SERIES

When a sequence or series of observations are ordered in a time, it is referred to as time series. It can also be explained as a sequence of observations of a periodic random variable. Examples of such sequence observations are the monthly demand for a product, the annual fresh men enrollment into a department of a university. It is used for operation research because they are the drivers of decision models. They can also be used easily for forecasting purposes because historical sequences of observations upon study variables are readily available from published secondary sources. An inventory model requires estimate of future demand, a course scheduling and staffing model for a university department requires estimates of future student inflow. A basic assumption in any time series analysis/modeling is that some aspects of the past pattern will continue to remain in the future. Also under this set up, often the time series process is assumed to be based on past values of the main variable but not on explanatory variables which may affect the variable system.

3.2 TYPES OF TIME SERIES

DEMOGRAPHIC TIME SERIES

This is a type of time series that occur in the study population. One significance of demographic time series is that it helps demographers to predict changes in population for as long as ten or twenty years into the future.

PHYSICAL TIME SERIES

These occur in the meteorology, marine sciences, physical sciences and in geophysics. Examples are rainfall on successive days and air temperatures measured in successive hours, days or months.

ECONOMIC TIME SERIES

This time series arises in the economic sector. Examples are share price on successive days, export totals in successive months, average income in successive months, company profits in successive years and many others.

MARKETING TIME SERIES

It is often essential to forecast future sales in other to plan production and may also be of interest to examine a relationship between sales and other time series such as advertising and expenditure

Marketing time series involves the analysis of sales figures in successive weeks or months. Marketing data has much in common with the economic data.

Observations of time series are collected at equally spaced, discrete time intervals. When there is only one variable upon which observations are made then we call them a single time series or more specifically a univariate time series. A basic assumption in any time series analysis/modeling is that some aspects of the past pattern will continue to remain in the future.

3.3.0 OBJECTIVES OF TIME SERIES

- One objective of time series is to predict the future values. After identification and validation of the model, a forecast model is obtained that can forecast or predict for one or several periods ahead. However, chances of forecast errors are inevitable as the period advances.
- Observations are made at specific times in measuring the quality of a process to give the desired results. Therefore time series is generated to support the process so that at any particular time we may know that a manufacturing process is performing well or moving off target and then appropriate action could be taken.
- Variation in one time series may be used to explain that of another especially when observations are considered on two or more variables. The purpose is to develop an in-depth understanding of the technique which give rise to the given time series.

3.4.0 TIME SERIES COMPONENTS

CYCLICAL

Some time series show oscillation at a fixed period due to some other physical cause. Example, daily variation of temperature. Others do not display oscillation at a fixed period but are predictable to some extent. Cyclical variation is distinguished from seasonal variations in that cycles extent over a longer period than seasonal fluctuations.

TREND

When there is a systematic movement of a time series that changes over time, it is referred to as a trend. It does not repeat within the time usually and is associated with climatic changes. Example includes the increasing of prices of goods during an occasion and having prices reducing immediately after that occasion.

SEASONALITY

Seasonal patterns or effects occur in time series data that are usually associated with climatic changes. Seasonality exists when a time series exhibit regular fluctuations or variation during a particular period of time usually year to year. Example, retail sales tend to peak during any festive season and then decline after the holidays.

ERRATIC OR RESIDUAL

Fluctuations not classified under the other time series components stated above, and are unpredictable or take place at various points in time by chance or randomly, falls under residual components. Residuals which may or may not be random are left when trend and cyclic variations are removed from a set of data. Floods and strikes illustrate irregular components of a time series.

Many data series include combinations of the preceding patterns. After separating out the existing patterns in any time series data, the pattern that remains unidentifiable form the random or error component. Time plot (data plotted over time) and seasonal plot (data plotted against individual seasons in which the data were observed) help in visualizing these patterns while exploring the data.

3.5.0 STATIONARITY OF TIME SERIES

A time series is stationary if values of the time series fluctuate around a constant mean with constant variation. If the values do not seem to fluctuate around a constant mean, then it is non- stationary.

In other words a time series is stationary if there is no systematic changes in mean (no trend) and variances, and strictly periodic variations have been removed. Mathematically, a time series is said to be stationary if the joint probability distribution of $x(t_1),.......x(t_n)$ is the same as the joint distribution of $x(t_1+n),.......x(t_n+n)$ for all $t_1,......t_n$. In other words, when the value of time origin, n is increased the change will have no effect on the joint distribution which must depend on the intervals.

ACHIEVING STATIONARITY

Due to the non-stationary nature of most business and economic time series, it is required that stationarity be achieved before building any model

- ❖ We can difference the data. That is , given the series Z_t , we create the new series $Y_t = Z_t - Z_{t-1}$

The differenced data will contain one less point than the original data.

Although you can difference the data more than once.

- ❖ For non-constant variance, taking the logarithm or square root of the series may stabilize the variance. For negative data, you can add a suitable constant to make the entire data positive before applying the transformation. This constant can then be subtracted from the model to obtain predicted (i.e., the fitted) values and forecasts for future points.

- ❖ If the data contains a trend, we can fit some type of curve to the data and then model the residuals from that fit. Since the purpose of the fit is to simply remove long term trend, a simple fit, such as a straight line, is typically used.

3.6.0 AUTOCORRELATION FUNCTION

Autocorrelation function is a function that describes the relationship between neighbouring data observation in time series. It helps obtain partial description of the process for developing a forecasting model. It also gives evidence about how successive values of the same variable relate to each other. This is achieved by measuring sample autocorrelation coefficient of the time series data.

For a time lag k, the autocorrelation is defined by

$$w_k = \frac{\left[E\left(Y_t - \mu_Y\right)\left(Y_{t+1} - \mu_Y\right)\right]}{\left[E\left(Y_t - \mu_Y\right)^2\left(Y_{t+1} - \mu_Y\right)\right]^{\frac{1}{2}}}$$

Autocorrelation is also sometimes called lagged correlation or serial correlation,. Positive autocorrelation might be considered a specific form of persistence, a tendency for a system to remain in the same state from one observation to the next.

Three tools for assessing the autocorrelation of a time series are:

(1) The time series plot,

(2) The lagged scatter plot, and

(3) The autocorrelation function

The autocorrelation function can be used for the following two purposes:

1. To detect non-randomness in data.

2. To identify an appropriate time series model if the data are not random.

Autocorrelation plots are formed by:

Vertical axis: Autocorrelation coefficient

Horizontal axis: Time lag k = 1, 2, 3... and Confidence bands

3.6.1 PARTIAL AUTOCORRELATION FUNCTION (PACF)

A partial autocorrelation coefficient of order k measures the strength of correlation among pairs of entries in the time series while accounting for all autocorrelation below order k. A partial autocorrelation permits the forecaster to identify the degree of the relationship between current and old values of a variable while holding the effects of all other lags constant.

For example, the partial autocorrelation coefficient for order k=5 is computed in such a manner that the effects of k = 1,2,3,4 partial autocorrelations have been excluded.

SAMPLING DISTRIBUTION OF PARTIAL AUTOCORRELATION FUNCTION

Partial autocorrelation function also follows normal distribution with mean =0 and variance = $\frac{1}{n}$, where n is the sample size.

Specifically, partial autocorrelations are useful in identifying the order of an autoregressive model. The partial autocorrelation of an AR (p) process is zero at lag p+1 and greater.

Partial Auto Correlation Plots are formed by:

Vertical axis: Partial autocorrelation at lag k

Horizontal axis: Time lag k(k=0,1,2,3)

In addition, 95% confidence interval bands are typically included on the plot.

3.6.2 TIME SERIES MODELS

Time series helps to predict the future values of time series based on its past behavior. The time series data should be random. A simple random model is one where the observation consists of an overall arithmetic mean and a random error component.

There are a number of approaches to modeling time series data. These include:

- Autoregressive (AR) Models
- Moving Average (MA) Models
- Autoregressive Moving Average(ARMA) Models
- Autoregressive Integrated Moving Average (ARIMA) Models.

3.6.3 AUTOREGRESSIVE (AR) MODELS

A model in which future values are forecasted purely on the basis of past values of the time series. An autoregressive model of order p, denoted by AR (p) with mean zero is generally given by the equation: $y_t = \phi_1 y_{(t-1)} + \phi_2 y_{(t-2)} + \phi_3 y_{(t-3)} + ... + \phi_p y_{(t-p)} + \varepsilon_t$

Where p is the order of the Autoregressive model, $\phi_1, \phi_2, \ldots .\phi_p$ are the autoregressive model parameter and ε_t is the random shock or white noise process.

Examples of AR models are

$y_t = \mu + \phi_1 y_{(t-1)} + \varepsilon_t$ ....AR (1)

$y_t = \mu + \phi_1 y_{(t-1)} + \phi_2 y_{(t-2)} + \varepsilon_t$ ....AR (2)

$y_t = \mu + \phi_1 y_{(t-1)} + \phi_2 y_{(t-2)} + \phi_3 y_{(t-3)} + \varepsilon_t$ ....AR (3)

In this general case, the ACF damps down and the PACF cuts off after p lags.

An AR (p) model is stationary if the roots of $\phi_L = 0$ all lie outside the unit circle. AR (p) process has only p partial autocorrelation that are statistically different from zero. However, autocorrelation coefficient of an AR (p) models dies down or trails off to zero.

3.6.4 MOVING AVERAGE (MA) MODELS

A model in which future values are forecast based on linear combination of past forecast errors is called moving average model.

A moving average model of order q, with mean zero, denoted by MA (q) is generally given by:

$$y_t = \theta_1 \varepsilon_{(t-1)} + \theta_2 \varepsilon_{(t-2)} + \theta_3 \varepsilon_{(t-3)} + ... + \theta_q \varepsilon_{(t-q)} + \varepsilon_t$$

Where q is the order of the Moving Average model, $\theta_1, \theta_2, \theta_3, \ldots, \theta_q \left(\theta_q \neq 0\right)$ are the moving average model parameter and ε_t is the random shock or white noise process.

Examples of moving average models are

$$y_t = \mu + \theta_1 \varepsilon_{(t-1)} + \varepsilon_t \text{.............................MA(1)}$$

$$y_t = \mu + \theta_1 \varepsilon_{(t-1)} + \theta_2 \varepsilon_{(t-2)} + \varepsilon_t \text{...................MA(2)}$$

$$y_t = \mu + \theta_1 \varepsilon_{(t-1)} + \theta_2 \varepsilon_{(t-2)} + \theta_3 \varepsilon_{(t-3)} + \varepsilon_t \text{..........MA(3)}$$

To identify an MA (q) model, autocorrelation coefficient of the MA (q) model cuts off after lag q but its partial autocorrelation dies off to zero. In this general case, the PACF damps down and the ACF cuts off after q lags. An MA (q) model is necessarily stationary if q is finite.

An MA (q) is said to be invertible if can be inverted, in other words if it can be expressed as an AR. An MA (q) is invertible if the roots of all lie outside the unit circle.

3.6.5 ARMA (p, q) MODELS

This is a combination of autoregressive and moving average models. The advantage of ARMA models over AR and MA is that a stationary time series may often be described by ARMA models involving fewer parameters than pure AR or MA models. The general form of ARMA containing p AR terms and q of MA terms is

$$y_t = \mu + \phi_1 y_{(t-1)} + ... + \phi_p y_{(t-p)} + \theta_1 \varepsilon_{(t-1)} + ... + \theta_q \varepsilon_{(t-q)} + \varepsilon_t$$

Where ϕ and θ are parameters, p and q are the orders of AR and MA models, μ is the arithmetic mean and ε_t is the random shock or white noise process.

Examples of ARMA (p, q) models are

$$y_t = \mu + \phi_1 y_{(t-1)} + \varepsilon_t + \theta_1 \varepsilon_{(t-1)} \text{.......... ARMA (1,1)}$$

$$y_t = \mu + \phi_1 y_{(t-1)} + \phi_2 y_{(t-2)} + \varepsilon_t + \theta_1 \varepsilon_{(t-1)} \text{.......... ARMA (2,1)}$$

$$y_t = \mu + \phi_1 y_{(t-1)} + \varepsilon_t + \theta_1 \varepsilon_{(t-1)} + \theta_2 \varepsilon_{(t-2)} \text{.......... ARMA (1,2)}$$

$$y_t = \mu + \phi_1 y_{(t-1)} + \phi_2 y_{(t-2)} + \varepsilon_t + \theta_1 \varepsilon_{(t-1)} + \theta_2 \varepsilon_{(t-2)} \text{.......... ARMA (2,2)}$$

To identify an ARMA model, both autocorrelation coefficients and partial autocorrelation coefficients decline to zero exponentially.

3.6.6 AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) MODELS

The time series models above are only used when the time series data is stationary. However many real time series are not stationary hence those models cannot be used for the data. Differencing the data one or two times will reduce the non-stationary time series to stationary series. ARIMA also called Box-Jenkins models are the models based on this idea.

In general, an ARIMA model is characterized by the notation ARIMA (p, d, q), where p, d and q denote orders of auto-regression, integration (differencing) and moving average respectively.

Examples of ARIMA models are

$$y_t = \phi_1 y_{t-1} + \varepsilon_t \text{.......... ARIMA (1,0,0)}$$

$$y_t = \phi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1} \text{.......... ARIMA (1,0,1)}$$

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \ldots\ldots\ldots\ldots\ldots \text{ARIMA } (2,0,1)$$

$$y_t = y_{t-1} + \phi_1 (y_{t-1} + y_{t-2}) + \varepsilon_t \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{ARIMA } (1,1,0)$$

3.6.7 PURELY SEASONAL MODELS

A purely seasonal time series is one that has only seasonal models the autocorrelation die down and partial autocorrelation cut off after one season lag for an SAR(1) model and after two seasonal lags for an SAR(2)model.

3.6.8 SEASONAL AUTOREGRESSIVE MODEL (SAR)

Seasonal autoregressive models are built with parameters called seasonal autoregressive parameters (SAR parameters). The SAR parameters represent autoregressive relationships that exist between time series data separated by multiples of number off periods per season. A general seasonal autoregressive model with p SAR parameters is written as

$$y_t = \phi_s y_{t-s} + \phi_{2s} y_{t-2s} + \ldots + \phi_{ps} y_{t-ps} + \varepsilon_t$$

$$y_t = \phi_s y_{t-s} + \phi_{2s} y_{t-2s} + \ldots. + \phi + \theta_s \varepsilon_{t-s} + \theta_{2s} \varepsilon_{t-2s} + \varepsilon_t \ldots. + \theta_{qs} \varepsilon_{t-qs}$$

$$y_t = \phi_s y_{t-s} + \varepsilon_t + \theta_s \varepsilon_{t-s} \ldots\ldots\ldots ARMA(1,1)^s$$

Examples of SAR models are

$$y_t = \phi_s y_{t-s} + \varepsilon_t \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{ARIMA } (1,0,0)^S$$

$$y_t = \phi_s y_{t-s} + \phi_{2s} y_{t-2s} + \varepsilon_t \ldots\ldots\ldots \text{ARIMA } (2,0,0)^S$$

3.6.9 SEASONAL MOVING AVERAGE (SMA)

Seasonal moving average models are constructed with seasonal moving average parameters (SMA Parameters). SMA parameters represent moving average relationships that exist among

the time series observations separated by a multiple of the number of periods per season. The general seasonal moving average model with q parameters is as follows;

$$y_t = \theta_s \varepsilon_{t-s} + \theta_{2s} \varepsilon_{t-2s} + \varepsilon_t .... + \theta_{qs} \varepsilon_{t-qs}$$

Examples of SAR models are

$$y_t = \theta_s \varepsilon_{t-s} + \varepsilon_t \text{..................................ARIMA } (0,0,1)^s$$

$$y_t = \theta_s \varepsilon_{t-s} + \theta_{2s} \varepsilon_{t-2s} + \varepsilon_t \text{..................ARIMA } (0,0,2)^s$$

3.7.0 SEASONAL AUTOREGRESSIVE MOVING AVERAGE (SARMA)

SARMA model is a mixture of SAR and SMA models and are built with parameters of both models. It is generally written as

$$y_t = \phi_s y_{t-s} + \phi_{2s} y_{t-2s} + ... + \phi_{ps} y_{t-ps} + \theta_s \varepsilon_{t-s} + \theta_{2s} \varepsilon_{t-2s} + \varepsilon_t .... + \theta_{qs} \varepsilon_{t-qs}$$

The order of SARMA model is expressed in terms of both ps and qs.

Examples of SARMA models are

$$y_t = \phi_s y_{t-s} + \varepsilon_t + \theta_s \varepsilon_{t-s} \text{..........} ARMA(1,1)^s$$

3.7.1 BOX AND JENKINS METHOD OF MODELLING

This method of forecasting by Box-Jenkins implements knowledge of autocorrelation analysis based on autoregressive integrated moving average models. This method is statistically sophisticated in analyzing and identifying a forecasting model that best fit the time series.

The procedure is of four distinct stages namely;

- Identification
- Estimation
- Diagnostic checking
- Forecasting

The first stage in building the model is the identification of the appropriate ARIMA models through the study of the autocorrelation and partial autocorrelation functions.

The next stage is to estimate the parameters of the ARIMA model chosen.

The third stage is the diagnostic checking of the model. The Q statistic is used for the model adequacy check.

If the model is not adequate then the forecaster goes to stage one to identify an alternative model and this is tested for adequacy and if adequate then the forecaster goes to the fourth and final stage of the process.

The fourth stage is where the analysis uses the model chosen to forecast and the process ends.

3.7.2 IDENTIFICATION STAGE

The first step in developing an ARIMA model is to determine if the series is stationary. Stationarity can be accessed from a time series plot and also from the autocorrelation plots. If the model is not stationary, we difference the data or perform the dickey fuller test.

Stationarity could also be achieved by some modes of transformation say log transformation. Once the data becomes stationary, we then identify the order of the model. Autocorrelation plots and Partial Autocorrelation plots are used to identify the order.

Other tools for model identification are

- Akaike Information Criterion (AIC)
- Schwarz's Bayesian Information (BIC)

3.7.3 ESTIMATION STAGE

Once a model is identified the next stage of the ARIMA model building process is to estimate the parameters. The purpose of the estimation is to find the parameter estimates that minimize the mean square error. An iterative non-linear least squares procedure is applied to the parameter estimates of an ARMA (p, q) model. The method minimizes the sum of squares of error given to form the model and data.

3.7.4 MODEL DIAGNOSTIC STAGE

Different models can be obtained for various combinations of AR and MA individually and collectively. The best model is obtained with the following diagnostics:

3.7.5 DIAGNOSTICS OF RESIDUALS

After selection of the best model the following diagnostics of the residuals are made:

- Time Plot of the Residuals
- Plot of Residual ACF
- The Normal Q-Q Plot

3.7.6 SELECTING FROM COMPETING MODELS

Sometimes, we can have a situation where one or more models best fit the series. In such a case, a parsimonious model is chosen first. When two parsimonious models best fit the data, we then go further and choose the model with minimum Akaikes information criteria value and minimum residual variance.

3.7.7 PARSIMONY

Even though both diagnostic procedures describe above aid the analysis to arrive at a forecasting model but neither procedure is considered the ultimate. We then introduce the principle of parsimony. It requires that a forecaster uses few parameters in the model. It also denotes that models should be simple and more accurate with the prediction, everything else being equal. It is also easier to find estimators for parameters in parsimonious models.

3.7.8 FORECASTING

Once we have decided on an appropriate time-series model, estimated its unknown parameters and established that the model fits well, we can turn to the problem of forecasting future values of the series.

Also the relationship between accident, casualties and fatalities were also analysed using the cointegration analysis. If two or more series are individually integrated but some linear combination of them has a lower order of integration, then the series are said to be integrated.

If two or more series are themselves non stationary but a linear combination of them is stationary, then the series are said to be co integrated. Testing for co integration implies testing for a long run relationship between the variables. In this chapter we discuss the various methods of testing for cointegration which includes the ff:

- The two step estimation procedure developed by Engle and granger in 1987.
- The Phillips ouliaris residual based test
- Johansen procedure

3.7.9 THE ENGLE GRANGER TWO STEP PROCEDURE

If two time series x_t and y_t are cointegrated, a linear combination of them must be stationary. We could test for stationarity using the Dickey Fuller test or the kpss test.

The cointegration test can be done in three steps as ff:

- Pretest the variables for the presence of unit root and check if they are integrated of the same order.
- Regress the model knowing the dependent and the independent variables.
- Predict the residuals of the regressed variable and test for the stationarity of the residuals. If the residuals contain a unit root, then it is not stationary, hence no long run relationship between the variables but if there is no unit root, then it indicates the stationarity of the variables, hence long run relationship between them.

CHAPTER 4

RESULTS AND ANALYSIS

This chapter outlines the results and analysis of accident cases, casualties and fatalities of road accidents in Ghana from 1991 to 2010.

The statistical software used was R statistical package. A series of ARIMA models were selected based on the diagnostics of the residuals of each model which include the time plot of the residual, plot of residual ACF, normal QQ plot. Other tools for the model identification are Akaike's information criterion (AIC), and Schwarz's Bayesian criterion (BIC).

Lastly, five years' forecasts were estimated using the best model.

4.1.0 DATA PRESENTATION

The road accident data contains the total number of accident cases, casualties and fatalities in Ghana and was compiled by the National Road Safety Commission (NRSC) and Building and Road Research Institute (BRRI). See appendix for the accident data.

