KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY, KUMASI, GHANA

COLLEGE OF SCIENCES

INSTITUTE OF DISTANCE LEARNING

DEPARTMENT OF MATHEMATICS

TIME SERIES ANALYSIS OF ROAD TRAFFIC ACCIDENTS IN GHANA, A CASE STUDY OF ACCRA – TEMA MOTORWAY, GREATER ACCRA REGION

THESIS SUBMITTED TO THE INSTITUTE OF DISTANCE LEARNING, KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN INDUSTRIAL MATHEMATICS

BY

OKUTU, JOHN KWADEY

(PG3016309)

B.ED MATHEMATICS (HONS.)

SUPERVISED BY

MR. EMMANUEL HARRIS

JULY, 2011.

DECLARATION

I wish to state that this work has never been submitted by anyone to any university or institution for the award of a degree or other purpose. I declare that this is my own original work. References from the work of others have been duly acknowledged.

Student	Supervisor
Signature………………….	Signature………………………
Name: Okutu, John Kwadey	Name: Mr. Emmanuel Harris
Date……………………….	Date…………………………

Head of Department, Mathematics	The Dean, Institute of Distance Learning
Signature…………………..	Signature……………………….
Name: F. K. Darkwah	Name: Professor I. K. Dontwi
Date……………………….	Date……………………………

DEDICATION

This work is dedicated to my son Philip Akafo Okutu and in memory of my late wife Dora Dadzie, may her soul continues to rest in absolute peace.

ACKNOWLEDGEMENTS

A good number of people have contributed in various ways to enable me go through the programme and carry out my project work successfully. I express my gratitude to them. I would like to thank all my lecturers especially Dr. F. T. Oduro, Dr. Osei Frimpong, Dr. S.K. Amponsah and the late Mr. E.Agyemang for being helpful to me in diverse ways during the period of the programme.

I am particularly grateful to my academic supervisor, Mr. Emmanuel Harris for his supervision, time and support during the course of the thesis in spite of the fact that he always has a busy schedule.

I also acknowledge the support and co-operation I enjoyed from D.S.P. Francis Obeng of the National Motor Traffic and Transport Unit (MTTU) and the MTTU officers of the Ashaiman and Airport (Accra) Divisional Command Units.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Road traffic accident in Ghana is increasing at an alarming rate and has raised major concerns. In this thesis, time series with Box – Jenkins method was applied to 31 years of annual road accident data from 1980 – 2010 to determine patterns of road traffic accident cases, injuries and deaths along the Accra – Tema motorway. Models were subsequently developed for accident cases, injuries and deaths. ARIMA (1,1,1) was tentatively used to model the injury and death data whilst ARIMA(0,1,2) was used to model the accident cases data. 10 years forecasts were made using the models developed and it showed that, road traffic accident cases, injuries and death would continue to increase.

CHAPTER 1

INTRODUCTION

1.1.0 BACKGROUND OF THE STUDY

A Road Traffic Accident (RTA) is when a road vehicle collides with another vehicle, pedestrian, animal or geographical or architectural obstacle. The Road Traffic Accidents can result in injury, property damage and death. Road Traffic Accident results in the deaths of about two million people worldwide and injuries about five times this number every year and it was estimated that approximately 3000 people die by road traffic accidents around the world on a given day (WHO, 2008). A projection of global leading causes of death from 2008 to 2030 by World Health Organization revealed that, if current trends continue, road traffic accidents will rise from the ninth to fifth of world leading cause of death 3.6% of global deaths, up from 2.2% in 2004(WHO, 2008).While disability-adjusted life years (DALYs) will rise from ninth with 2.7% of total DALYs in 2004 to third and 4.9% of total DALYs in 2030 (WHO, 2004b).

In this study, a road traffic accident is defined as accident which took place on the road between two or more objects, one of which must be any kind of a moving vehicle (Jha et al, 2004). Road Traffic Accidents are increasing with a rapid pace and presently these are one of the leading causes of death in developing countries.