4.1.1 ANALYSIS OF ACCIDENT CASES IN GHANA

DESCRIPTIVE ANALYSIS OF ACCIDENT CASES DATA

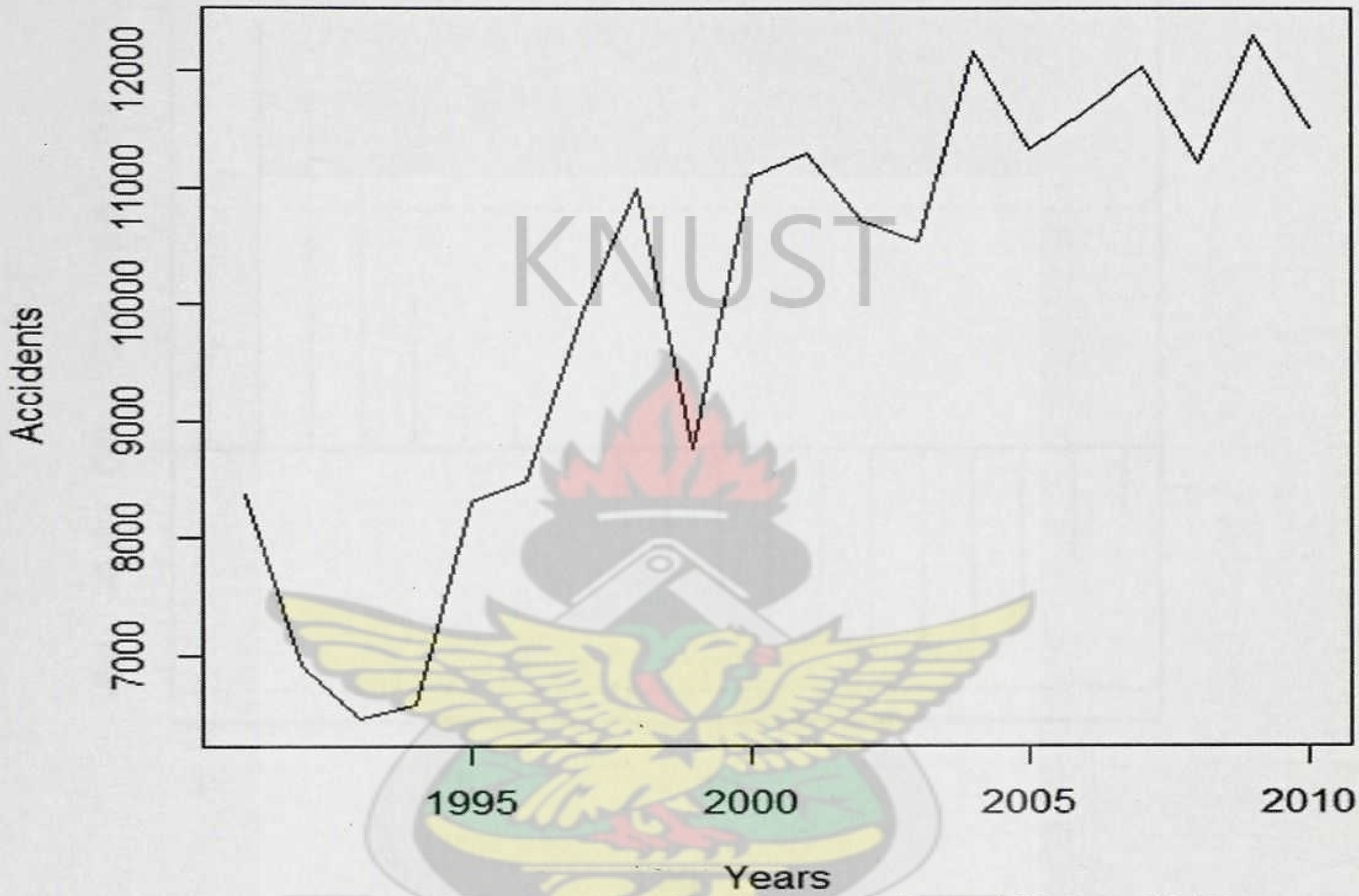

fig 4.1:Trend of accident cases from 1991 to 2010

FIG 4.1 : Time plot of accident cases in Ghana from 1991-2010.

The Figure shows the time plot of accident cases in Ghana from 1991-2010. The time plot shows a trend (when there is a systematic change in the plot).

Accident cases in Ghana decreased from 1991 to 1994 but increased sharply from 1995 to 1998.An irregular and inconsistent pattern was observed from 2000 to 2010. The accident cases were not stationary and do not exhibit seasonal variation. Minimum peak of 6467 accident cases occurred in 1993. In 2009, a maximum peak of accident cases was recorded which amounted to 12299.

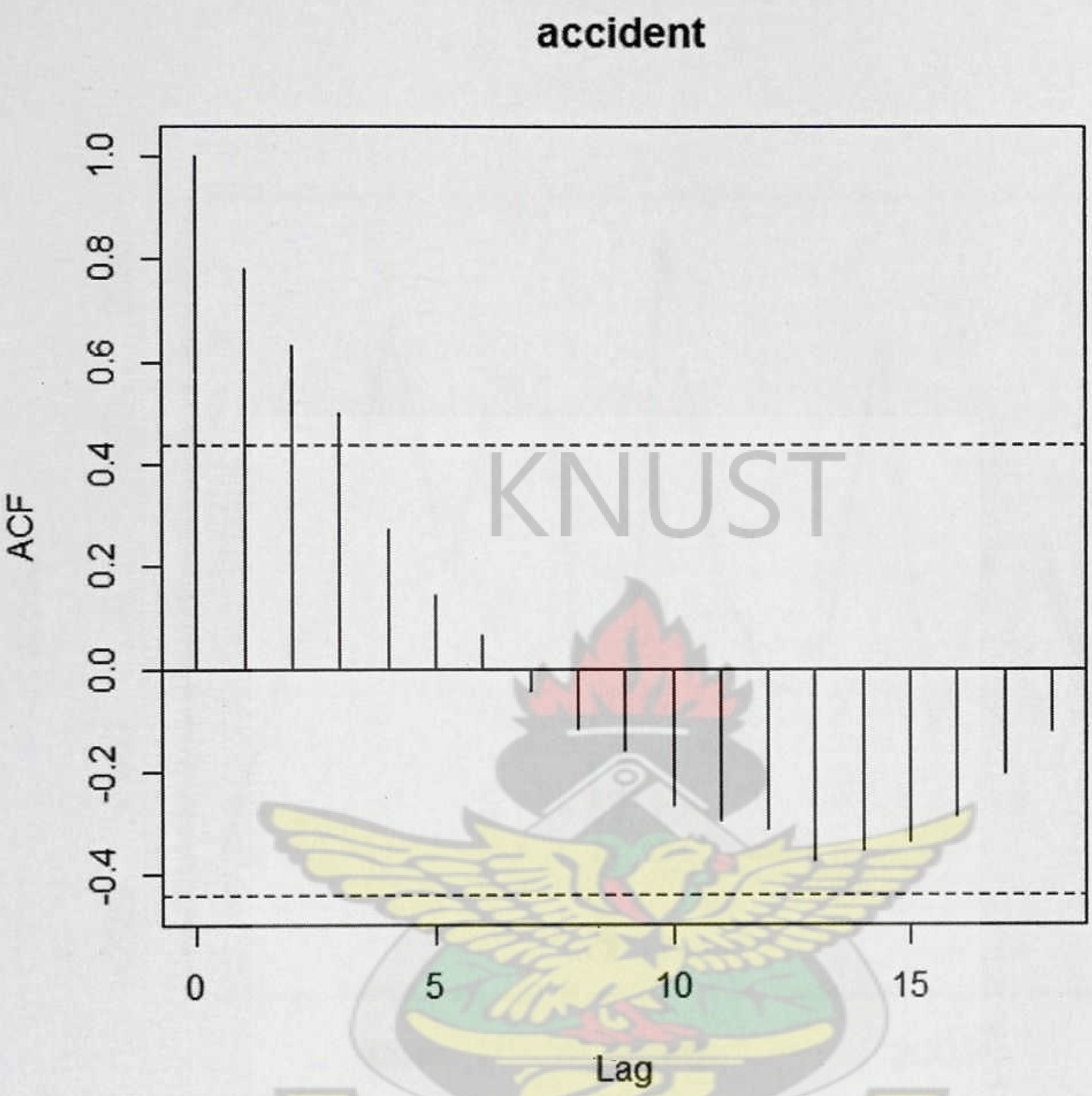

FIG 4.2: Autocorrelation of accident cases in Ghana.

The Autocorrelation function of accident cases in Ghana is shown in Fig 4.2 which describes the correlation between values of the accident cases in Ghana at different point in time. There was a trend and the time series was non-stationary.

4.1.2 FIRST DIFFERENCING OF THE ACCIDENT DATA CASE

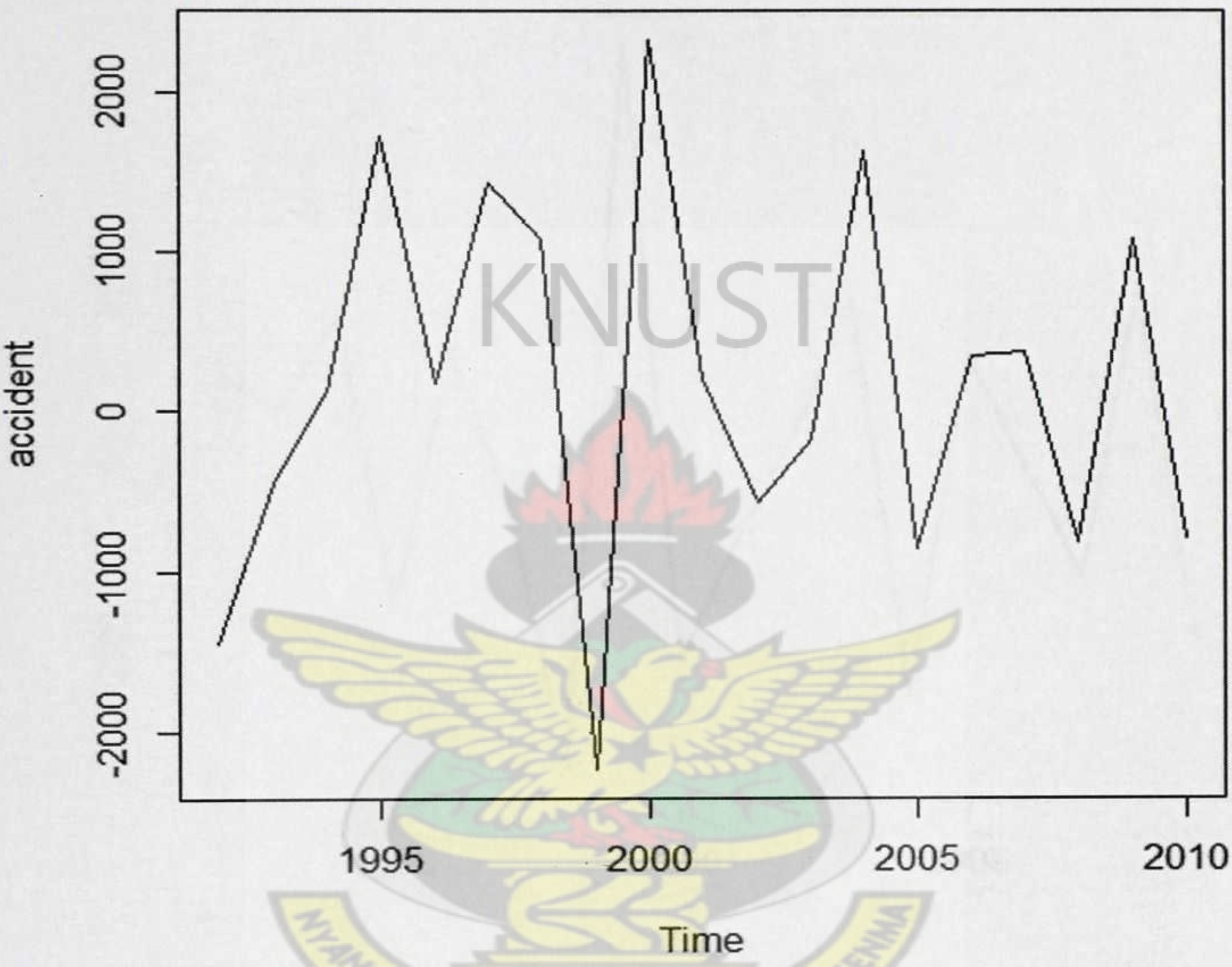

FIG 4.3 : First difference of Accident cases in Ghana.

First differencing was performed to remove trend component in the original data.

4.1.3 SECOND DIFFERENCING OF THE ACCIDENT DATA CASES

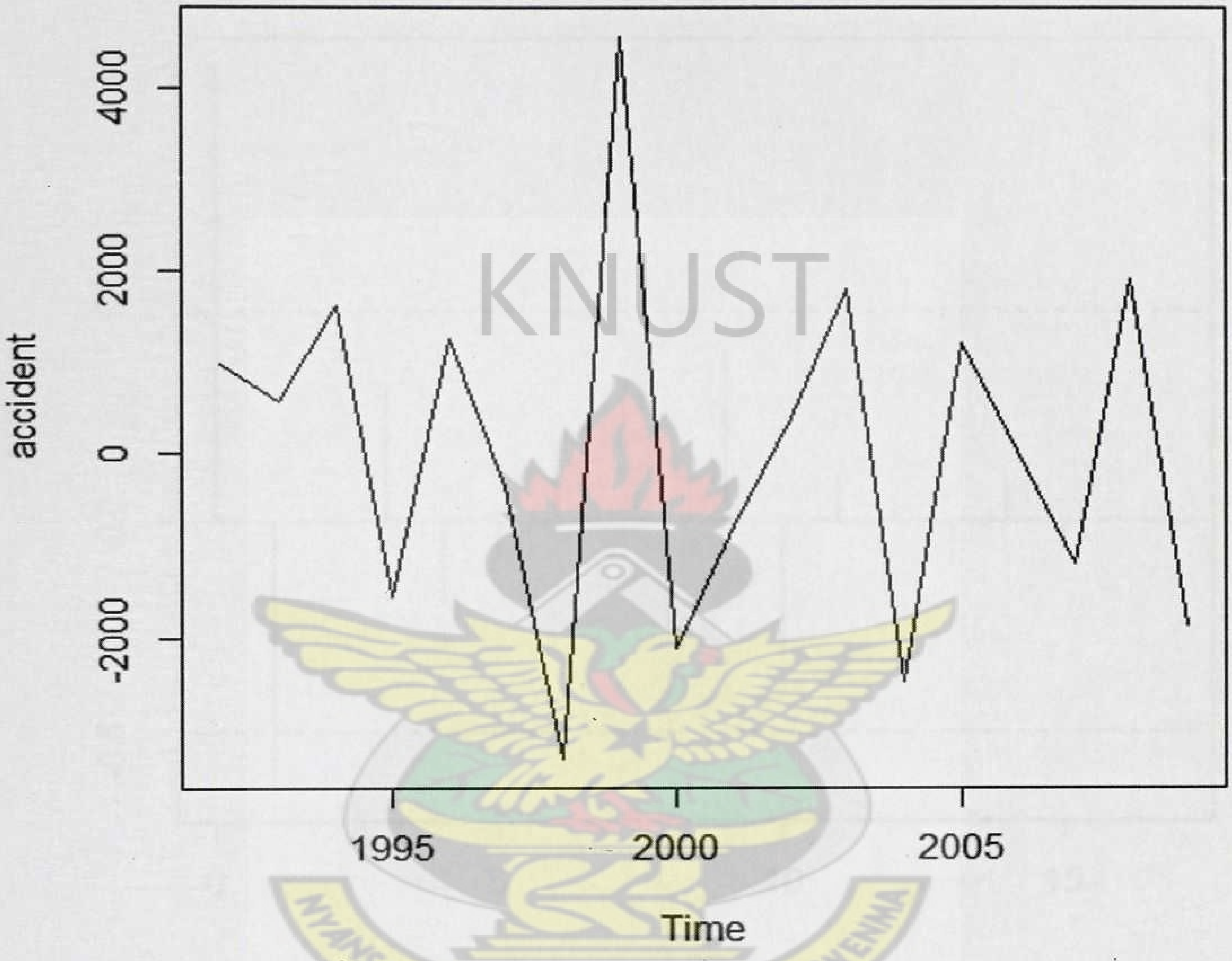

FIG 4.4 : Second differencing of accident cases in Ghana.

Second differencing was ~~done to~~ achieve better stationarity. The variability was approximately constant and the accident cases in Ghana now looks to be approximately stable.

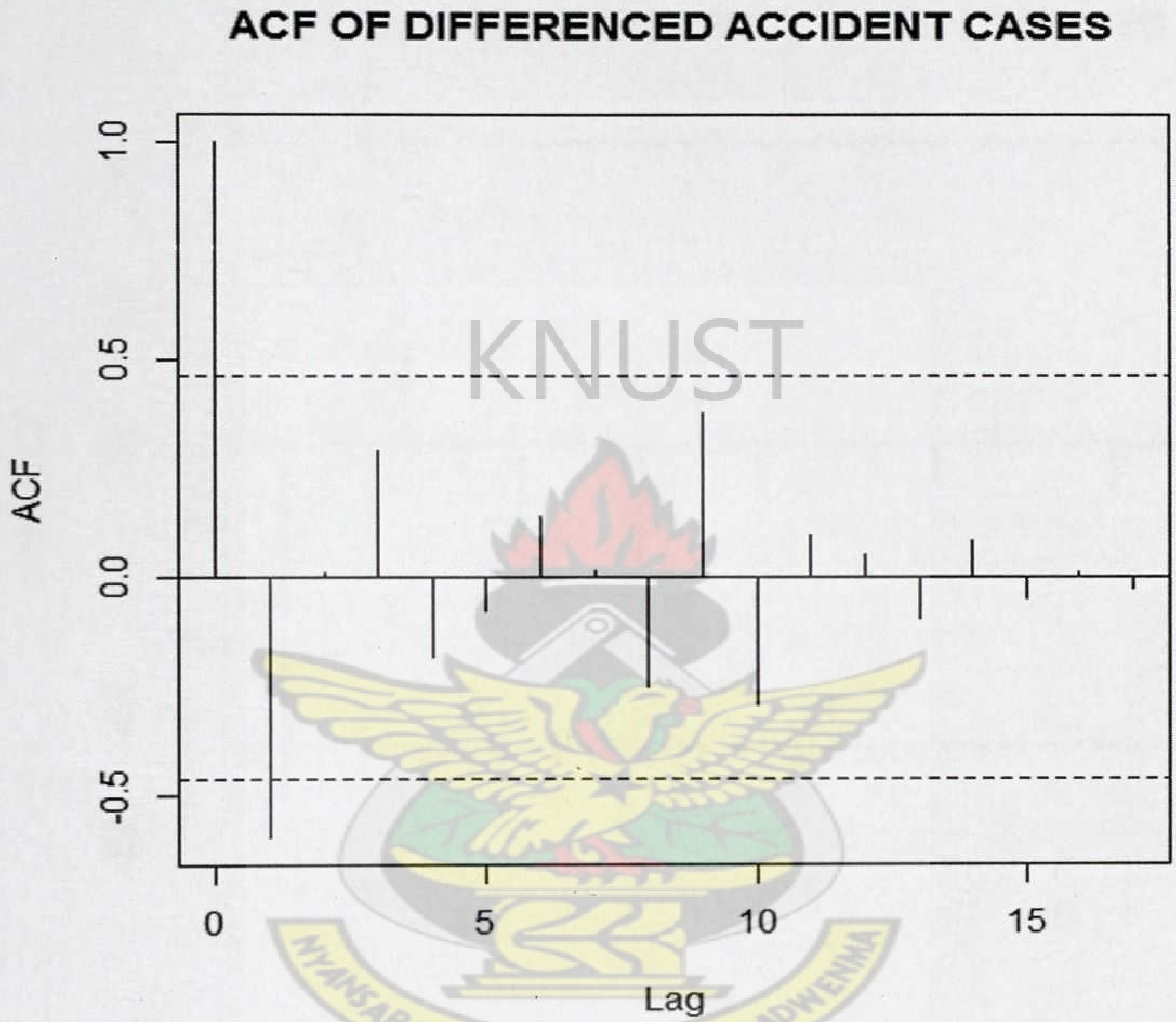

FIG 4.5 : ACF plots of the second differencing of accident cases in Ghana.

From the Autocorrelation function plots, we could see that only lag 1 was significant meaning a moving average of 1 (MA 1).

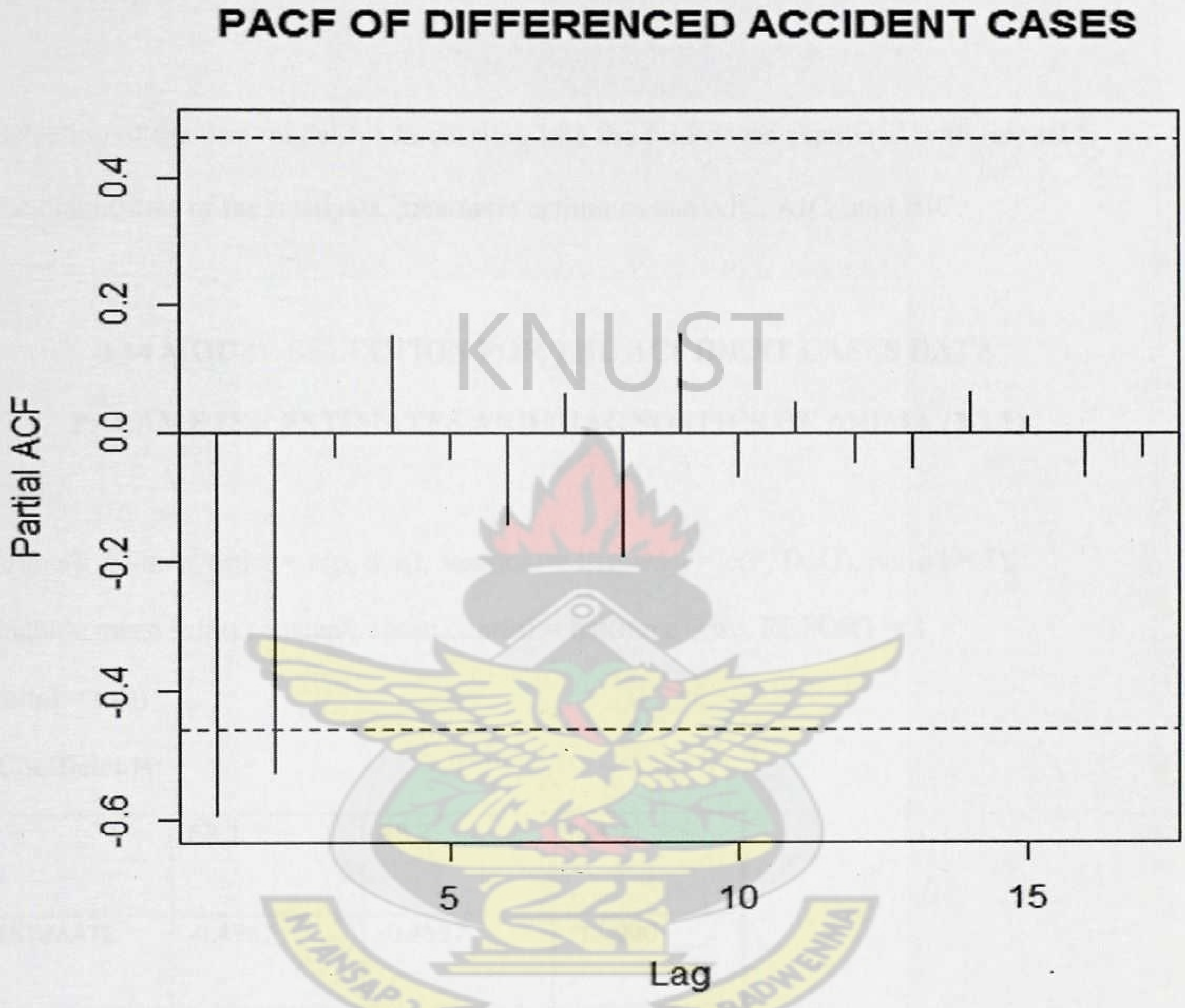

FIG 4.51 : PACF plots of the second differencing of accident cases in Ghana.

Comparing the partial Autocorrelation function with the error limits, we could see that lag 1 and lag 2 were significant meaning an Autoregressive model of 1 and 2 (AR 1 and AR 2) but we choose AR (2).

The following models were suggested.

ARIMA (2,2,1)

ARIMA (2,2,0)

ARIMA (0,2,1)

Selection of the best model for forecasting into the future was examined with respect to the diagnostics of the residuals, parameter estimates and AIC, AIC_C and BIC.

4.14 MODEL SELECTION FOR THE ACCIDENT CASES DATA

PARAMETER ESTIMATES AND DIAGNOSTICS OF ARIMA (2,2,1)

Call:

arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S),

include.mean = !no.constant, optim.control = list(trace = trc, REPORT = 1,

reltol = tol))

Coefficients:

	AR 1	AR 2	MA 1
ESTIMATE	-0.4997	-0.2597	-1.0000
S.E	0.2539	0.2493	0.2764
T-VALUE	1.97	1.04	3.62

sigma^2 estimated as 1073010: log likelihood = -152.68, aic = 313.35

$AIC

[1] 15.18598

AICc

[1] 15.41931

BIC

[1] 14.33534

The parameters based on the t-value estimate in MA 1 was statistically significant since the t-value was greater than 2 but the parameters in AR 1 and AR 2 were not significant.

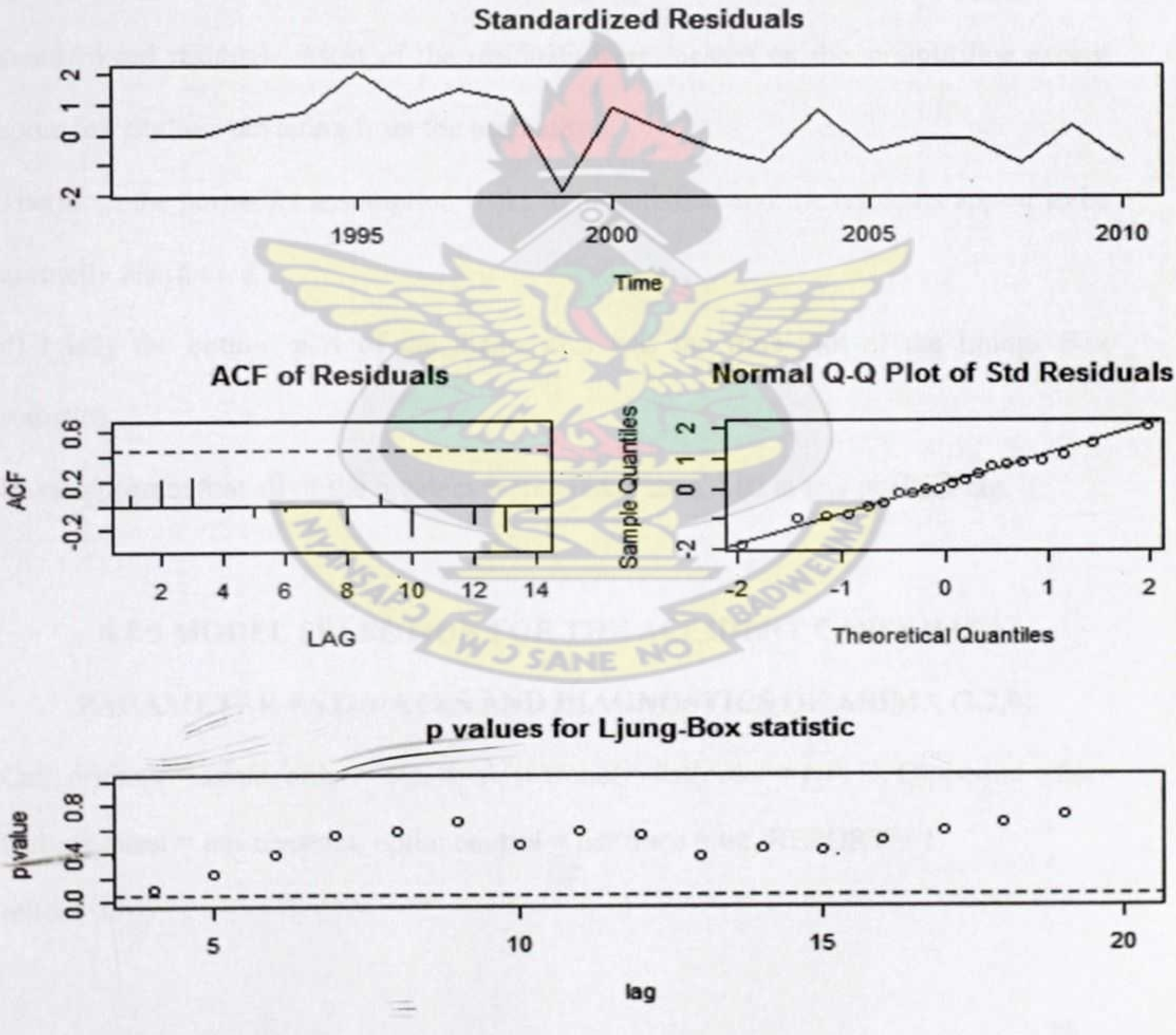

Fig 4.52: Diagnostics of residuals of ARIMA (2,2,1)

Diagnostics of the residuals from ARIMA (2, 2, 1) is shown in the Figure above.

a) The time plot of the standardized residuals of ARIMA (2, 2, 1) appears at the upper part. It contains a few outliers with a mean of zero and is independent and identical distribution. The standardized residuals plot shows no obvious trend and pattern.

b) The plot of the ACF of the residuals of the diagnostics appears at the middle part and we could see that at any positive lag, there was no evidence of significant correlation in the residuals.

c) At the right side of the middle of the diagnostics is the normal Q – Q plot of the standardized residuals. Most of the residuals were located on the straight line except some few outliers deviating from the normality.

Therefore the normality assumption looks to be satisfied and the residuals appear to be normally distributed.

d) Lastly the bottom part of the diagnostics was the time plot of the Ljung- Box statistics.

It was apparent that all of the p values were greater than 0.05 at any positive lag.

4.1.5 MODEL SELECTION FOR THE ACCIDENT CASES DATA

PARAMETER ESTIMATES AND DIAGNOSTICS OF ARIMA (2,2,0)

Call: arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S), include.mean = !no.constant, optim.control = list(trace = trc, REPORT = 1, reltol = tol))

Coefficients:

	AR 1	AR 2
ESTIMATE	-0.9459	-0.5514
S.E	0.1911	0.1863
T-VALUE	4.95	2.96

sigma^2 estimated as 1489656: log likelihood = -154.06, aic = 314.12

$AIC

[1] 15.41406

$AICc

[1] 15.58906

$BIC

[1] 14.51363

The parameters based on the t-value estimate in AR 1 and AR 2 were statistically significant since the t-value were greater than 2.

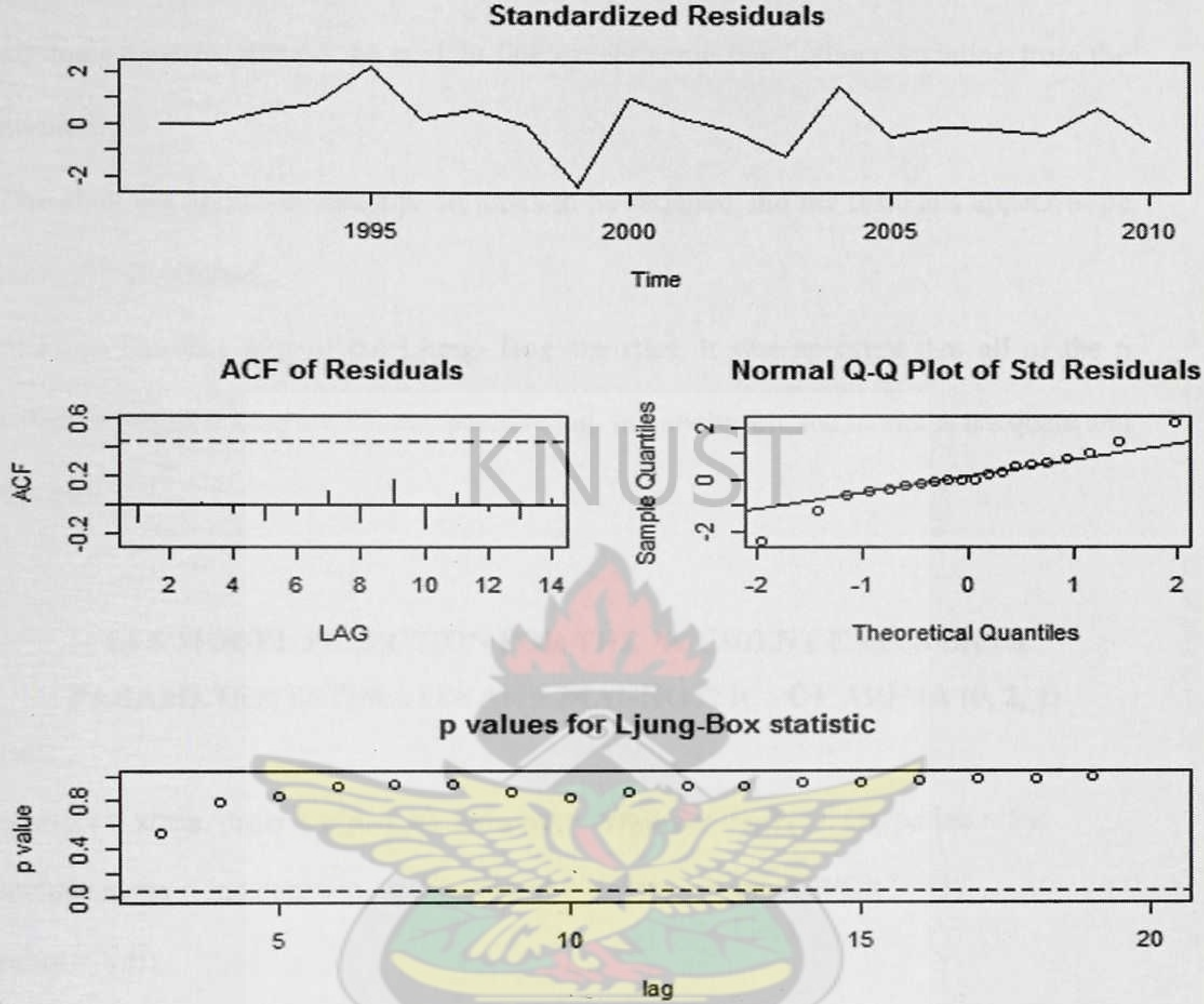

FIG 4.53: Diagnostics of residuals of ARIMA (2,2,0)

Diagnostics of the residuals from ARIMA (2, 2, 0) is shown in the Figure above.

a) The standardized residual plot contains a few outliers with a mean of zero and is independent and identical distribution. It shows no obvious trend and pattern.

b) The plot of the ACF of the residuals of the diagnostics appears at the middle part and we could see that at any positive lag, there was no evidence of significant correlation in the residuals.

c) The normal Q – Q plot of the standardized residuals is shown above. Most of the residuals were located on the straight line except some few outliers deviating from the normality.

Therefore the normality assumption looks to be satisfied and the residuals appear to be normally distributed.

d) From the time plot of the Ljung- Box statistics, It was apparent that all of the p values are greater than 0.05 at any positive lag. In conclusion, the model is adequate and fits well.

4.1.6 MODEL SELECTION FOR THE ACCIDENT CASES DATA PARAMETER ESTIMATES AND DIAGNOSTICS OF ARIMA (0, 2, 1)

call:

arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S), include.mean = !no.constant, optim.control = list(trace = trc, REPORT = 1, reltol = tol))

Coefficients:

	MA 1
ESTIMATE	-1.000
S.E	0.1563
T-VALUE	6.40

sigma^2 estimated as 1386739: log likelihood = -154.3, aic = 312.59

$AIC

[1] 15.24247

$AICc

[1] 15.37776

$BIC

[1] 14.29225

The parameters based on the t-value estimate in MA 1 was statistically significant since the t-value was greater than 2.

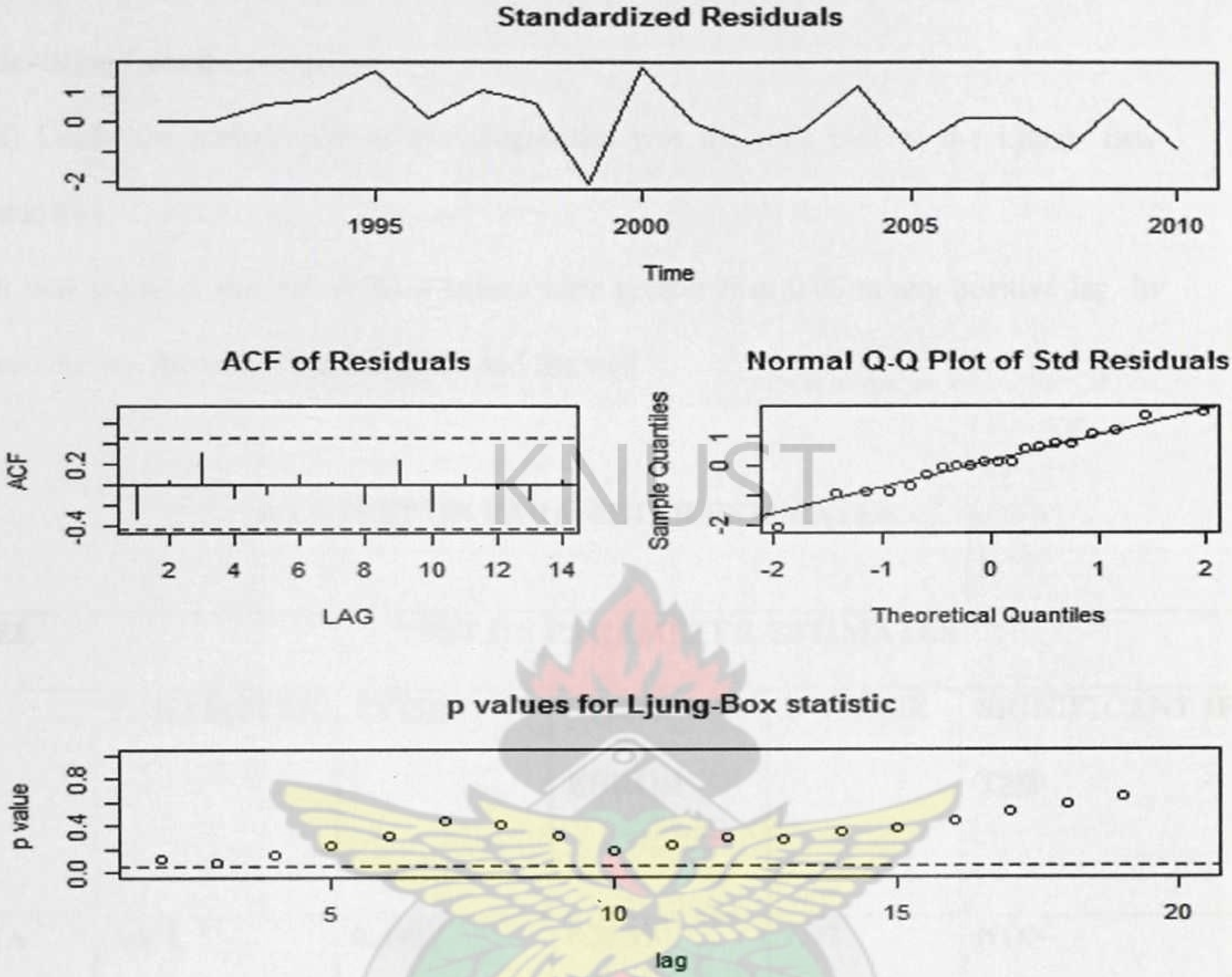

FIG 4.54: Diagnostics of residuals of ARIMA (0, 2, 1)

Diagnostics of the residuals from ARIMA (0, 2, 1) is shown in the Figure above.

a) The standardized residuals plot shows no obvious trend and pattern and looks like an independent and identical distribution.

b) The plot of the ACF of the residuals of the diagnostics appears at the middle part and we could see that at any positive lag, there was no evidence of significant correlation in the residuals.

c) Most of the residuals were located on the straight line except some few outliers deviating from the normality.

d) Lastly the bottom part of the diagnostics was the time plot of the Ljung- Box statistics.

It was apparent that all of the p values were greater than 0.05 at any positive lag. In conclusion, the model was adequate and fits well.

4.1.7 TEST ON PARAMETER ESTIMATES

MODEL	TEST ON PARAMETER ESTIMATES				
	PARAMETER	ESTIMATE	STANDARD ERROR	T-VALUE	SIGNIFICANT IF T≥|2|
ARIMA (2,2,1)	AR 1	0.4997	0.2539	1.97	NON-SIGNIFICANT
	AR 2	0.2597	0.2493	1.04	NON-SIGNIFICANT
	MA 1	1.000	0.2764	3.62	SIGNIFICANT
ARIMA (2,2,0)	AR 1	0.9459	0.1911	4.95	SIGNIFICANT
	AR 2	0.5514	0.1863	2.96	SIGNIFICANT
ARIMA (0,2,1)	MA 1	1.000	0.1563	6.40	SIGNIFICANT

DIAGNOSTICS		
	ARIMA(2,2,0)	ARIMA(0,2,1)
RESIDUAL VARIANCE	1489656	1386739
AIC	314.12	312.59
AIC_C	15.19	15.15
BIC	14.51	14.30

4.1.8 SELECTION OF BEST MODEL FOR FORECASTING ACCIDENT CASES IN GHANA.

The standardized residuals plots of all the models were independently and identically distributed with mean zero and some few outliers. There was no evidence of significance in the autocorrelation functions of the residuals of all the models and the residuals appear to be normally distributed in all the models. The Ljung – Box statistics were not significant at any positive lag for all the models.

From the table above all the parameters in the coefficients of ARIMA(2,2,0) and ARIMA(0,2,1) models were significant except ARIMA (2,2,1).

The AIC, AIC_C, BIC and residual variance were good for all the models but they favor ARIMA (0, 2, 1) model. Since ARIMA (2,2,1) was not significant, I compared the parameters of ARIMA (2,2,0) and ARIMA (0,2,1)

Comparing the AIC of the models, we could observe that ARIMA (0,2,1) had the minimum..

The best model was chosen based on the minimum residual variance and we could observe that for all the models, ARIMA (0,2,1) has the minimum variance.

From the discussion above it was clear that ARIMA (0, 2, 1) model was the best model for forecasting accident cases in Ghana.

4.1.9 FITTING THE ACCIDENT CASES MODEL

ARIMA (0,2,1) model was the best model for forecasting accident cases in Ghana. It has two level of differencing without an AR part. In terms of the observed series, the model becomes:

$$y_t - 2y_{t-1} + y_{t-2} = \alpha + \varepsilon_t - \theta_1 \varepsilon_{(t-1)}$$

The point estimate of the parameter of ARIMA (0,2,1) is $\theta_1 = -1.0000$

All the estimates were significant except the constant term, hence it was dropped.

The fitted ARIMA (0,2,1) model for forecasting accident cases from 1991-2010 was given by

$$\hat{y}_t = 2y_{t-1} - y_{t-2} + \varepsilon_t + \varepsilon_{(t-1)}$$

Where ε_t has an estimated variance of 1386739

4.2.0 FORECASTING ACCIDENT CASES IN GHANA

5 steps prediction into the future;

$pred

Time Series:

Start = 2011

End = 2015

Frequency = 1

[1] 11671.05 11836.11 12001.16 12166.22 12331.27

$se

Time Series:

Start = 2011

End = 2015

Frequency = 1

[1] 1208.190 1750.833 2194.786 2591.278 2959.448

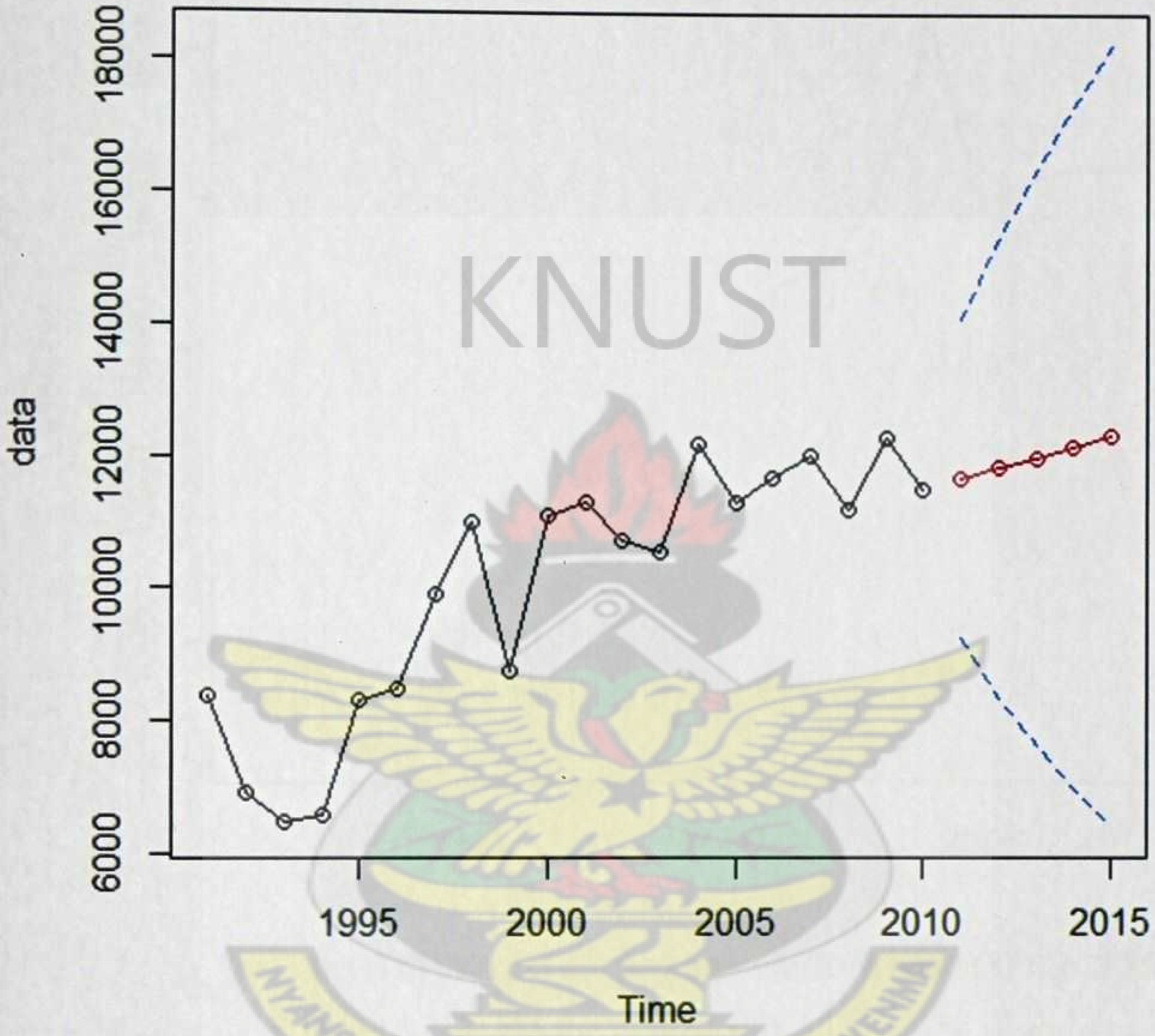

Figure 4.5.5 : Graph of the accident cases, its forecasts and confidence intervals

The Figure gives the visual representation of the accident cases data (black line), its forecasts (red line) and confidence interval (blue short dashes lines).