The morbidity and mortality burden in developing countries is rising due to a combination of factors, including rapid motorization, poor road and traffic infrastructure as well as the behavior of road users (Nantulya and Reich, 2002). This contrasts with technologically advanced countries where the indices are reducing (Osram eta al, 1994; Oneil and Mohan, 2002).

According to AUSTROADS (1994), road accidents occur as a result of one, or more than one of the following factors; Human factors; Vehicular factors; Road and environmental factors. Driving faster or slower than the flow of traffic, which may or may not accord with the posted speed limit has robustly been demonstrated to increase the likelihood and severity of crashes, as shown by the Solomon Curve (OOIDA, 2003).The factors of traffic accidents are driver, the highway and motor vehicles (Aaron and Strasser, 1990; Balogun and Abereoje, 1992; Mock et al, 1999).Most traffic accidents often involve the three elements. Most road traffic accidents involve motor vehicles but bicycles or pedestrians accidents can occur without vehicles (Stutts and Hunter, 1999). A high proportion of road traffic accidents can be apportioned to unsafe human acts. The drunken drivers of motor vehicles make the clearest example (Hijar et al. 2000). Reckless and dangerous driving, alcoholism, faulty pedestrian attitude, etc. constitute the major causes in Ghana (Mensah, 1986, Oduro, 1998).

According to the Motor Traffic and Transport Unit (MTTU) of the Ghana Police, from January 1, 1992 to December 31, 2001(a period of ten years), a total of 104,420 accidents cases were recorded giving an annual average of 10,442 cases. And about 145,331 vehicles were involved and 10,106 people lost their lives with 80,022 people injured. However, between January and March 2003 a period of three months the capital, Accra alone recorded 1,417 motor accidents with 200 deaths, 373 serious injuries and 966 minor cases. And according to the National Road Safety Commission from January 1, 2006 to September in the same year, 2,185 people lost their lives through accidents representing an increase of about 6% over the figure for the whole of 2005 when 1,778 lives were claimed in road accidents. From 2007 to

2010, a total of 6,213 lives had been lost while 39,797 others had sustained various degrees of injury in motor accidents.

According to records compiled by the unit 1,760 people died while 11,147 people injured in 12,981 motor accidents involving 18,589 vehicles in 2010.

In Greater Accra, a slightly heavily motorized region in Ghana with some poor road conditions and transport systems has a high rate of RTAs and the tendency is on the increase. The recognition of road traffic accidents as a crisis in Ghana inspired the establishment of the National Road Safety Commission. The commission was charged with responsibilities for among others, policy making, organization and administration of road safety in Ghana.

Despite increased road safety campaign by the commission, the rate at which accidents occur on our road is very alarming. It is truism that one of the major challenges this country is still battling with is motor accidents. Professor Agyeman Badu Akosor, the former Director General of Ghana Health Service rightly hammered home this fact when he stated that the most deadly disease in Ghana at the moment is motor accident.

For developing measures aimed at reducing the rate of road traffic accidents and the consequent injuries and fatalities, there is the need for regular evaluation of the road traffic accidents in terms of developing statistical models for forecasting future number of accident cases, fatalities rates and injuries and this is the purpose of the thesis.

The Accra-Tema motorway was opened to traffic in 1964 by the then government of Ghana as part of the country's programme of transforming Tema into an industrial hub of the newly independent nation. The 19km motorway is the oldest paved road in

Ghana. Being a concrete pavement, it is more expensive to construct than asphalt or other bituminous surface roads but it is more economical to operate over a longer time; it is longer lasting stronger and requires minimal maintenance.

Among the features of the motorway was a dual carriageway with a median or a central reservation area that completely separated the two carriageways. Like all motorways, the Accra-Tema motorway was designed prohibiting pedestrian movement, parking areas or U-turn. Moreover, until recently, no road joined the motorway at any other section except the entry and the exit points.