From the prediction values and the graph above, it can be observed that, accident cases in Ghana would continue to increase in the next 5 years.

4.2.1 ANALYSIS OF CASUALTY CASES IN GHANA

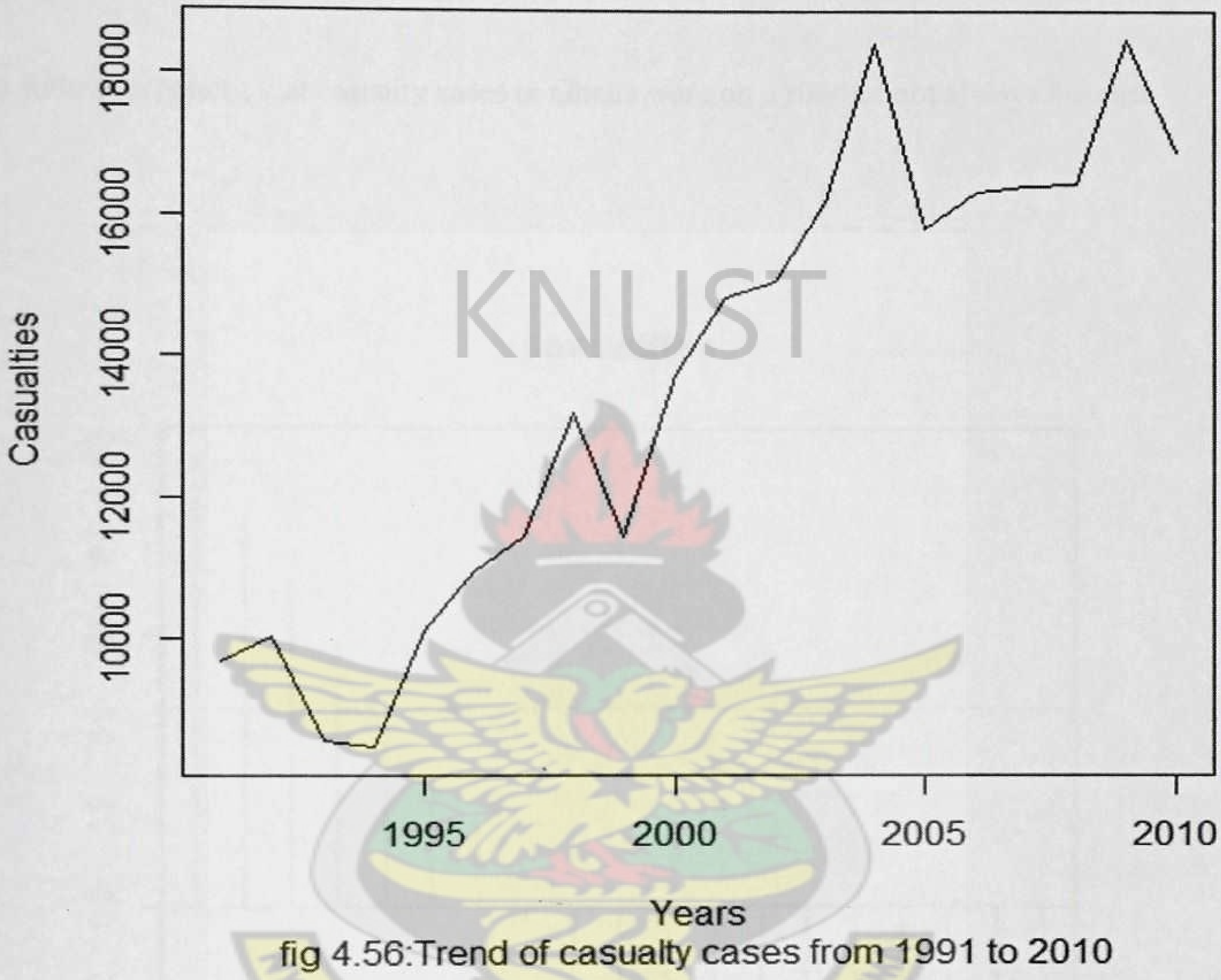

FIG 4.56 : Time plot of casualty cases in Ghana from 1991-2010.

The Figure shows the time plot of casualty cases in Ghana from 1991 to 2010. The time plot was not periodic and there was a systematic change known as a trend.

Casualty cases in Ghana increased from 1991 to 1992 followed by decrease in 1993 to 1994.

There was a sharp increase from 1995 to 1998 but reverted to a decrease in 1999. It was then followed by an increase in 2000 to 2004. Casualty cases rose from 2005 to 2009 and lastly declined in 2010.

It follows a pattern that casualty cases in Ghana were on a rise but not always the case

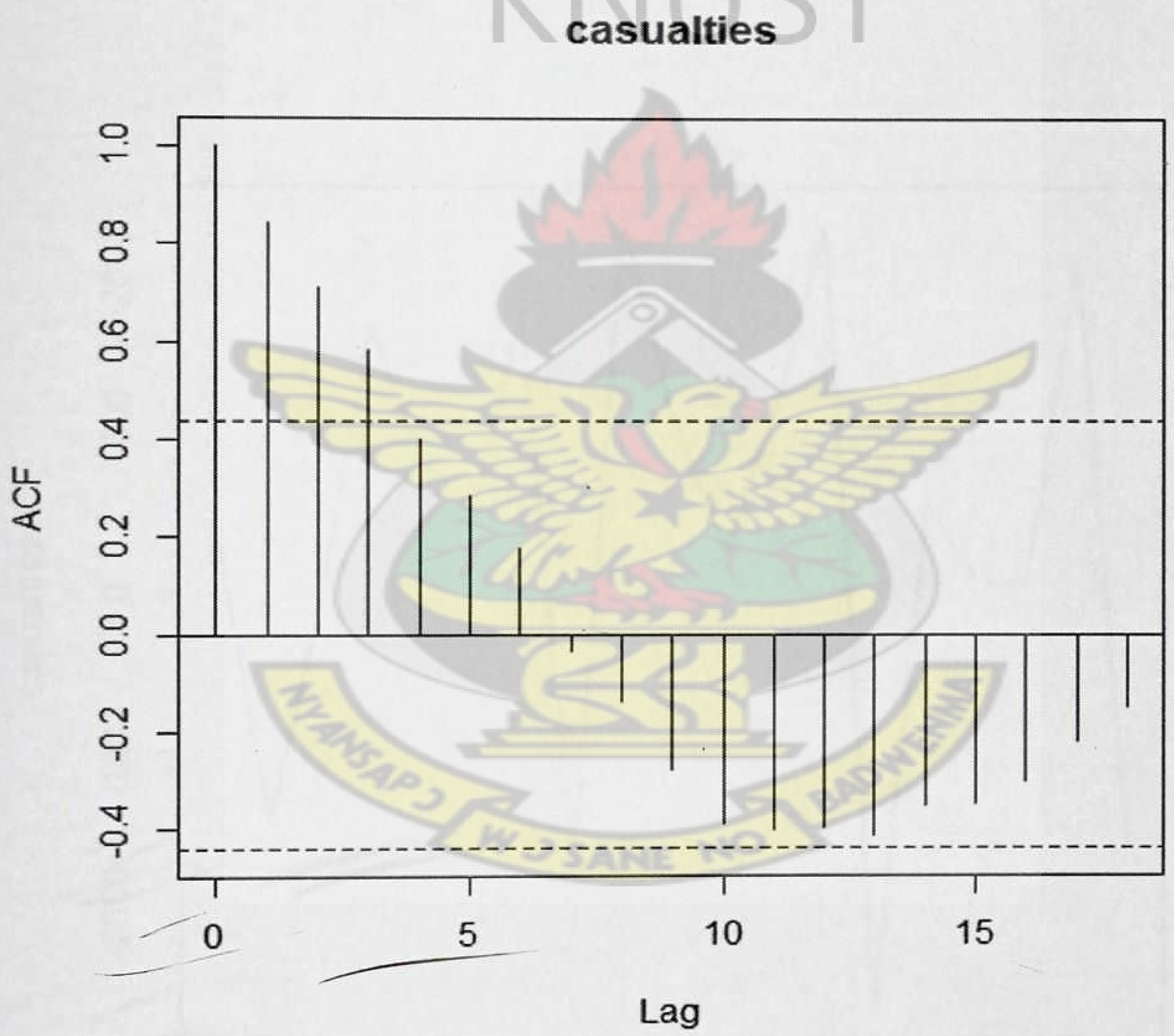

FIG 4.57 : Autocorrelation function of casualty cases in Ghana from 1991 to 2010.

The Autocorrelation function of casualty cases in Ghana is shown in Fig 4.57 which describes the correlation between values of the casualty cases in Ghana at different point in time.

There was a trend and the time series was non-stationary. It was apparent that autocorrelatiom function was increasing gradually.

4.2.2 FIRST DIFFERENCING OF THE CASUALTY DATA CASES

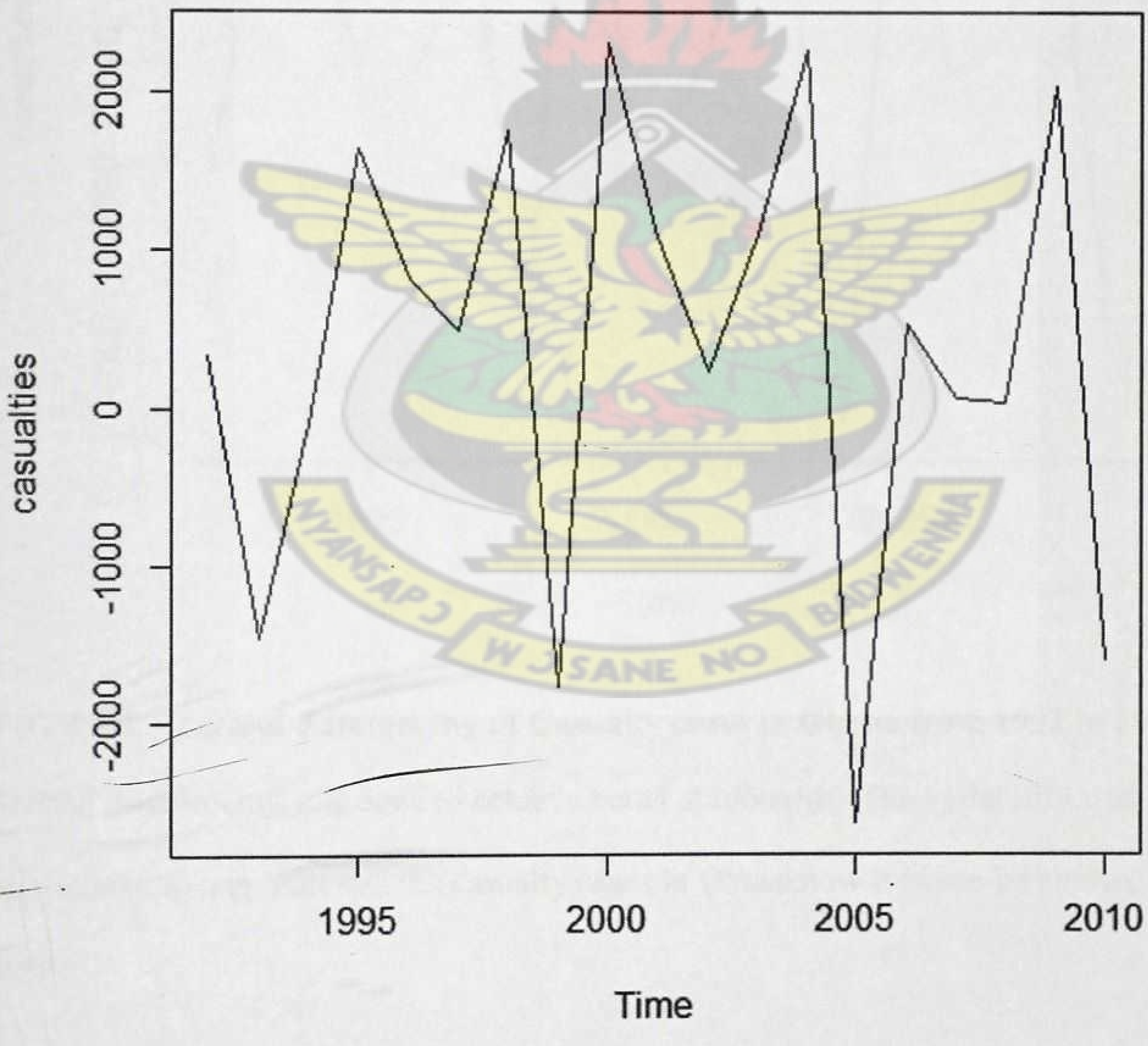

FIG 4.58 : First difference of casualty cases in Ghana.

First differencing was performed to remove trend component in the original data.

4.2.3 SECOND DIFFERENCING OF THE CASUALTY DATA CASES

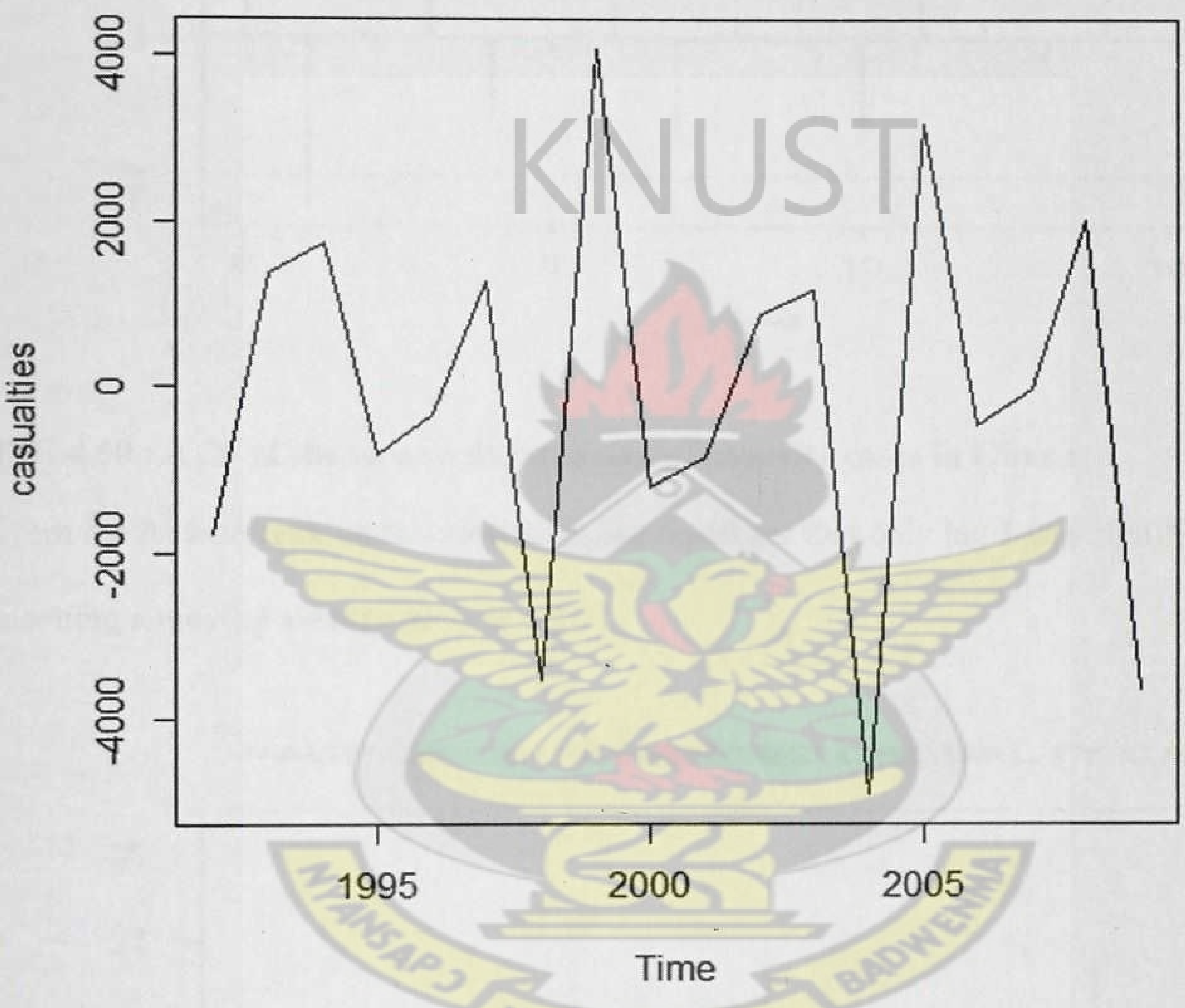

FIG 4.5.9 : Second differencing of Casualty cases in Ghana from 1991 to 2010.

Second differencing was done to achieve better stationarity. The variability was approximately constant and the casualty cases in Ghana now looks to be approximately stable.

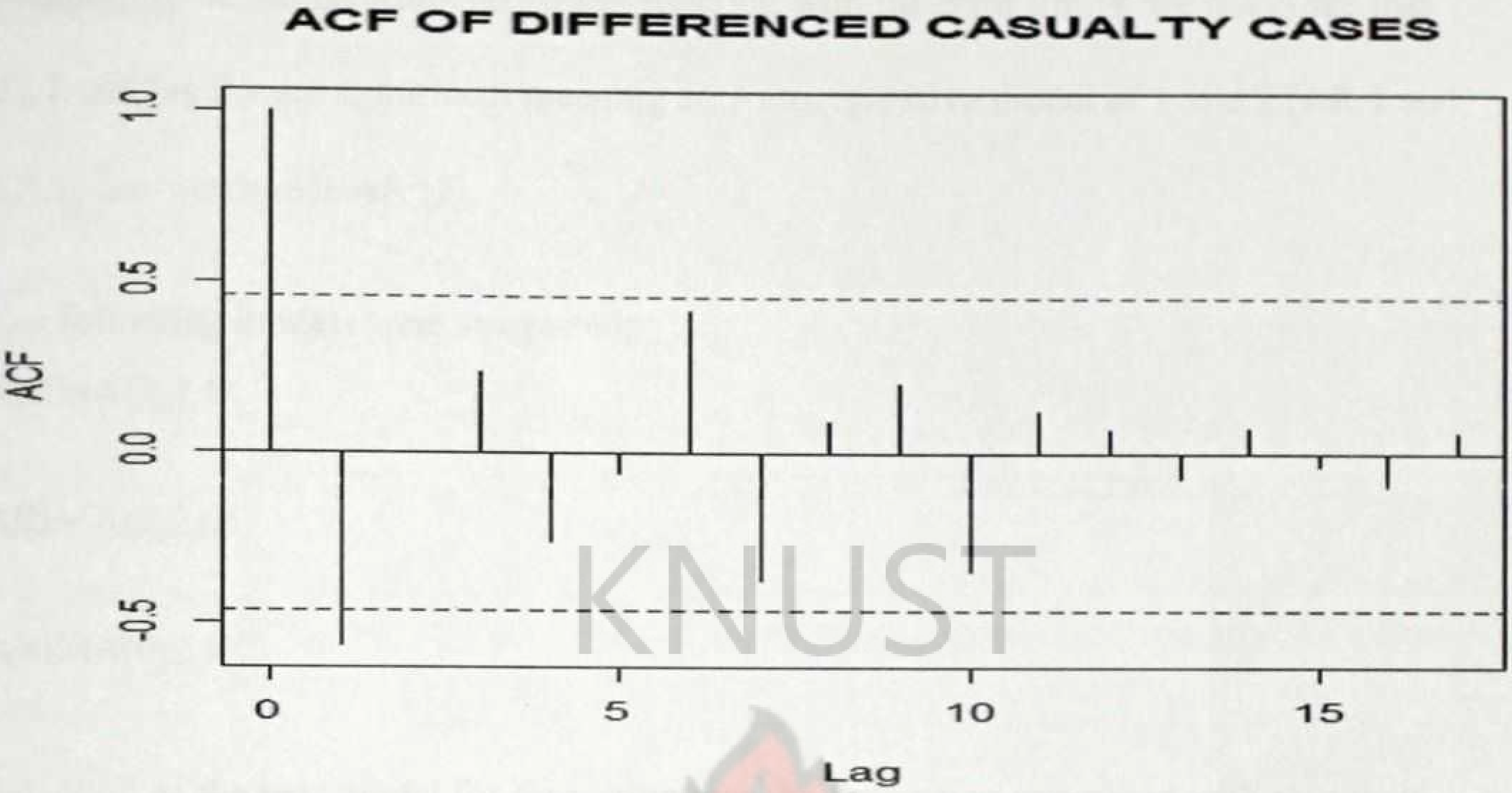

FIG 4.60 : ACF of the second differencing of casualty cases in Ghana.

From the Autocorrelation function plots, we could see that only lag 1 was significant meaning a moving average of 1 (MA 1).

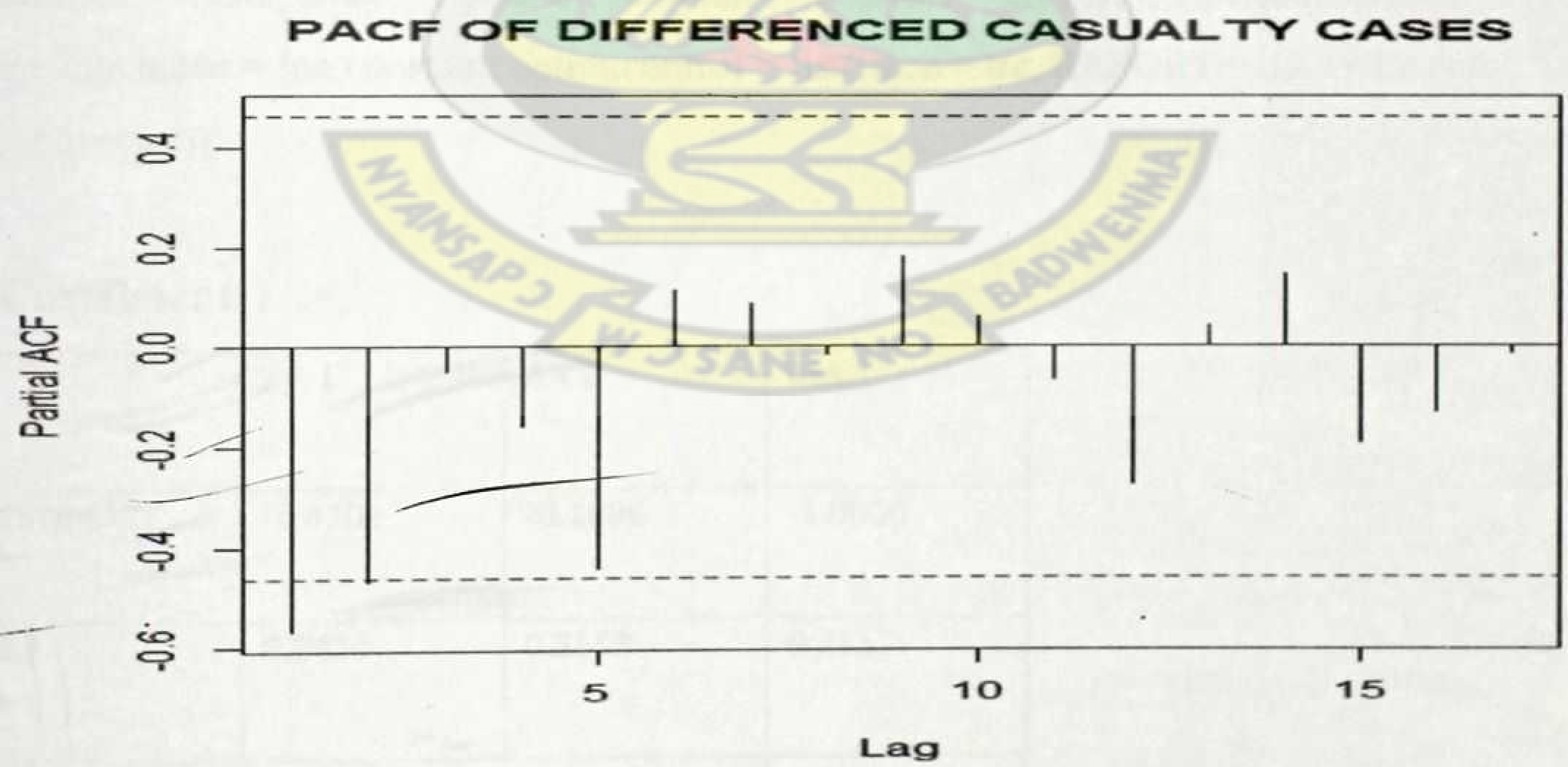

FIG 4.61 : PACF of the second differencing of casualty cases in Ghana.

Comparing the partial Autocorrelation function with the error limits, we could see that lag 1 and lag 2 were significant meaning an Autoregressive model of 1 and 2 (AR 1 and AR 2) but we choose AR (2).

The following models were suggested;
ARIMA(2,2,1)

ARIMA(2,2,0)

ARIMA(0,2,1)

Selection of the best model for forcasting into the future was examined with respect to the diagnostics of the residuals, parameter estimates and AIC, AIC_C and BIC.

4.2.4 MODEL SELECTION FOR THE CASUALTY CASES DATA PARAMETER ESTIMATES AND DIAGNOSTICS OF ARIMA (2,2,1)

call:
arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S),
include.mean = !no.constant, optim.control = list(trace = trc, REPORT = 1,
reltol = tol))

Coefficients:

	AR 1	AR 2	MA 1
ESTIMATE	-0.4301	-0.1886	-1.0000
S.E	0.2410	0.2553	0.2352
T-VALUE	1.78	0.74	4.25

sigma^2 estimated as 1620075: log likelihood = -156.26, aic = 320.51

$AIC

[1] 15.59798

$AICc

[1] 15.83132

$BIC

[1] 14.74734

The parameters based on the t-value estimate in MA 1 was statistically significant since the t-value was greater than 2 but the parameters in AR 1 and AR 2 were not significant.

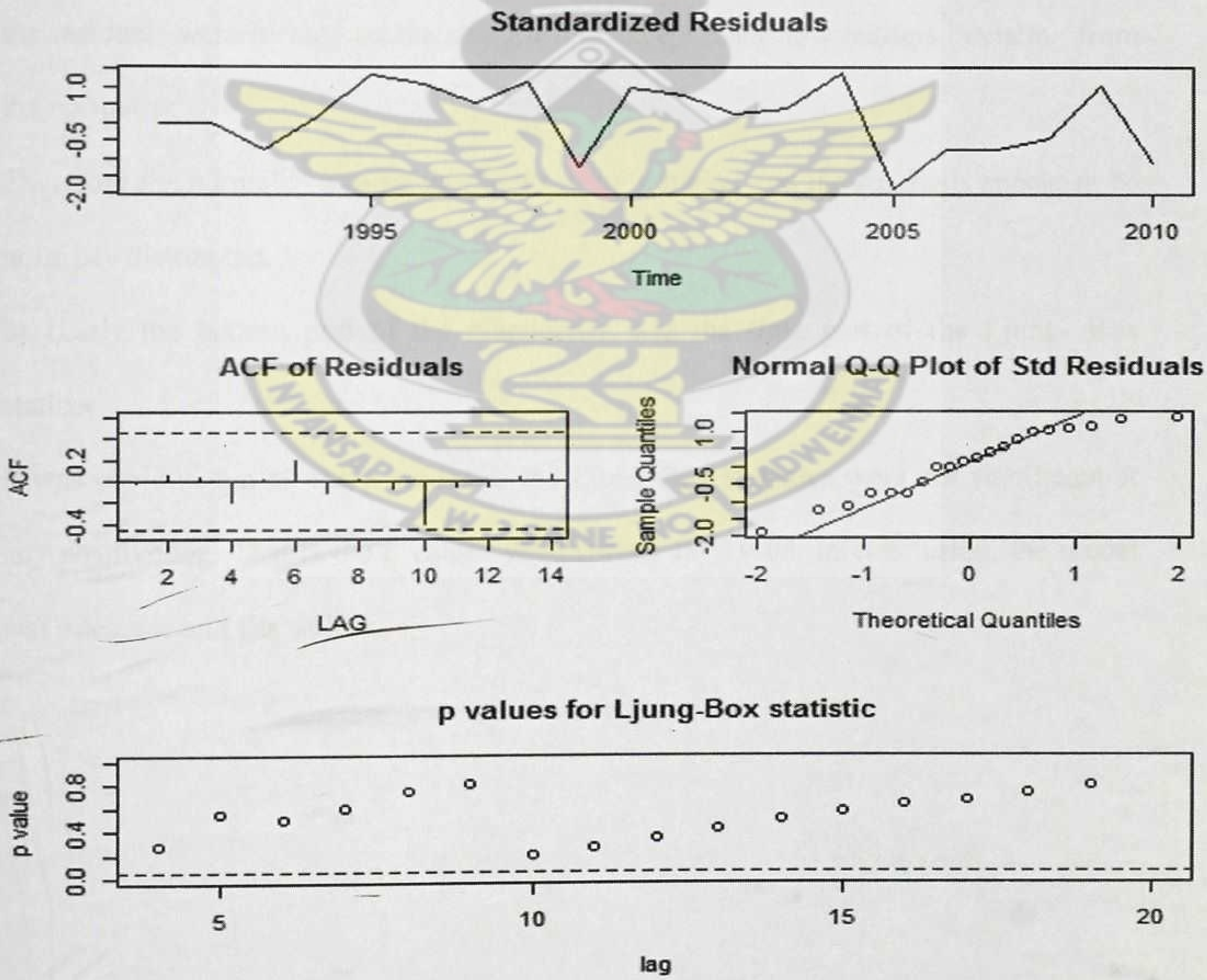

FIG : 4.6.2 Diagnostics of the residuals from ARIMA (2, 2, 1).

a) The time plot of the standardized residuals of ARIMA (2, 2, 1) appears at the top. The standardized residuals plot shows no obvious trend and pattern.

It contains a few outliers with a mean of zero and is independent and identical distribution.

b) At the middle was the plot of the ACF of the residuals of the diagnostics and we could see that at any positive lag, there was no evidence of significant correlation in the residuals.

c) The normal Q – Q plot of the standardized residuals appears at the right side. Most of the residuals were located on the straight line except some few outliers deviating from the normality.

Therefore the normality assumption looks to be satisfied and the residuals appear to be normally distributed.

d) Lastly the bottom part of the diagnostics was the time plot of the Ljung- Box statistics.

It was obvious that all the p values of the Ljung-Box statistics were not significant at any positive lag. That is the p values were greater than 0.05. In conclusion, the model was adequate and fits well

4.2.5 MODEL SELECTION FOR THE CASUALTY CASES DATA

PARAMETER ESTIMATES AND DIAGNOSTICS OF ARIMA (2,2,0)

Call:

arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S), include.mean = !no.constant, optim.control = list(trace = trc, REPORT = 1, reltol = tol))

Coefficients:

	AR 1	AR 2
ESTIMATE	-0.9846	-0.5626
S.E	0.1967	0.1908
T-VALUE	5.00	2.95

sigma^2 estimated as 2166926: log likelihood = -157.47, aic = 320.95

$AIC

[1] 15.78882

$AICc

[1] 15.96382

$BIC

[1] 14.88839

The parameters based on the t-value estimate in AR 1 and AR 2 were statistically significant since the t-value were greater than 2.

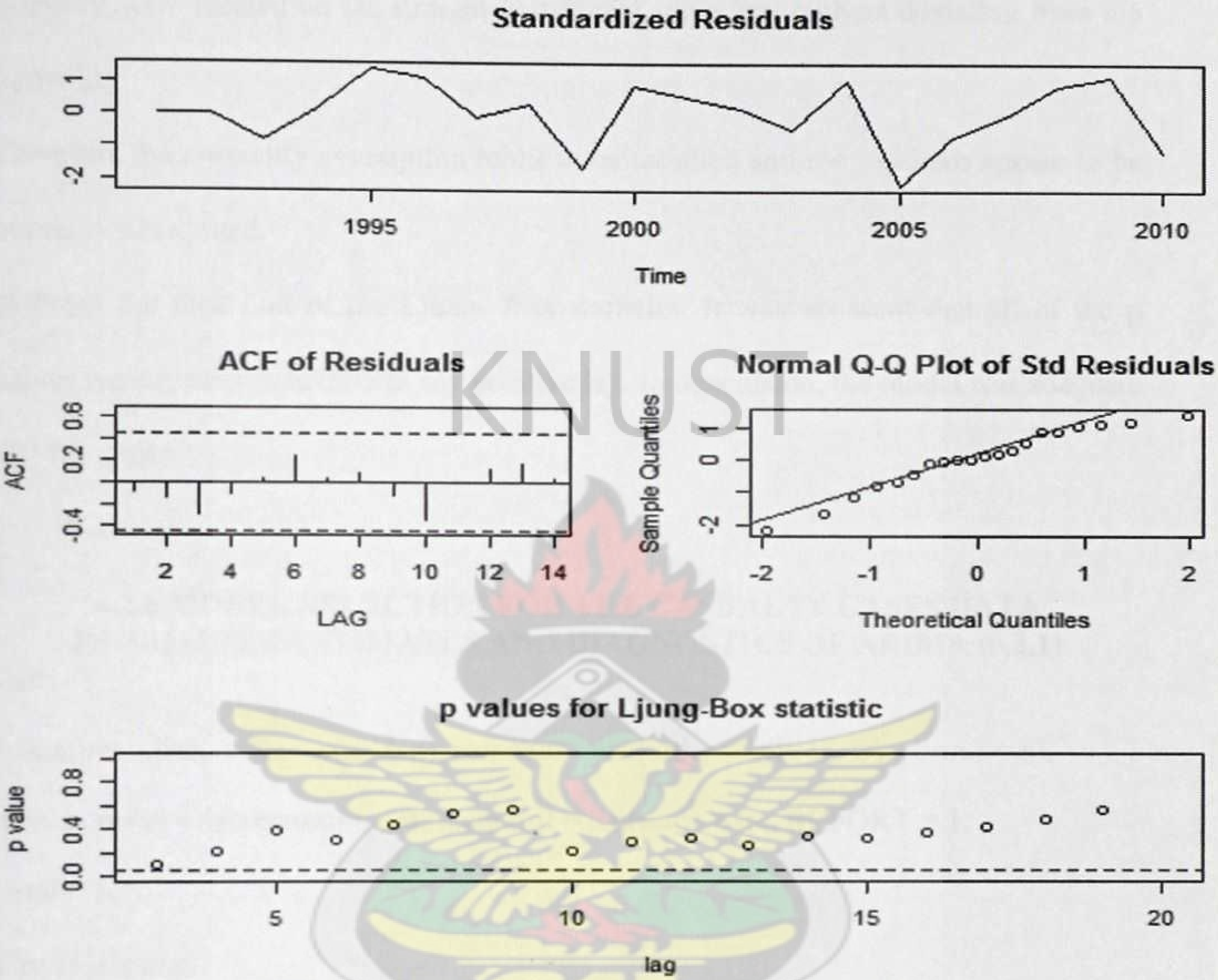

FIG 4.6.3: Diagnostics of the residuals from ARIMA (2, 2, 0).

Diagnostics of the residuals from ARIMA (2, 2, 0) is shown in the Figure above.

a) The standardized residual plot contains a few outliers with a mean of zero and is independent and identical distribution. It shows no obvious trend and pattern.

b) The plot of the ACF of the residuals of the diagnostics appears at the middle part and we could see that at any positive lag, there was no evidence of significant correlation in the residuals.

c) The normal Q – Q plot of the standardized residuals is shown above. Most of the residuals were located on the straight line except some few outliers deviating from the normality.

Therefore the normality assumption looks to be satisfied and the residuals appear to be normally distributed.

d) From the time plot of the Ljung- Box statistics, It was apparent that all of the p values were greater than 0.05 at any positive lag. In conclusion, the model was adequate and fits well.

4.2.6 MODEL SELECTION FOR THE CASUALTY CASES DATA PARAMETER ESTIMATES AND DIAGNOSTICS OF ARIMA (0,2,1)

Call:

arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S),

include.mean = !no.constant, optim.control = list(trace = trc, REPORT = 1,

reltol = tol))

Coefficients:

	MA 1
ESTIMATE	-1.0000
S.E	0.1592
T-VALUE	6.28

sigma^2 estimated as 2008315: log likelihood = -157.63, aic = 319.26

$AIC

[1] 15.61281

$AICc

[1] 15.7481

$BIC

[1] 14.66259

The parameters based on the t-value estimate in MA 1 was statistically significant since the t-value was greater than 2.

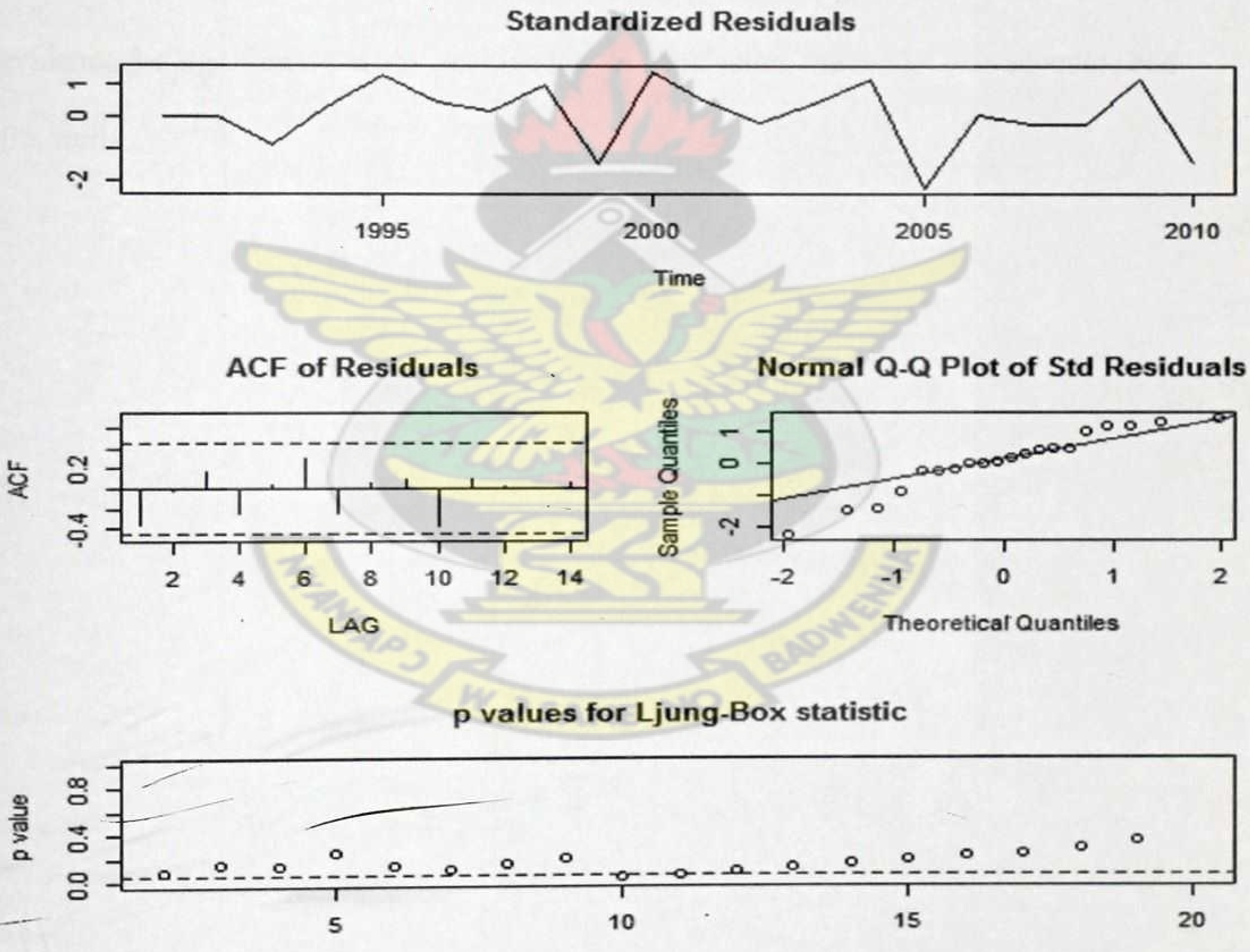

FIG 4.6.4 : Diagnostics of the residuals from ARIMA (0, 2, 1).

Diagnostics of the residuals from ARIMA (0, 2, 1) is shown in the Figure above.

a) The standardized residuals plot shows no obvious trend and pattern and looks like an independent and identical distribution.

b) The plot of the ACF of the residuals of the diagnostics appears at the middle part and we could see that at any positive lag, there was no evidence of significant correlation in the residuals.

c) Most of the residuals are located on the straight line except some few outliers deviating from the normality.

d) Few of the p values were below 0.05 at any positive lag. This does not provide evidence for significance at any positive lag. In conclusion, the model was adequate and fits well.

4.2.7 TEST ON PARAMETER ESTIMATES

MODEL	TEST ON PARAMETER ESTIMATES				
	PARAMETER	ESTIMATE	STANDARD ERROR	T-VALUE	SIGNIFICANT IF T≥|2|
ARIMA (2,2,1)	AR 1	0.4301	0.2410	1.78	NON-SIGNIFICANT
	AR 2	0.1886	0.2553	0.74	NON-SIGNIFICANT
	MA 1	1.000	0.2352	4.25	SIGNIFICANT
ARIMA (2,2,0)	AR 1	0.9846	0.1967	5.00	SIGNIFICANT
	AR 2	0.5626	0.1908	2.95	SIGNIFICANT
ARIMA (0,2,1)	MA 1	1.000	0.1592	6.28	SIGNIFICANT

DIAGNOSTICS		
	ARIMA(2,2,0)	ARIMA(0,2,1)
RESIDUAL VARIANCE	2166926	2008315
AIC	320.95	319.26
AIC_C	15.96	15.75
BIC	14.89	14.66

4.2.8 SELECTION OF BEST MODEL FOR FORECASTING CASUALTY CASES IN GHANA.

The standardized residuals plots of all the models were independently and identically distributed with mean zero and some few outliers. There was no evidence of significance in the autocorrelation functions of the residuals of all the models and the residuals appear to be normally distributed in all the models. The Ljung – Box statistics were not significant at any positive lag for all the models.

From the table above all the parameters in the coefficients of ARIMA(2,2,0) and ARIMA(0,2,1) models were significant except ARIMA (2,2,1).

The AIC, AIC_C, BIC and residual variance were good for all the models but they favor ARIMA (0, 2, 1) model. Since ARIMA (2,2,1) was not significant, I compared the parameters of ARIMA (2,2,0) and ARIMA (0,2,1)

Comparing the AIC of the models, we could observe that ARIMA (0,2,1) had the minimum..

The best model was chosen based on the minimum residual variance and we could observe that for all the models, ARIMA(0,2,1) has the minimum variance.

From the discussion above it was clear that ARIMA (0, 2, 1) model was the best model for forecasting casualty cases in Ghana.

4.2.9 FITTING THE CASUALTY CASES MODEL

ARIMA (0,2,1) model was the best model for forecasting casualty cases in Ghana. It has two level of differencing without an AR part. In terms of the observed series, the model becomes:

$$y_t - 2y_{t-1} + y_{t-2} = \alpha + \varepsilon_t - \theta_1 \varepsilon_{(t-1)}$$

The point estimate of the parameter of ARIMA (0,2,1) is $\theta_1 = -1.0000$

All the estimates were significant except the constant term, hence it was dropped.