The economic importance of the motorway cannot be over emphasized as it is the main route for transportation of goods to and from the Tema Harbour and also passengers and goods to the Volta and Northern regions from Accra, and countries east of Ghana.

Considering the importance of the road and the increased level of road traffic accidents in recent years along the road, there is the need for this study aimed at characterizing the RTAs to provide an enabling base for the development of countermeasures by the government and the traffic control agents to reduce incidences of road traffic accident on the road.

1.2.0 STATEMENT OF THE PROBLEM

Over the years, the education and research department of the national MTTU uses descriptive statistics techniques and charts such as bar graphs, histograms and frequency polygons to organize accidents data on the Accra – Tema motorway. This statistical approach of analyzing the data does not inform the department about the estimates of RTA cases, injuries and deaths in the future. Hence the national MTTU and RSC cannot make projections about RTA along the Accra – Tema motorway.

In light of problems associated with this method of analyzing RTA data by the MTTU, it is necessary to use time series techniques which can better describe and model the accident data.

1.3.0 OBJECTIVES OF THE STUDY

The specific objectives of this thesis are as follows:

- To identify patterns of road traffic accident cases, injuries and death along the Accra –Tema motorway over the period of 1980-2010.
- Develop a suitable time series forecasting model for number of road traffic accident cases, injuries and deaths on the Accra-Tema motorway over the period 1980 - 2010 and use it to estimate 10 years forecast.

1.4.0 JUSTIFICATION FOR THE STUDY

This research has become necessary to conduct following the increasing rate of road fatalities and injuries in Ghana especially along the Accra-Tema motorway. The results of this thesis would be useful for road safety planning on the Accra –Tema motorway.

Another usefulness of the work of this thesis was to provide a better opportunity for the national MTTU 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.

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

1.5.0 METHODOLOGY

Data for the study was secondary; a historical annual traffic crash data for the years 1980 through 2010 was compiled from the Airport and Ashaiman divisional commands of the Motor Traffic and Transport Unit (MTTU) of the Ghana Police Service. These two stations are responsible for compiling accident data on the Accra – Tema motorway. 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 consists of three steps iterative cycle of:

- Model identification
- Model estimation
- Diagnostics checks on model adequacy followed by forecasting.

1.6.0 SCOPE AND LIMITATION

The scope of the thesis was limited to the following:

- Thirty one (31) years of annual number of accident cases, injuries and accident deaths along the Accra- Tema motorway for the period 1980-2010.
- Using Box-Jenkins methodology to develop a time series model for both descriptive and forecasting purposes. In this case the explanatory capacity of the model was not addressed, as no additional independent variable was used for modeling the time series data.

1.7.0 ORGANIZATION OF THE THESIS

The remaining portion of the thesis was organized as follows:

Chapter 2 presents literature review, followed by Chapter 3 on methodology. Chapter 4 focused on data collection and analysis. Conclusions and recommendations were presented in Chapter 5.

CHAPTER 2

LITERATURE REVIEW

This chapter focused on review of earlier forecasting models in road traffic accidents and related statistical (including time series) models used in analyzing road traffic accident data. The following are some references reviewed which are related to the research topic that has been studied.

2.1.0 EARLIER FORECASTING MODELS OF ROAD TRAFFIC ACCIDENTS

Many researchers including Smeed (1949) have devoted their research to the area of road accidents and reported pioneering work on the analysis of road accidents. Smeed examined the relationship on a number of road fatalities with those of motor vehicles and the population of twenty countries in 1938 in the following form:

$$D/N=0.0003(N/P)^{0.67} \qquad (1)$$

where D, N, P are deaths, number of motor vehicles and population respectively. Using the same method as Smeed, Jacobs and Cutting (1986) carried out analysis of fatalities in developing countries for different years and established significant relationships between fatality rates and levels of vehicle ownership. The analysis was repeated for the years 1980 using data from 20 developing countries and a relationship was derived which is as follows

$$D/N=0.00036(N/P)^{0.65} \qquad (2)$$

Smeed's analysis was heavily criticized by Andreessen (1985) for model accuracy.