The fitted ARIMA (0,2,1) model for forecasting casualty cases from 1991-2010 was given by

$$\hat{y}_t = 2y_{t-1} - y_{t-2} + \varepsilon_t + \varepsilon_{(t-1)}$$

Where ε_t has an estimated variance of 2008315.

4.3.0 FORECASTING CASUALTY CASES IN GHANA

$pred

Time Series:

Start = 2011

End = 2015

Frequency = 1

[1] 17283.53 17663.06 18042.58 18422.11 18801.64

Time Series:

Start = 2011

End = 2015

Frequency = 1

[1] 1453.965 2106.995 2641.258 3118.407 3561.472

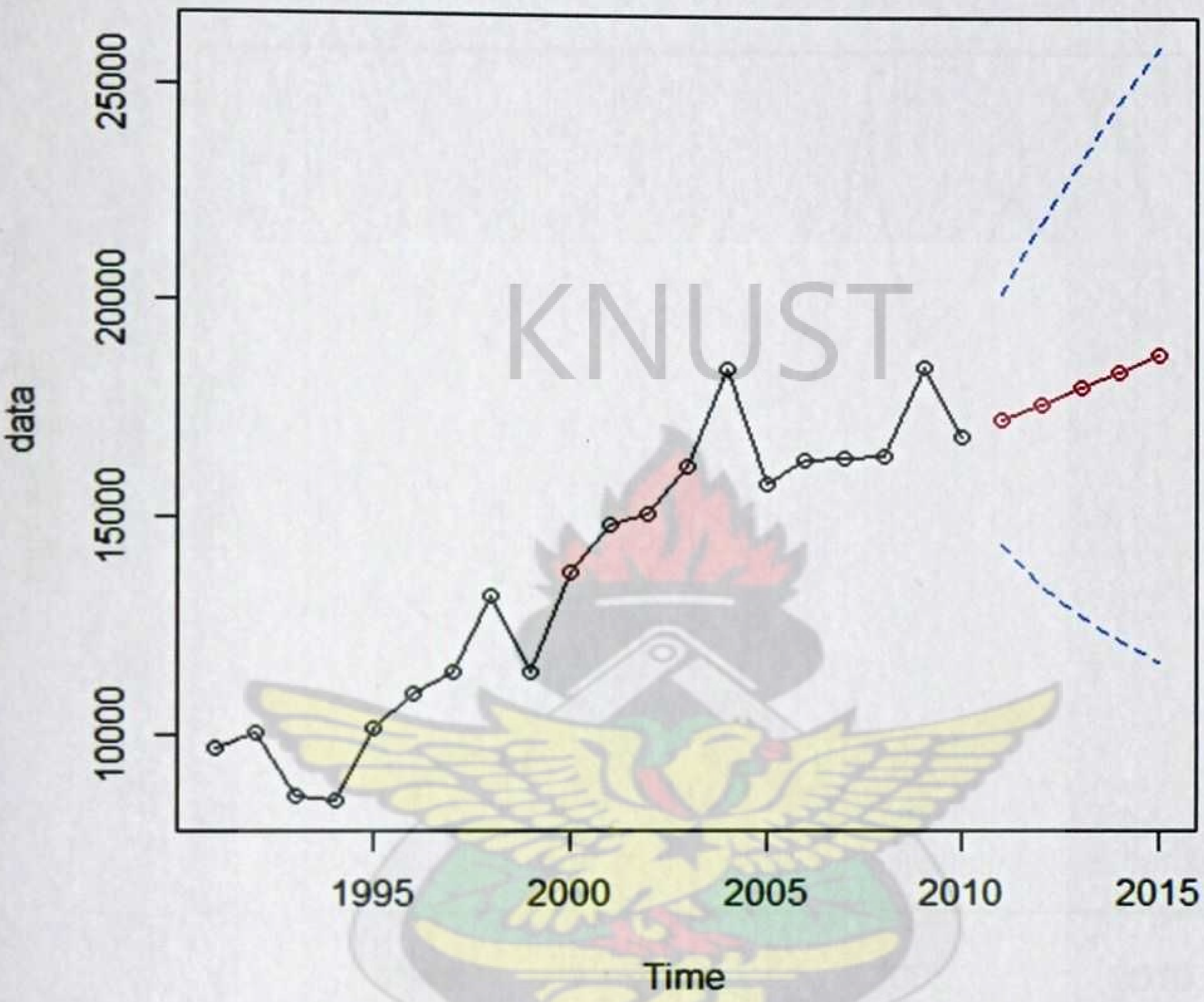

FIG 4.6.5 : Graph of the casualty cases, its forecasts and confidence intervals

The Figure gives the visual representation of the original casualty cases data (black line), its forecasts (red line) and confidence interval (blue short dashes lines).

From the prediction values and the graph above, it can be observed that, casualty cases in Ghana would continue to increase in the next 5 years.

4.3.1 ANALYSIS OF FATALITY CASES IN GHANA

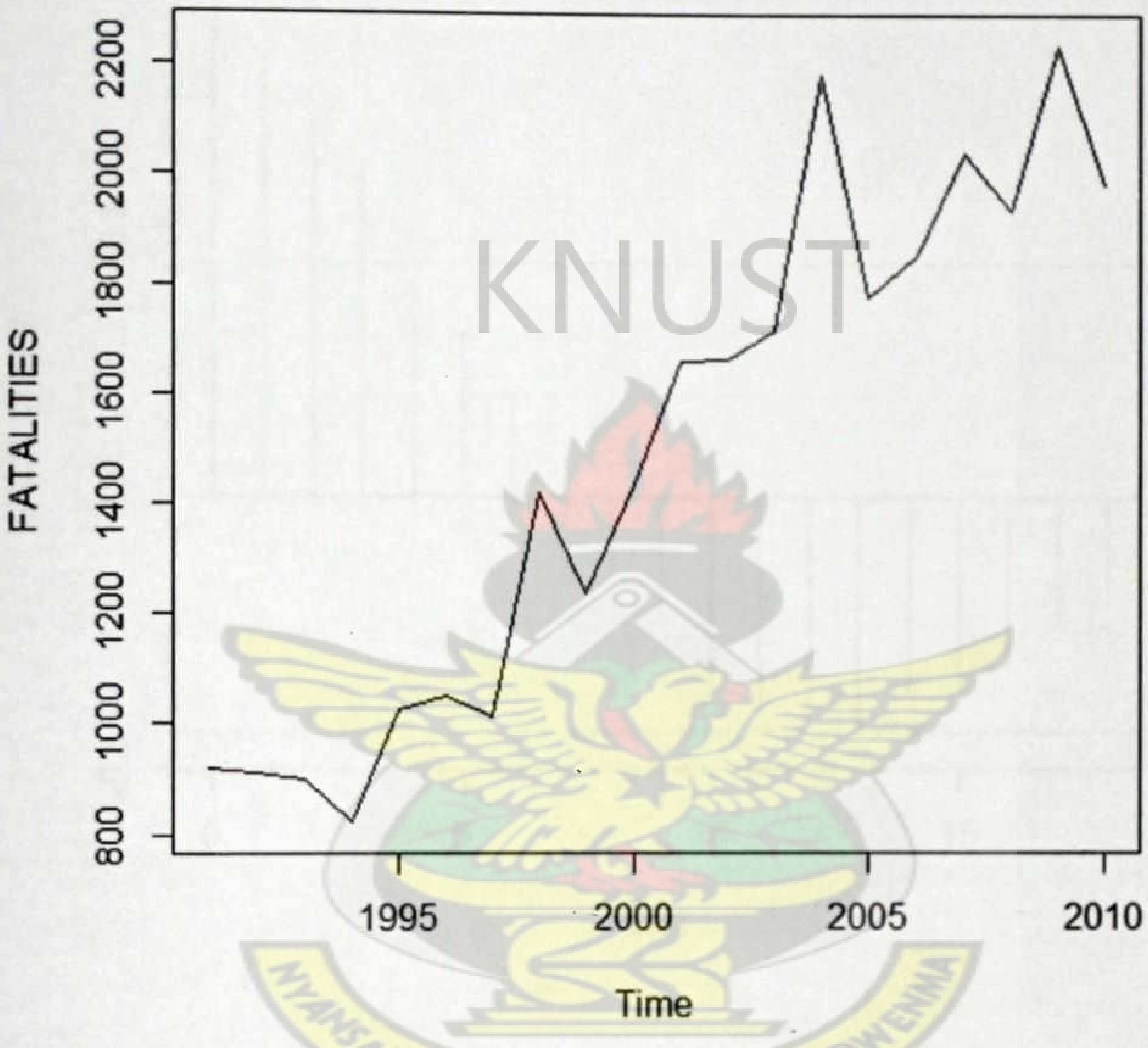

FIG 4.6.6 : Time plot of fatality cases in Ghana from 1991 to 2010.

Fatalities in Ghana decreased from 1991 to 1994 and paltry increased from 1995 to 1996. It then declined from 1997 to 1998 but rose sharply from 1999 to 2004. An irregular pattern was then observed from 2005 to 2010.

In general, the trend of fatality cases in Ghana was on a rise but not always the case. The fatality time plot in fig was not stationary due to the trend component.

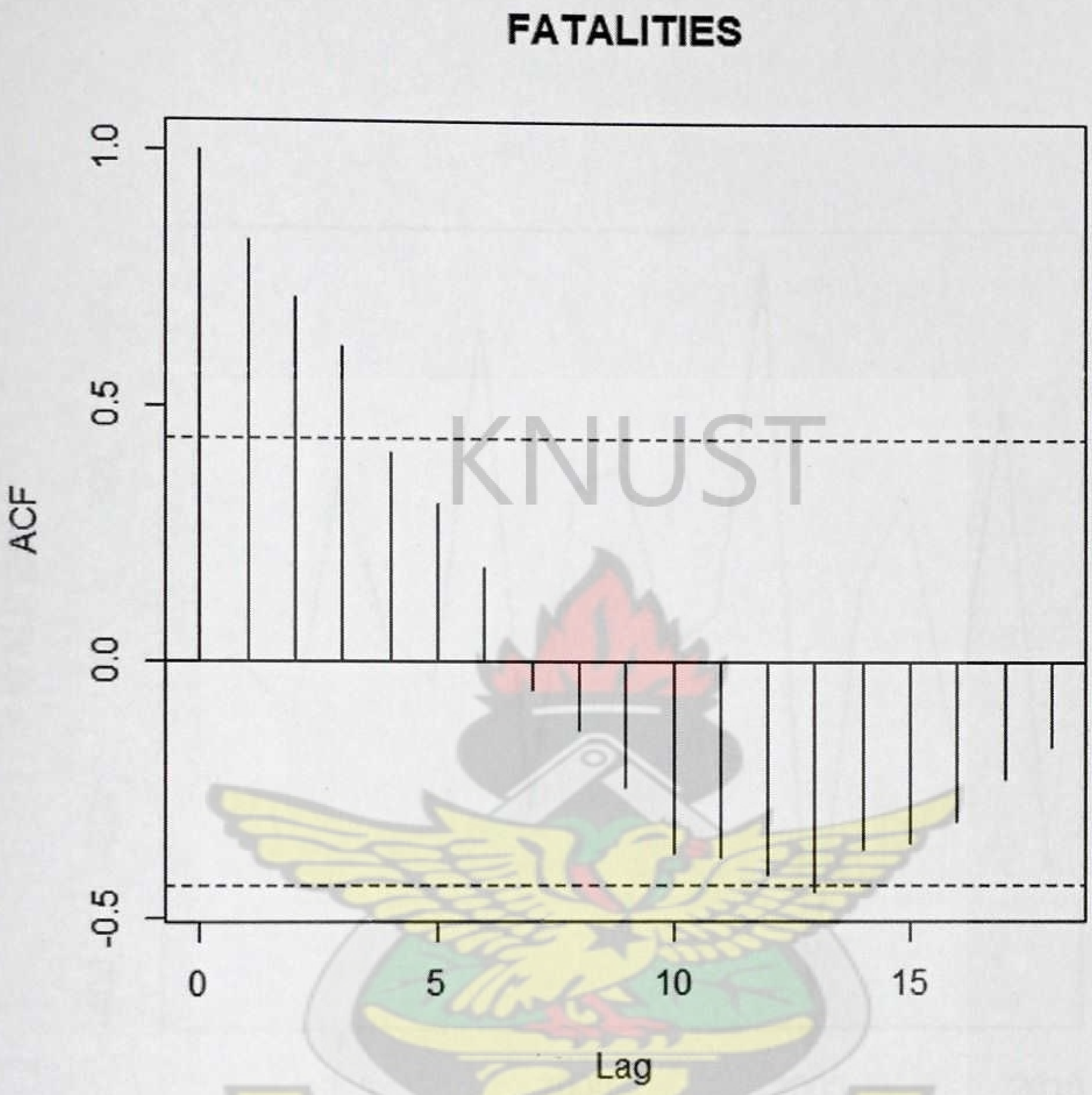

FIG 4.6.7 : Autocorrelation function of fatality cases in Ghana.

4.3.2 FIRST DIFFERENCING OF THE FATALITY DATA CASES

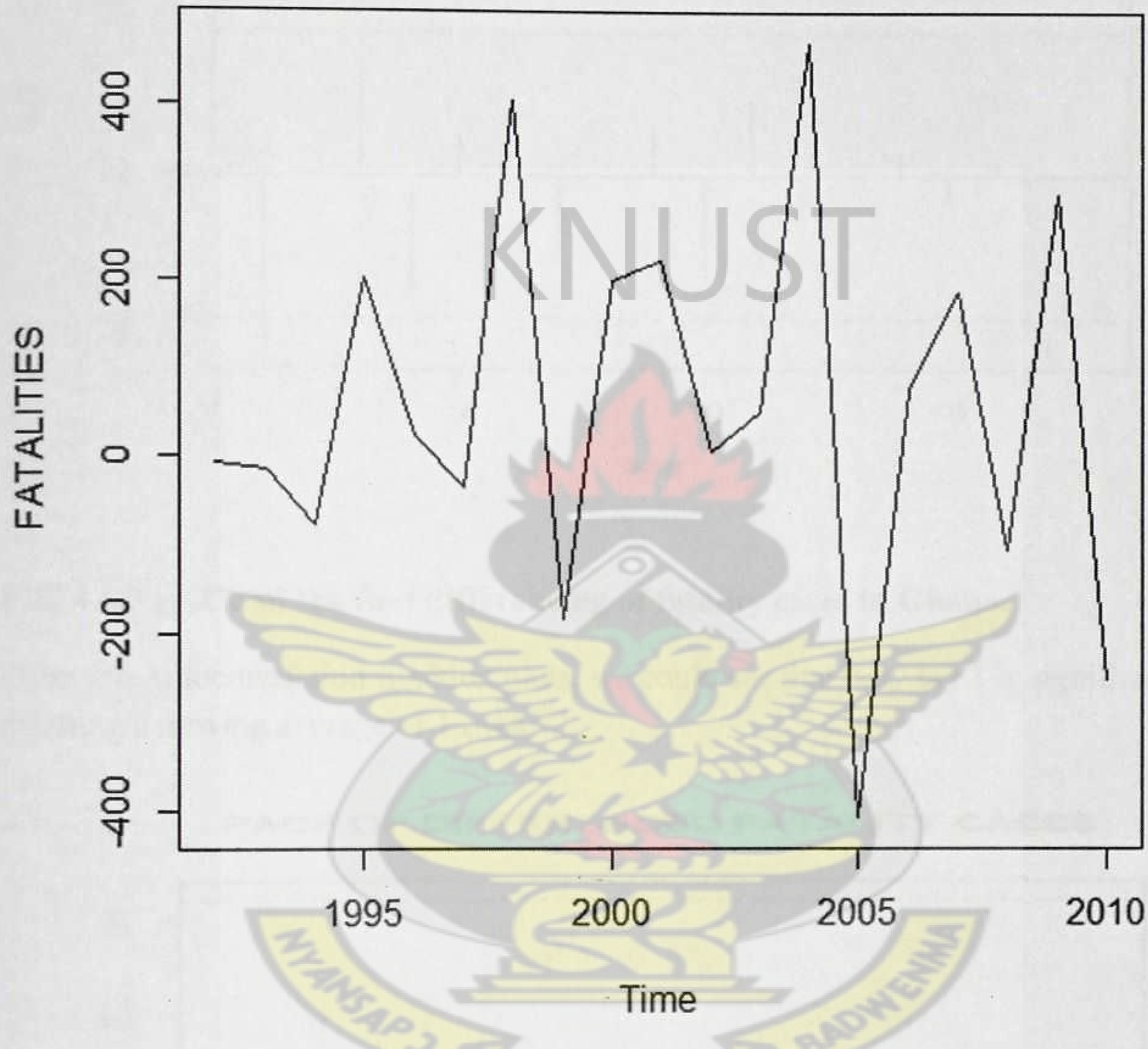

FIG 4.6.8 : First differencing of fatality cases in Ghana from 1991 to 2010.

First differencing method was performed to transform the fatality data by removing the trend component as shown in Figure above. The observations move irregularly but revert to its mean value. The fatality data now looks to be approximately stable.

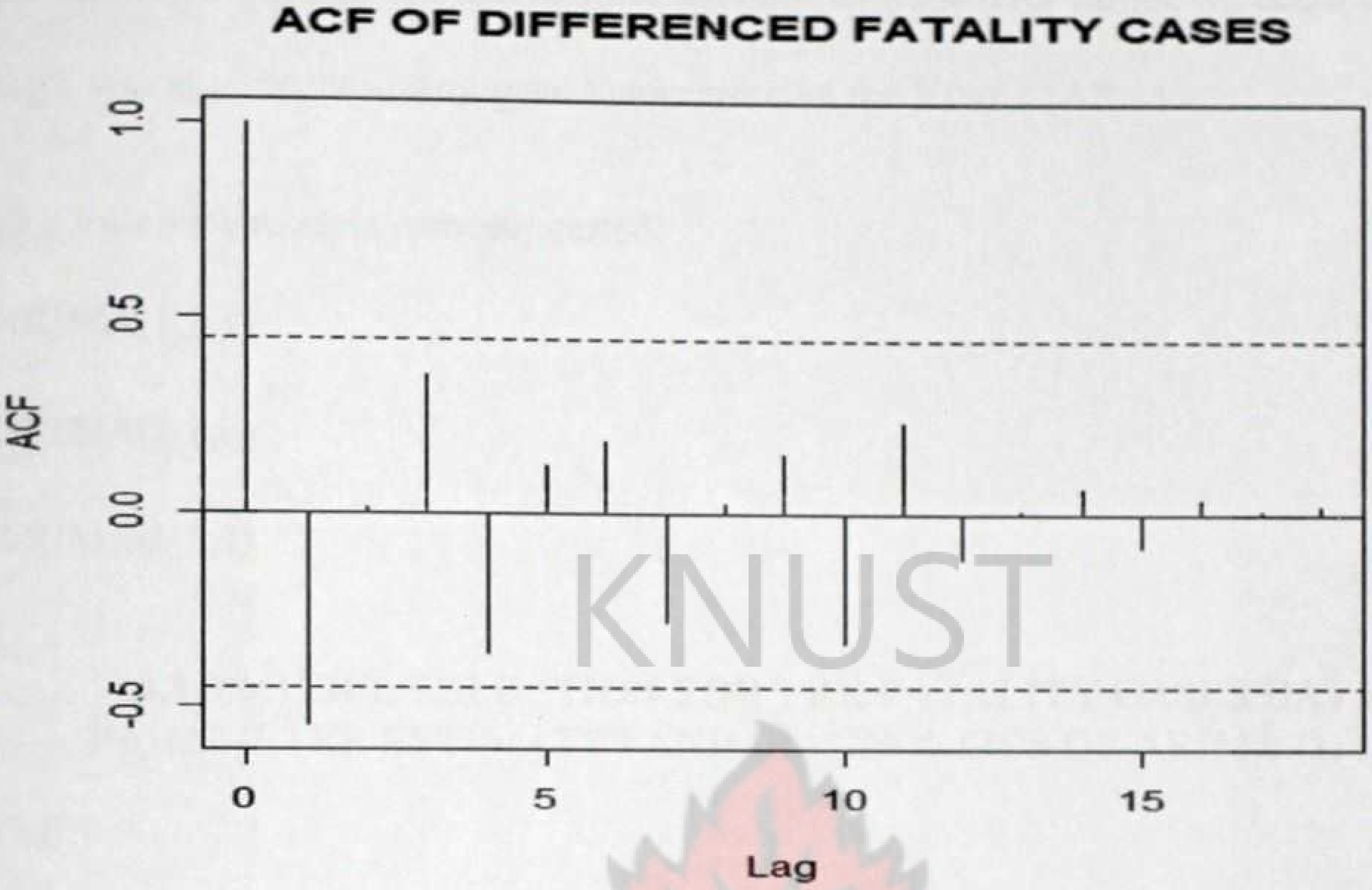

FIG 4.6.9 : ACF of the first differencing of fatality cases in Ghana.

From the Autocorrelation function plots, we could see that only lag 1 is significant meaning a moving average of 1 (MA 1).