He argued that the Smeed's formula cannot be applied universally to all countries. The fatality model similar to Smeed's equation produced by Andreessen in 1985 is of the form:

$$\text{Death}=0.000112(\text{Population})^{0.73259}(\text{Number of Vehicles})^{0.33293} \quad (3)$$

Mekky (1985) used the same time series data for the analysis and studied the effects of a rapid increase in motorization levels on fatality rates in some developing countries. Kim (1990) developed a similar model in Korea and suggested the following equation:

$$\text{Death}=0.25451(\text{Population})^{0.699196}(\text{Number of Vehicles})^{0.251414} \quad (4)$$

In Malaysia two models had been proposed. Aminuddin (1990) proposed a simple linear model and projected 4950 deaths by 2000. Rehan (1995) however improved Aminuddin's model and suggested a similar model to Smeed's and derived the following equation:

$$\text{Death}=0.08193(\text{Population x Number of Vehicles})^{0.335} \quad (5)$$

Using employment and population data, Partyka (1984) developed simple models with a view to understand the various factors affecting the increase in accidents in developing countries. The study on the effects of speed limits on road accidents has been carried out by Fieldwick and Brown (1987). It was found that speed limits have considerable effects on safety in urban and rural areas. Minter (1987) discussed an application of the two models (Wright and Towell) for road safety problems and finally developed a model for estimating the road accidents in U.K.

Pramada and Sarkar (1993) investigated the variations in the pattern of road accidents in various Union Territories of India. Emenalo et al (1987) established the trend curves for the road accidents casualties, and other relevant quantities for Zambia. Pramada and Sarkar (1997) again developed a road accidents model by using the additional parameter of road length. Ameen and Naji (2001) presented a general modeling strategy to forecast road accident fatalities in Yemen.

2.2.0 TIME SERIES AND OTHER STATISTICAL MODELS USED IN RTA

Numerous cross-sectional studies have been conducted in varying scales and scopes in order to understand the relationships between factors and traffic accidents by combining several years of data and performing statistical analysis and constructing statistical models. The multiple regression and Poisson regression are commonly used for modeling the mortality rates and number of deaths in a specific population. However Pocock et -al. (1981) pointed out that unweighted multiple regression is not appropriate for modeling mortality rates in different areas which vary in population size. In addition fully weighted regression is usually too extreme. Thus they introduced an intermediate solution via maximum likelihood for modeling death rates. Tsauo et al. (1996) examined the effect of age, period of death and birth cohort in motor vehicle mortality in Taiwan from 1974 – 1992, used data from vital statistic. Log-linear regression was used for fitting the model to perform the effects of variables. Kardara and Kondakis (1997) identified trends of road traffic accident deaths and injuries rates in Greece from 1981-1991 by using linear regression with logarithmic transformation. LaScala et al. (2000) examined correlations between demographic and environmental versus pedestrian injury rates by using a spatial autocorrelation corrected regression model with applying the logarithmic transformation for the injuries rates. Evans (2003) conducted statistical modeling for estimating road traffics and railways accident fatal rates based on past accident data in Great Britain during 1967-2000. In addition Lix et al. (2004) used Poisson regression to investigate the relation of demographic, geographical, and temporal explanatory variables with mortality in difference regions of Manitoba, Canada between 1985and 1999, used data from Vital Statistics records and the provincial health registry.

Yang et al. (2005) used Poisson regression modeling to examine and compare age- and sex-specific mortality rates due to injuries in the Guangxi Province in South Western China in 2002, based on death certificates data. However this study focused only on small areas.