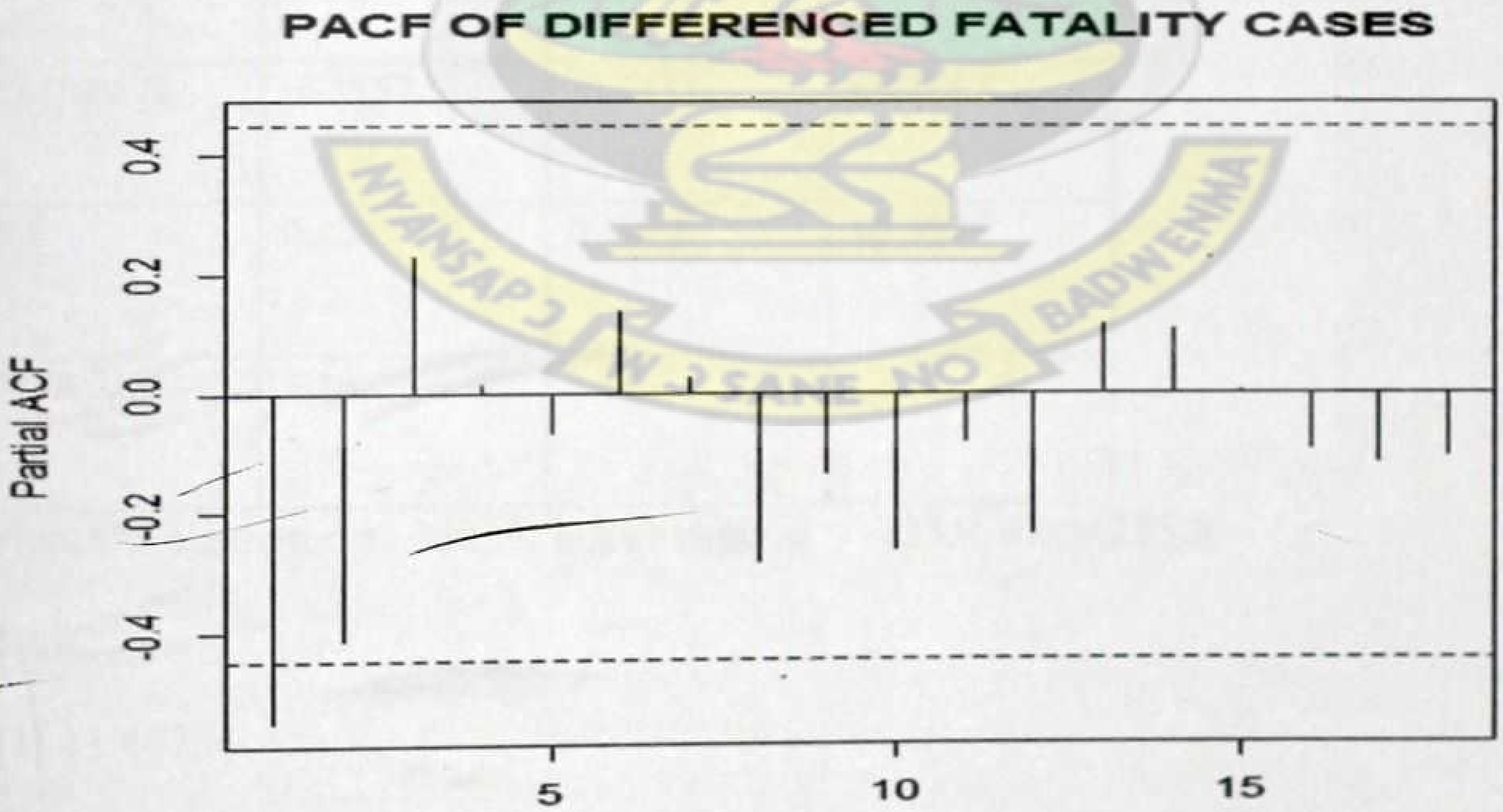

FIG 4.7.0: PACF **of the first differencing of fatality cases in Ghana.**

Comparing the partial Autocorrelation function with the error limits, we could see that lag 1 was significant meaning an Autoregressive model of 1 (AR 1)

The following models were suggested;

ARIMA(1,1,1)

ARIMA(1,1,0)

ARIMA(0,1,1)

4.3.3 MODEL SELECTION FOR THE FATALITY CASES DATA PARAMETER ESTIMATES AND DIAGNOSTICS OF ARIMA (1, 1, 1)

Call:

arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S), xreg = constant, optim.control = list(trace = trc, REPORT = 1, reltol = tol))

Coefficients:

	AR 1	MA 1	CONSTANT
ESTIMATE	-0.3557	-0.4737	68.1867
S.E	0.3063	0.3072	15.9926
T-VALUE	1.16	1.54	

sigma^2 estimated as 26039: log likelihood = -123.9, aic = 255.8

$AIC

[1] 11.46734

$AICc

[1] 11.70068

$BIC

[1] 10.6167

The parameters based on the t-value estimates were not significant since each value was less than 2.

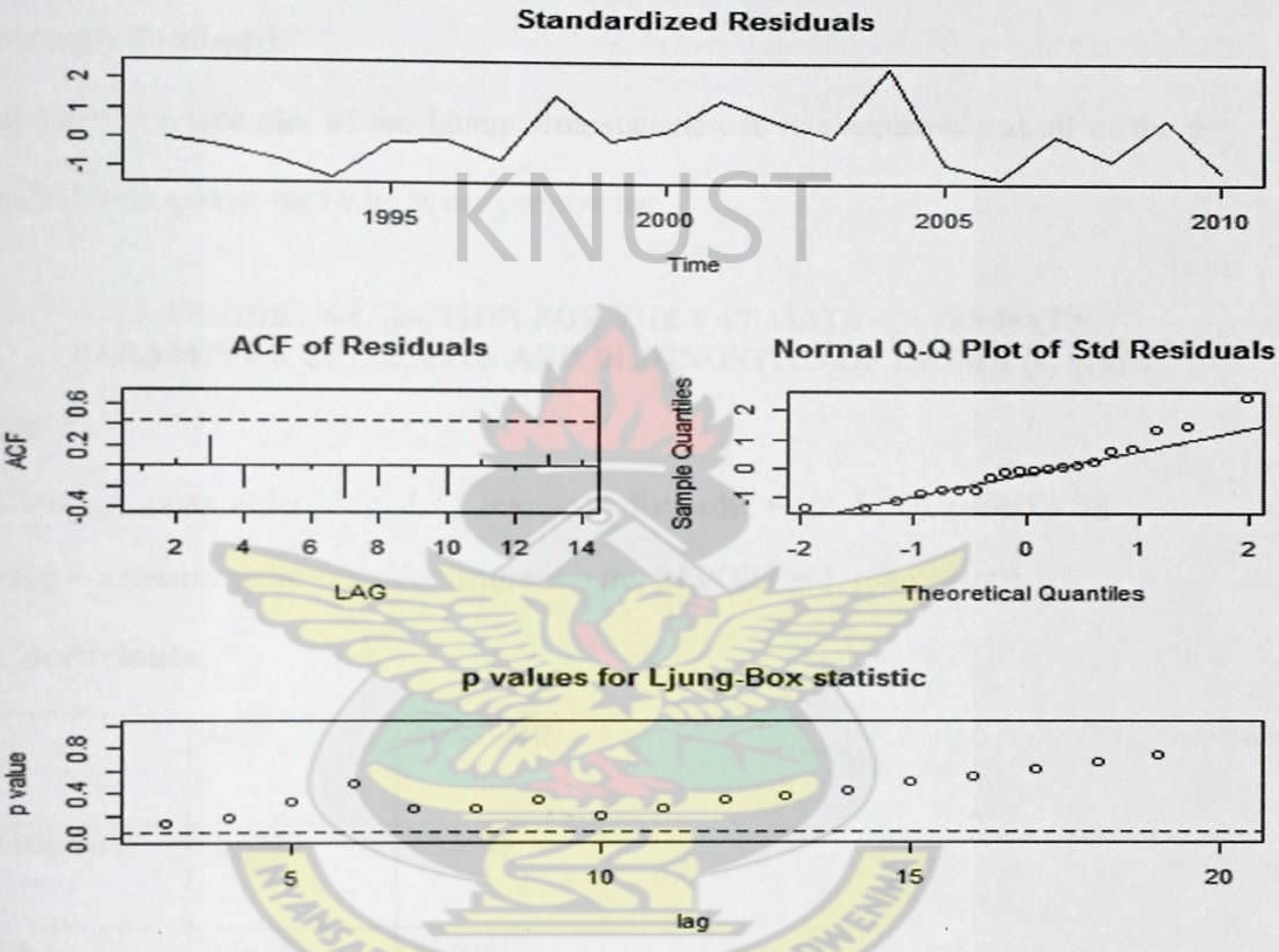

FIG 4.7.1 : Diagnostics of the residuals from ARIMA (1, 1, 1)

Diagnostics of the residuals from ARIMA (1, 1, 1) is shown in the Figure above.

a) The standardized residual plot contains a few outliers with a mean of zero and was independent and identical distribution. It shows no obvious trend and pattern.

b) The plot of the ACF of the residuals of the diagnostics appears at the middle part and we could see that at any positive lag, there was no evidence of significant correlation in the residuals.

c) The normal Q – Q plot of the standardized residuals is shown above. Most of the residuals were located on the straight line except some few outliers deviating from the normality.

Therefore the normality assumption looks to be satisfied and the residuals appear to be normally distributed.

d) From the time plot of the Ljung- Box statistics, It was apparent that all of the p values were greater than 0.05 at any positive lag.

4.3.4 MODEL SELECTION FOR THE FATALITY CASES DATA PARAMETER ESTIMATES AND DIAGNOSTICS OF ARIMA (1, 1, 0)

Call:

arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S), xreg = constant, optim.control = list(trace = trc, REPORT = 1, reltol = tol))

Coefficients:

	AR 1	CONSTANT
ESTIMATE	-0.5886	63.5969
S.E	0.1878	25.6141
T-VALUE	3.13	

sigma^2 estimated as 30119: log likelihood = -125.15, aic = 256.29

$AIC

[1] 11.51292

$AICc

[1] 11.68792

$BIC

[1] 10.6125

The parameters based on the t-value estimate in AR 1 was statistically significant since the t-value was greater than 2.

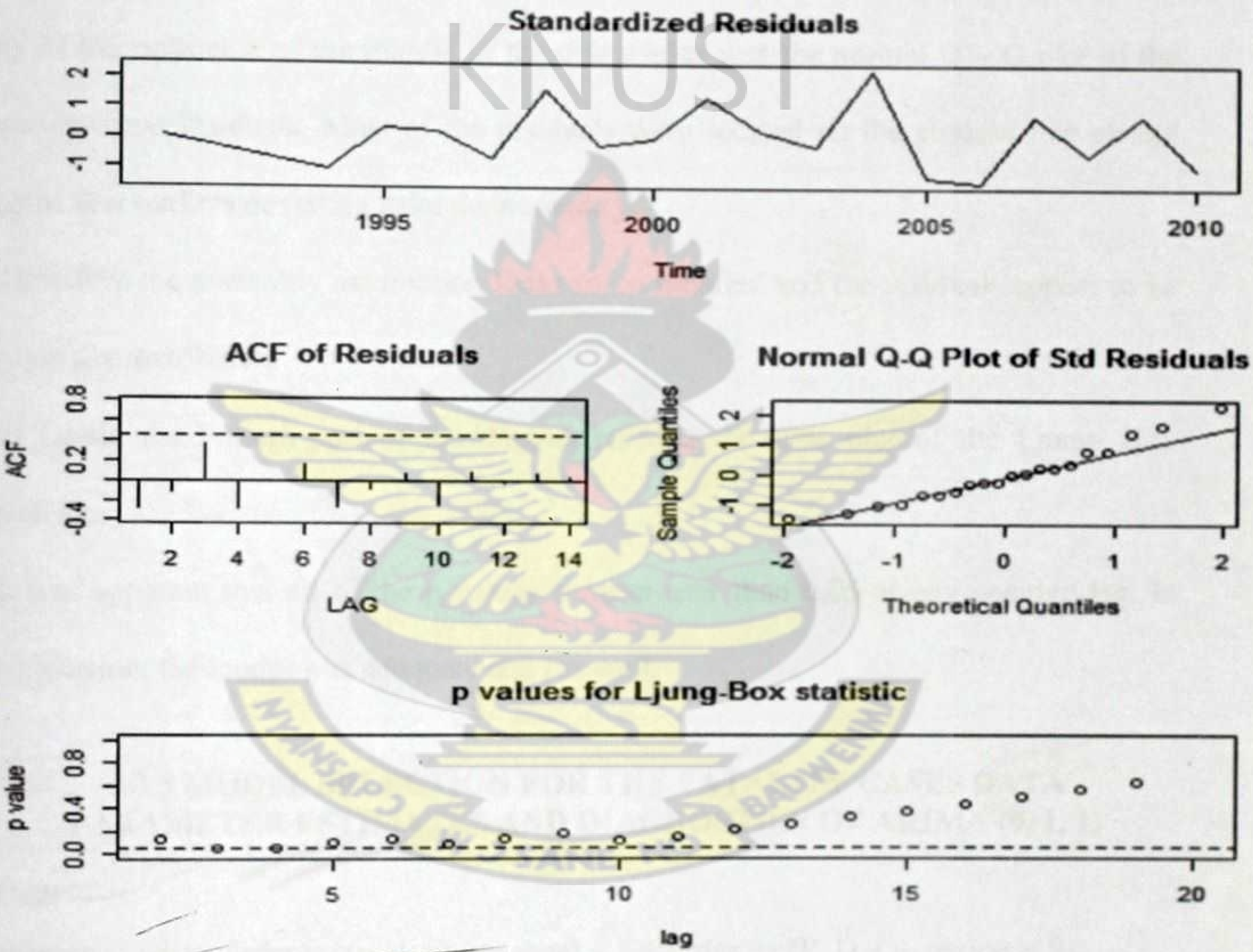

FIG 4.7.2 : Diagnostics of the residuals from ARIMA (1, 1, 0)

Diagnostics of the residuals from ARIMA (1, 1, 0) is shown in the Figure above.

a)The time plot of the standardized residuals of ARIMA (1, 1, 0) appears at the upper part. It contains a few outliers with a mean of zero and was independent and identical distribution. The standardized residuals plot shows no obvious trend and pattern.

b) The plot of the ACF of the residuals of the diagnostics appears at the middle part and we could see that at any positive lag, there was no evidence of significant correlation in the residuals.

c) At the right side of the middle of the diagnostics was the normal Q – Q plot of the standardized residuals. Most of the residuals were located on the straight line except some few outliers deviating from the normality.

Therefore the normality assumption looks to be satisfied and the residuals appear to be normally distributed.

d) Lastly the bottom part of the diagnostics was the time plot of the Ljung- Box statistics.

It was apparent that all of the p values were greater than 0.05 at any positive lag. In conclusion, the model was adequate and fits well.

4.3.5 MODEL SELECTION FOR THE FATALITY CASES DATA PARAMETER ESTIMATES AND DIAGNOSTICS OF ARIMA (0, 1, 1)

Call:

arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S), xreg = constant, optim.control = list(trace = trc, REPORT = 1, reltol = tol))

Coefficients:

	MA 1	CONSTANT
ESTIMATE	-1.000	75.9670

S.E	0.1967	5.8995
T-VALUE	5.08	

sigma^2 estimated as 23145: log likelihood = -123.93, aic = 253.86

$AIC

[1] 11.24951

$AICc

[1] 11.42451

$BIC

[1] 10.34909

The parameters based on the t-value estimate in MA 1 was statistically significant since the t-value was greater than 2.

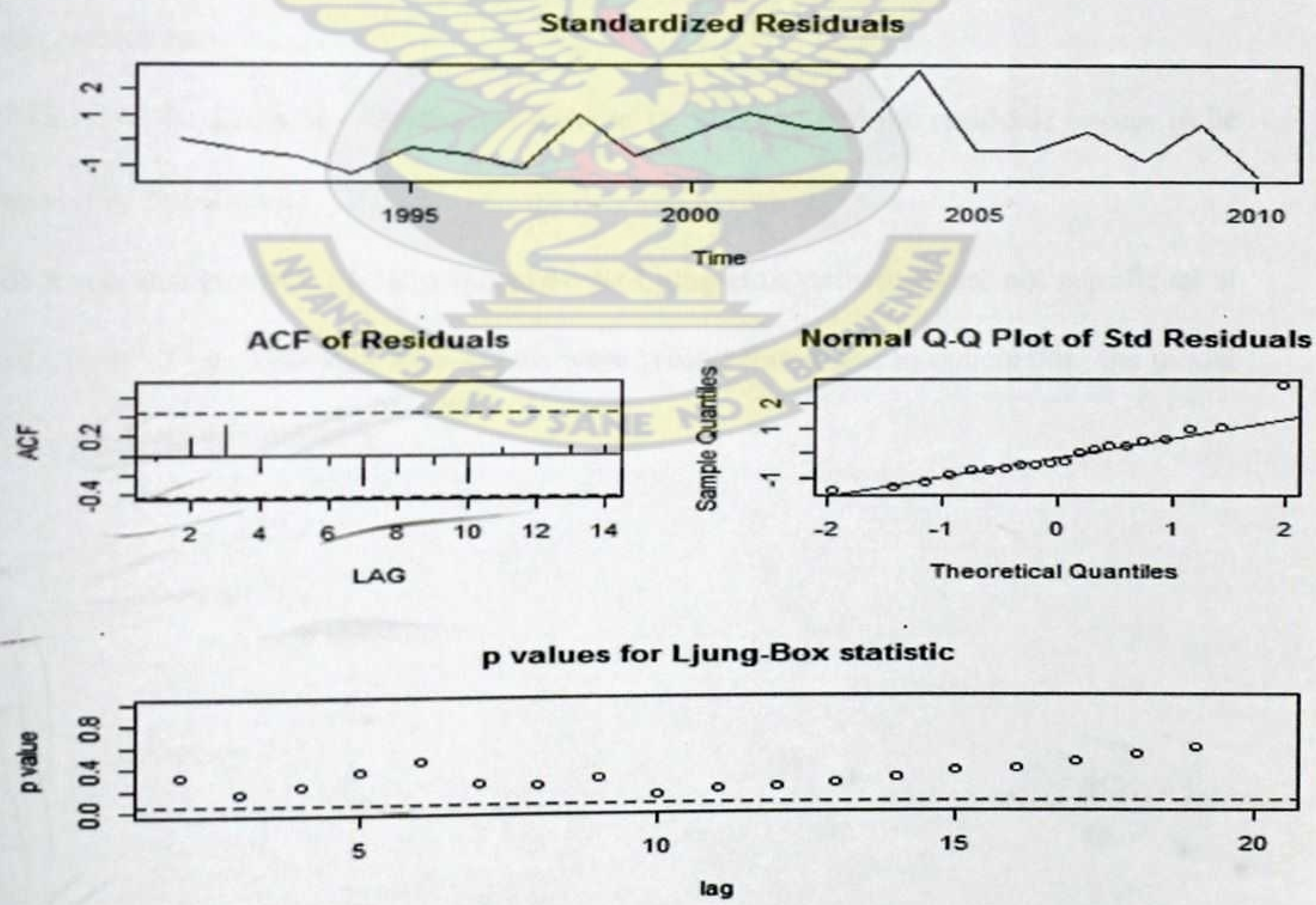

FIG 4.7.3 : Diagnostics of the residuals from ARIMA (0, 1, 1)

Diagnostics of the residuals from ARIMA (0, 1, 1) is shown in the Figure above.

a) The time plot of the standardized residuals of ARIMA (0, 1, 1) appears at the top. The standardized residuals plot shows no obvious trend and pattern.

It contains a few outliers with a mean of zero and is independent and identical distribution.

b) At the middle was the plot of the ACF of the residuals of the diagnostics and we could see that at any positive lag, there was no evidence of significant correlation in the residuals.

c) The normal Q – Q plot of the standardized residuals appears at the right side. Most of the residuals were located on the straight line except some few outliers deviating from the normality.

Therefore the normality assumption looks to be satisfied and the residuals appear to be normally distributed.

d) It was obvious that all the p values of the Ljung-Box statistics were not significant at any positive lag. That was the p values were greater than 0.05. In conclusion, the model was adequate and fits well.

4.3.6 TEST ON PARAMETER ESTIMATES

MODEL	TEST ON PARAMETER ESTIMATES				
	PARAMETER	ESTIMATE	STANDARD ERROR	T-VALUE	SIGNIFICANT IF T≥|2|
ARIMA(1,1,1)	AR 1	0.3557	0.3063	1.16	NON-SIGNIFICANT
	MA1	0.4737	0.3072	1.54	NON-SIGNIFICANT
ARIMA(1,1,0)	AR 1	0.5886	0.1878	3.13	SIGNIFICANT
ARIMA(0,1,1)	MA 1	1.000	0.1967	5.08	SIGNIFICANT

DIAGNOSTICS		
	ARIMA(1,1,0)	ARIMA(0,1,1)
RESIDUAL VARIANCE	30119	23145
AIC	256.29	253.86
AIC_C	11.69	11.42
BIC	10.61	10.34

4.3.7 SELECTION OF THE BEST MODEL

The standardized residuals plots of all the models were independently and identically distributed with mean zero and some few outliers. There was no evidence of significance in the autocorrelation functions of the residuals of all the models and the residuals appear to be normally distributed in all the models. The Ljung – Box statistics were not significant at any positive lag for all the models.

From the above all the parameters in the coefficients of ARIMA(0,1,1) and ARIMA(1,1,0) models were significant except ARIMA (1,1,1).

The AIC, AIC_C and BIC were good for all the models but they favor ARIMA (0, 1, 1) model since it has the smallest Aic and residual variance.

From the discussion above it was clear that ARIMA (0, 1, 1) model was the best model for forecasting fatality cases in Ghana.

4.3.8 FITTING THE FATALITY CASES MODEL

ARIMA (0,1,1) model wais the best model for forecasting fatality cases in Ghana. It has one level of differencing without an AR part. In terms of the observed series, the model becomes:

$$y_t - y_{t-1} = \alpha + \varepsilon_t - \theta_1 \varepsilon_{(t-1)}$$

The point estimate of the parameter of ARIMA (0,1,1) is $\theta_1 = -1.0000$ and

$\alpha = 75.9670$

All the estimates were significant.

The fitted ARIMA (0,1,1) model for forecasting fatality cases from 1991-2010 was given by

$$\hat{y}_t = y_{t-1} + \varepsilon_t + 75.9670 + \varepsilon_{(t-1)}$$

Where ε_t has an estimated variance of 23145.

4.3.9 FORECASTING FATALITY CASES IN GHANA

Time Series:

Start = 2011

End = 2015

Frequency = 1

[1] 2288.053 2364.020 2439.987 2515.954 2591.921

$se

Time Series:

Start = 2011

End = 2015

Frequency = 1

[1] 155.8902 155.8902 155.8902 155.8902 155.8902

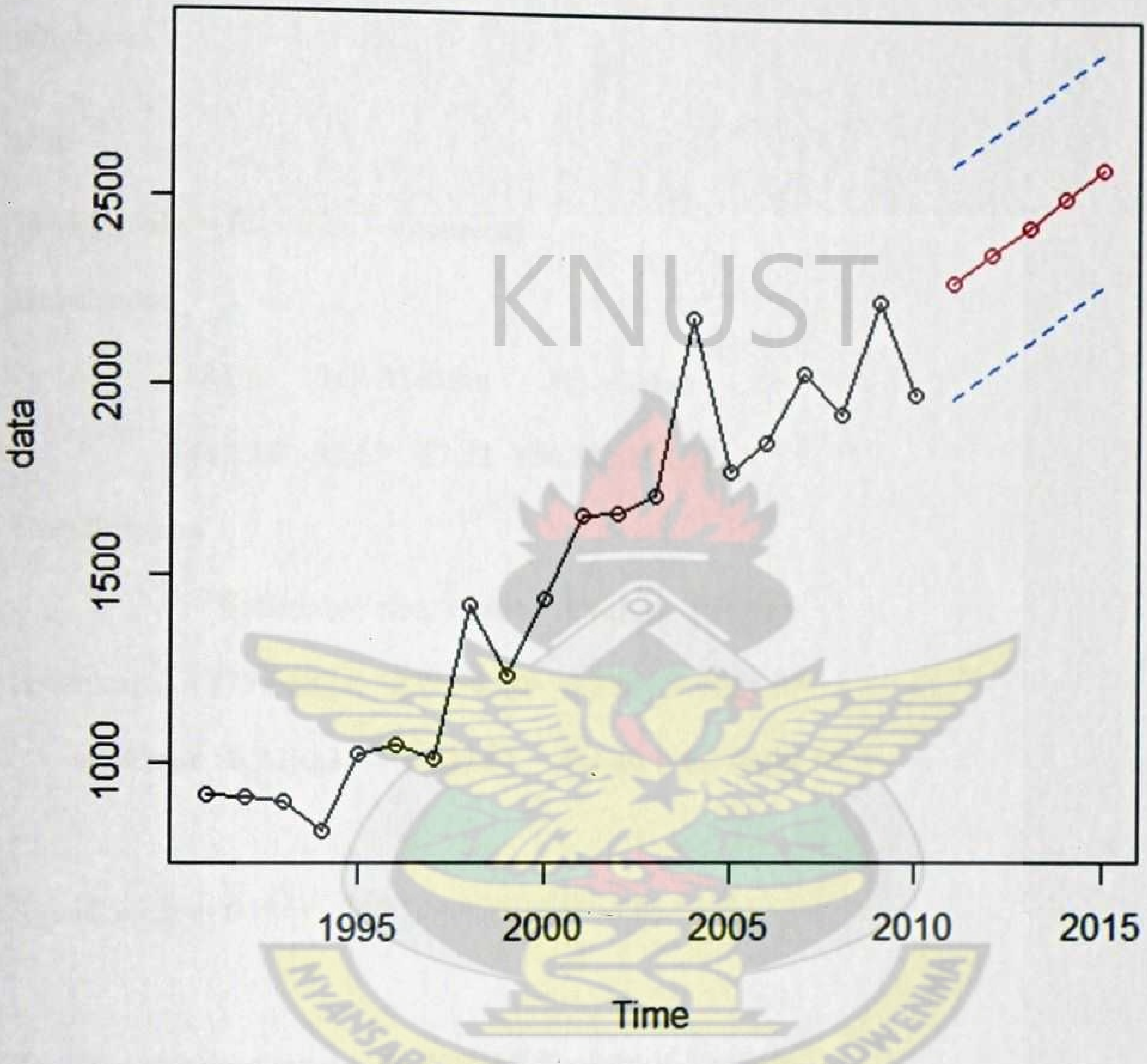

FIG 4.7.4 : Graph of the casualty cases, its forecasts and confidence intervals

The Figure gives the visual representation of the original fatality cases data (black line), its forecasts (red line) and confidence interval (blue short dashes lines).

From the prediction values and the graph above, it can be observed that, fatality cases in Ghana would continue to increase in the next 5 years.

To implement the Engle Granger method, we begin by regressing the accident cases, casualties and fatalities on each other and then access the model of fit. Taking fatalities as dependent variable and accident as independent variable, the following output was obtained:

Call:

lm(formula = fatalities ~ accident)

Residuals:

Min	1Q	Median	3Q	Max
-449.18	-97.57	27.31	136.34	234.17

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-779.55824	239.54298	-3.254	0.0044	**
accident	0.22623	0.02347	9.640	1.57e-08	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 196.4 on 18 degrees of freedom

Multiple R-squared: 0.8377, Adjusted R-squared: 0.8287

F-statistic: 92.92 on 1 and 18 DF, p-value: 1.567e-08

The following equation is obtained:

Fatalities = -779.55824+0.22632 accident

The p-values of the independent variables were very small, this means that these regression coefficients were statistically significant at the 0.05 level of significance. The R squared value was 0.8377 meaning 83.77% of the variations in the fatality cases were explained by changes in the accident cases. To determine the long run relationship between fatalities and accidents, we test whether the residuals of the regression relationship were stationary. The residuals are obtained as follows:

1	2	3	4	5	6
-193.977721	127.601737	217.535890	114.067108	-75.082673	-91.672732
7	8	9	10	11	12
-449.180070	-289.054832	34.114319	-291.641663	-115.244817	20.515491
13	14	15	16	17	18
110.653092	211.221286	-2.352998	-4.080658	99.214646	180.627266
19	20				
234.168902	162.568425				

These residuals were subjected to stationarity by testing them using the dickey fuller test or kpss test.

Augmented Dickey Fuller Test

Data:

Dickey-Fuller=-1.9568, Lag order=2, p-value=0.5859

Alternative hypothesis: stationary

Since the p-value was greater than the significance level, we fail to reject the null hypothesis that the residuals were not stationary since there was the presence of a unit root, and hence no long run relationship between fatalities and accident.

Taking casualties as dependent variable and accident as independent variable, the following output was obtained :

Call:

lm(formula = Casualties ~ Accident)

Residuals:

Min	1Q	Median	3Q	Max
-2000.61	**-757.42**	**98.03**	**907.36**	**1733.14**

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	**-2497.3359**	**1416.0292**	**-1.764**	**0.0948 .**
Accident	**1.6078**	**0.1387**	**11.589**	**8.83e-10** ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1161 on 18 degrees of freedom

Multiple R-squared: 0.8818, Adjusted R-squared: 0.8753

F-statistic: 134.3 on 1 and 18 DF, p-value: 8.826e-10

The following equation is obtained:

Casualties= -2497.3359+1.6078 accident

The p-values of the independent variables were very small, this means that these regression coefficients were statistically significant at the 0.05 level of significance. The R squared value was 0.8818 meaning 88.18% of the variations in the casualty cases were explained by changes in the accident cases. To determine the long run relationship between casualties and accidents, we test whether the residuals of the regression relationship were stationary. The residuals were obtained as follows:

1	2	3	4	5	6
-1266.76782	1398.29504	677.83143	399.72408	-736.12446	-197.48565
7	8	9	10	11	12
-2000.60850	-1976.79343	-152.62466	-1581.10124	-821.30356	346.99225
13	14	15	16	17	18
1733.13789	1377.63605	110.28643	85.77961	-441.09833	922.71092
19	20				
1219.27155	902.23968				

These residuals were subjected to stationarity by testing them using the dickey fuller test or kpss test.

Augmented Dickey Fuller Test

Data:

Dickey-Fuller= -2.3127, Lag order=2, p-value=0.4533

Alternative hypothesis: stationary

Since the p-value was greater than the significance level, we fail to reject the null hypothesis that the residuals were not stationary since there was the presence of a unit root, and hence no long run relationship between casualties and accident.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1.0 CONCLUSIONS

Road traffic accident in Ghana was increasing at an alarming rate and has raised major concerns. The NRSC recognizes the contributions road safety researches make to development of accident reduction initiatives. It was against this background that this thesis was carried out in order to identify the patterns of Road Traffic Accident cases, casualties and fatalities and to develop a time series ARIMA models to predict 5 years Road Traffic Accident cases, casualties and fatalities in Ghana.

Time series analysis of the data from the years 1991 – 2010 showed that patterns of Road Traffic Accident cases, casualties and fatalities were increasing in Ghana.

ARIMA models were subsequently developed for the accident data cases, casualties and fatalities over the period 1991 – 2010, after identifying various tentative models.

ARIMA (0, 2, 1) was identified to be suitable model for forecasting in to the future of the accident and casualty cases whist ARIMA(0,1,1) was found to be suitable model for forecasting fatality cases in Ghana.

The study also revealed that road traffic accident cases, casualties and fatalities in Ghana would continue to increase over the next 5 years.

5.2.0 RECOMMENDATION

- An improved and better policies of National road safety commission should be introduced with much emphasis on publicisation and education to ensure maximum reduction in Road accident cases in Ghana.
- More formal statistical methods should be used to analyse road accident cases in Ghana rather than the use of just descriptive statistics. This will enhance further projections and analysis.
- The model should not be used to forecast long time ahead. This is because long time forecasting could lead to arbitrary large forecast values.
- Laws on overspeeding should be strictly enforced. The use of seat belt should be checked. The compulsory seat belt law for all drivers and front seat passengers should be highly enforced.
- There should be an enforcement of traffic safety campaign. Poor enforcement of traffic safety regulations can be due to lack of well trained staff, inadequate physical resources, administrative problems and corruption. Enforcement must be meaningful and be maintained over a long period of time to increase fear of punishment amongst drivers.
- It is very important that roads in Ghana should be properly maintained in terms of the use of appropriate materials for patching pot holes.
- Finally, it is also recommended that further research should be conducted to take care of drastic government interventions.

REFRENCES

1. Abdel-Aty, M., 2003, "Analysis of driver injury severity levels at multiple locations using ordered probit models", Journal of Safety Research, 34, 597.
2. Abdeh, Hammid Abbassi, 2005, Stats Methods of predicting Fatalities of road traffic accident in Kuwait. " Egyptian population and family review. Pp.2-19 Accident analysis and prevention, 18(2): pp.119-127.
3. Abhishek, Misra, 2000, "Analysis of run-off the road crashes involving overcorrection". Electronic theses, treatises and Dissertations. Paper 2391.
4. Agran, P., D. Winn, C. Anderson, and C. Del Valle, 1998, "Family, Social, and Cultural Factors in Pedestrian Injuries among Hispanic Children." Injury Prevention, Vol. 4, No. 3: 188–193.
5. Al-Ghamdi, A.S., 1996b. Road accidents in Saudi Arabia: analysis of urban intersections in Riyadh (accident analysis and prevention 35 (2003) 717-724)
6. Alshuler, A. 1965, A town Planning Process: A Political Analysis. Cornell University, U.S.A.
7. Asogwa, S.E. 1992, Road traffic accidents in Nigeria a review and reappraisal. Accident Analysis and Prevention ; 24 (2):149-55.
8. Astrom, J. S, Kent, M.P. and Jovin, R. D. 2006, Signatures of Four Generations of Road Safety Planning in Nairobi City, Kenya In: Journal of Eastern African Research and Development. Vo. l20, pp. 186-201.
9. Aworemi, Joshua Remi, A. 2010, " Causal factors of road traffic in southwestern nigeria". Educational Research Vol. 1(4) pp. 118-124 May 2010. http//www.interesjournals.org/ER

10. Ballestros, M.F, P.C Dischinger and P. Langenberg. "pedestrian injuries and vehicle type in maryland, 1995-1999. "Accident analysis and prevention.vol 36,No.1(2004):73-81 Barret R, 1989, "National urban and town road safety comparison". Adis Ababa, 16-20

11 Bedard, M., Guyatt, G. H., Stones, M. J. and Hirdes, J. P., 2002. "The independent contribution of driver, crash, and vehicle characteristics to driver fatalities". Accident Analysis and Prevention Vol 34 No 6, pp.717-727.

12 Carson, J., and Mannering, F., 2001, "The effect of ice warning signs on iceaccident frequencies and severities", Accident Analysis and Prevention, 33, 99

13 Cejun, Lin chen, 2004, Time series analysis and forecast of Annual crash fatalities. " National Centre for statistics and Analysis"

14. Chang, L.-Y., and Mannering, F., 1999, "Analysis of injury severity and vehicle occupancy in truck- and non-truck-involved accidents", Accident Analysis and Prevention, vol 31, number 5 pp 579-592

15. Clifton, K., C. Burnier, and G. Akar, 2009, "Severity of Injury Resulting from Pedestrian- Vehicle Crashes: What Can We Learn from Examining the Built Environment?" Transportation Research: Part D: Transport and Environment, Vol. 14, No. 6 : 425–436.

16. Creighton, R.L (1979). Urban Transportation Planning, University of Illinois, Chicago. Page 375. ISBN: 0-252-00069-2.

17. Dissanayake, S., Lu, J., 2002. Analysis of severity of young driver crashes using sequential binary logistic regression modeling. Journal of Transportation Research Board 1784, 108–114.

18. Donnell, E. T., Harwood, D. W., Bauer, K. M., Mason, J. M. Jr., and Pietrucha, M. T. 2002, "Cross-Median Collisions on Pennsylvania Interstates and Expressways". Transportation Research Record 1784, Transportation Research Board, Washington D.C., pp91 – 99.

19. Down, J. 1997, Ideology and Town Planning in Tanzania. In: Journal of the Geographical Association of Tanzania. Vol. 15,pp. 20-39.

20. Duncan, C. S., Khattak, A. J., and Council, F. M., 1998, "Applying the ordered probit model to injury severity in truck-passenger car rear-end collisions". Transportation Research Record, 1635, paper number 98-1237

21. Herbert, D.S 1979, Urban. Development in Third world, Policy Guidelines, Praeger. 9 038 987654321. Publisher, New York.

22. Hill, R.W., 2003. Master Thesis: Statistical Analysis of Fatal Traffic Accident Data, Texas Tech University, Lubbock, Texas.

23. Holubowycz, O.T, 1995, "Age, Sex, and Blood Alcohol Concentration of Killed and Injured Pedestrians." Accident Analysis and Prevention, Vol. 27, No. 3 (1995): 417–422.

24. Jehle, D. and E. Cottington, 1998, "Effect of Alcohol Consumption on Outcome of Pedestrian Victims." Annals of Emergency Medicine, Vol. 17, No. 9 (1988): 953–956.

25. Jovanis, P. and Chang, H, 1979, "Modeling the relationship of accident to miles travelled". Transport Research Record 1086, page 42-51.

26. Khattak, A. j. 2001, "Injury severity in multivehicle rear-end crashes", Transportation Research Record, 1746, 59

27. Khorashadi, A., Niemeier, D., Shankar, V., and Mannering, F., 2005, "Differences in rural and urban driver-injury severities in accidents. Accident Analysis and Prevention, 37, 910

28. Kim, K., Kim, S. and Yamashita, E. 2000, "Alcohol-Impaired Motorcycle Crashes in Hawaii, 1986 to 1995". Transportation Research Record 1704, Transportation Research Board, Washington D.C., page 77 – 85.

29. Kockelman, K. M., and Kweon, Y.-J., 2002, "Driver injury severity: an application of ordered probit models", Accident Analysis and Prevention, 34, 313-321

30. Kweon, Y.-J., and Kockelman, K. M., 2003, "Overall injury risk to different drivers: combining exposure, frequency, and severity models", Accident Analysis and Prevention, 35, 441

31. Leon, S.R. (1996).Reducing death on the Road .The effects of minimum safety standards, public heath, 86(1):31-3.

32. MacLeod, K.E., et al. "Factors Associated with Hit-and-Run Pedestrian Fatalities and Driver Identification." TRB Annual Meeting 2010, Paper #10-3715, 2010.

33. Melcher, D. J., 2003. "Driver Oversteer and Directional Control Loss: Physical Fundamentals and Transportation Safety Applications".

34. Mercier, C. R., Shelley, M. C., Rimkus, J., and Mercier, J. M, 1997, Age and gender as predictors of injury severity in Head-On Highway Vehicular Collisions. The 76th Annual meeting, Transport Research Board, Washington, DC, Transportation research record 1581.

35. Milton, J. Mannering, F, 1997, The Relationship Among Highway Geometric, Traffic-Related Elements and Motor-Vehicle Accident Frequencies. The 76th Annual Meeting, Transportation Research Board, Washington, D.C.

36. Moudon, A.V. 2011, "The Risk of Pedestrian Injury and Fatality in Collisions with Motor Vehicles, A Social Ecological Study of State Routes and City Streets in King County, Washington." Accident Analysis and Prevention, Vol. 43, No. 1: 11–24.

37. Nassar, S.A, Saccomanno, F.F., and Shortreed, J.H, 1997, Integrated Risk Model (ARM) of Ontario. The 76th Annual Meeting, Transport Research Board, Washinton, D.C.

38. Nataliya, V, Malyshkina, 2006, "Influence of posted speed limit on roadway safety" Accident Analysis and Prevention, 33, 569

39. National Road Safety Commission (NRSC) Ghana, newsletter 2005.

40. National Road Safety Commission (NRSC) Ghana, newsletter 2009.

41. Odero, W., Garner, P. and Zwi, A. 1997, Road traffic injuries in the developing countries: a comprehensive review of epidemiological studies. Trop.med.int.health.2:445-460.

42. Ouyang, Y, Shankar, V., and Yamamoto, T. 2002, "Modeling the Simultaneity in Injury Causation in Multi-vehicle Collisions". Transportation Research Record 1704, Transportation Research Board, Washington D.C., pp143 – 152.

43. Peden, M. (Ed), 2004, World Report on Road Traffic Injury Prevention. World Health Organisation, Geneva.

44. (Peden) World Health Organization (WHO) report 2004

45. Pierce, Q and Maunder, P, 1998, Road and Traffic Authority Statistical report "Road Traffic Accidents in NSW-1997".[Motor Accidents Authorities of NSW].

46. Piet, R and Shefer, D, 1997, "Congestion and Safety on Highways: Towards an Analytical Model, in Ronan, P and Leyer (eds.). An International Journal for Research in Urban and Regional Studies, University of Glassow, Scotland, Vol 34 (4).

47. Pless, IB., R. Verreault, and S. Tenina, 1989, "A Case-control Study of Pedestrian and Bicyclist Injuries in Childhood." American Journal of Public Health, Vol. 79, No. 8: 995–998.

48. Plurad, David, Ginger, 2006, "Pedestrian Injuries: The Association of Alcohol Consumption with the Type and Severity of Injuries and Outcomes." Journal of the American College of Surgery, Vol. 202, No. 6 (2006): 919–927.

49. Roynolds, D,J. 1966), Economics, Town Planning and Traffic, Worcester, Ltd, London

50. Savolainen, P. T., 2006, "An evaluation of motorcycle safety in Indiana", Ph.D. Dissertation, Purdue University, UMI Dissertation Publishing

51. Savolainen, P. T., and Mannering, F., 2006a, "Additional evidence on the effectiveness of motorcycle training and motorcyclists' risk-taking behavior", submitted to Transportation Research Record.

52. Shankar, V., Mannering, F., and Barfield, W., 1996, "Statistical analysis of accident severity on rural freeways", Accident Analysis and Prevention, 28, 391.

53. Siddiqui, N.A., X. Chu, and M. Guttenplan, 2006, "Crossing Locations, Light Conditions, and Pedestrian Injury Severity." Transportation Research Record, Vol. 1982 (2006): 141–149. 23.

54. Soderlund, N. and Zwi, A. B. 1995, Traffic-related mortality in industrialized and less developed countries. Bulletin of the World Health Organisation 73 (2), 175-182.

55. Stares, S and Zhi Liu, 1997, "Chaina's Urban Transport Development Strategy, in Stares, S and ZhiLiu (eds.).Proceedings of a symposium in Beijing, World Bank.

56. Sunanda Dissanayake, P.E. contributory causes related to large trucks 2010. Report number 25-1121-0001-111. MATC-KSU:111

57. Taylor, W. C., Lim, I., and Lighthizer, D. R. 2001, "Effect on Crashes After Construction of Directional Median Crossovers". Transportation Research Record 1758, Transportation Research Board, Washington D.C., pp30 – 35.

58. Turner, P. A. and Georggi, N. 2001, " Analysis of Alcohol-Related Motorcycle Crashes in Florida and Recommended Countermeasures". Transportation Research Record 1779, Transportation Research Board, Washington D.C., pp189 – 196.

59. Ulfarsson, G. F., 2001, "Injury severity analysis for car, pickup, sport utility vehicle and minivan drivers: male and female differences", Ph.D. Dissertation. University of Washington, UMI Dissertation Publishing 61

60. Ulfarsson, G. F., and Mannering, F. L., 2004, "Differences in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger car accidents", Accident Analysis and Prevention, 36, 135

61. Unpublicised crash test, seat belts and alcohol. Am J Public Health; 86(1):31-3.

62. World Health Organization, 2000, Global Pollution and Health Related Environmental Monitoring, London, U.K.

63. World Health Organization (WHO), global report on road safety, 2004.

64. World Health Organization (WHO), global report on road safety, 2009.

65. Wough, D, 1990, Geography. An Integrated Approach, Thomas Nelson Ltd, Hong Kong.

66. Xiao, J., Kulakowski, B. T., and El-Gindy, M. 2000, "Prediction of Risk of Wet-Pavement Crashes, Fuzzy Logic Model". Transportation Research Record 1717, Transportation Research Board, Washington D.C., pp28 – 36.

67. Yamamoto, T., and Shankar, V. N., 2004, "Bivariate ordered-response probit model of driver's and passanger's injury severities in collisions with fixed objects", Accident Analysis and Prevention, 36, 869

68. Yuan, F., Ivan, J. N., Qin, X., Garrick, N. W., and Davis, C. F. 2001, "Safety Benefits of Intersection Approach Realignment on Rural Two-Lane Highways". Transportation Research Record 1785, Transportation Research Board, Washington D.C., pp21 – 29.

69. Zajac, S.S. and J.N. Ivan, 2003, "Factors Influencing Injury Severity of Motor Vehicle-Crossing Pedestrian Crashes in Rural Connecticut." Accident Analysis and Prevention, Vol. 35, No. 3 (2003): 369–379.

70. Zhang ,J., Fraser ,S., Lindsay ,J. ,Clarke, K, Mao, Y.1998, Age specific pattern of factors related to fatal motor vehicle traffic crashes: Focus on young and elderly, drivers.Public Health; 112(5): 289-95.

APPENDICE

ACCIDENT, CASUALTY AND FATALITY DATA IN GHANA FROM 1991-2010

YEAR	ACCIDENT	CASUALTIES	FATALITIES
1991	8370	9693	920
1992	6922	10030	914
1993	6467	8578	901
1994	6584	8488	824
1995	8313	10132	1026
1996	8488	10952	1049
1997	9918	11448	1015
1998	10996	13205	1419
1999	8763	11439	1237
2000	11087	13747	1437
2001	11293	14838	1660
2002	10715	15077	1665
2003	10542	16185	1716
2004	12175	18445	2186
2005	11320	15813	1779
2006	11668	16348	1856
2007	12038	16416	2043
2008	11214	16455	1938
2009	12299	18496	2237
2010	11506	16904	1986

SOURCE: BRRI/NRSC, GHANA