In light of problems associated with ordinary (regression) methods because of the assumption that the observations overtime 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) studied road accidents in Kuwait. He used an ARIMA model and compared it with ANN to predict fatalities in Kuwait. He concluded ANN was better in case of long term series without seasonal fluctuations of accidents or autocorrelations' components. Wen et al (2005) established a procedure of Road Traffic Injury (RTI) in China by using RTI data from 1951 to 2003. A series of predictive equations on RTI were established based on ARIMA models. They concluded that time series models thus established proves to be of significant usefulness in RTI prediction. Cejun and Chiou-Lin (2004) used two time series techniques; ARMA and Holt-Winters (HW) algorithm to predict annual motor vehicle crash fatalities. They concluded that the values predicted by ARMA models are a little bit higher than the ones obtained by HW algorithm. 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 3

METHODOLOGY

In this chapter we have looked at some theoretical background of time series and the method of time series which we have used in modeling road traffic accident data along the Accra –Tema motorway during the period 1980-2010.

3.1.0 TIME SERIES

A time series is a sequence of observations (Y_t) ordered in time. That is Y_1,Y_2 ,Y_3,…,Y_N or {Y_t}, $t \in N$,where 1,2,3,… denote time steps. Mostly these observations 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. 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.

Time series models have advantages in certain situations. They can be used easily for forecasting purposes because historical sequences of observations upon study variables are readily available from published secondary sources. These successive observations are statistically dependent and time series modeling is concerned with techniques for the analysis of such dependencies. Thus in time series modeling, the prediction of values for the future periods is based on the pattern of past values of the variable under study, but not generally on explanatory variables which may affect the system.

3.2.0 TIME SERIES COMPONENTS

An important step in analyzing time series data is to consider the types of data patterns, so that the models most appropriate to those patterns can be utilized. Four types of time series components can be distinguished. They are:

(i) Horizontal – when data values fluctuate around constant value

(ii) Trend – when there is long term increase or decrease in the data

(iii) Seasonal – when a series is influenced by seasonal factors and recurs on a regular periodic basis.

(iv) Cyclic – when the data exhibit rises and falls that are not of a fixed period.

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.

3.3.0 STATIONARITY OF TIME SERIES

A time series is stationary if: $\mathrm{E}(\mathrm{Y}_t) = \mu$, $\mathrm{Var}(\mathrm{Y}_t) = \sigma^2$ for all t. In other words, if $y_1, y_1, y_1, \dots, y_n$ values of the time series fluctuate around a constant mean with constant variation, the time series is stationary. If the n values do not seem to fluctuate around a constant mean, then it is non- stationary.

If the time series is not stationary, we can transform it to stationarity with one of the following techniques.

- We can difference the data. That is , given the series Z_t , we create the new series

  $$Y_{(t)} = Z_{(t)} - Z_{(t-1)} \qquad (6)$$

  $t = 2,3,4, \ldots$

  The differenced data will contain one less point than the original data. Although you can difference the data more than once.

- 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.
- 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.

The above techniques are intended to generate series with constant location and scale.

3.4.0 AUTOCORELATION FUNCTION

Autocorrelation refers to the correlation of a time series with its own past and future values. Autocorrelation is also sometimes called "lagged correlation" or "serial correlation", which refers to the correlation between members of a series of numbers arranged in time. 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

An important guide to the persistence in a time series is given by the series of quantities called the sample autocorrelation coefficients, which measure the correlation between observations at different times. The set of autocorrelation coefficients arranged as a function of separation in time is the sample autocorrelation function, or the acf.

The first- order autocorrelation coefficient is the simple coefficient of the first N – 1 observations, x_t t=1, 2,…, N -1 and the next N - 1 observations, x_t, t=2,3,…, N .

The correlation between x_t and x_{t+1} is given by:

$$r_1 = \frac{\sum_{t=1}^{N-1}(x_t-\bar{x}_{(1)})(x_{t+1}-\bar{x}_{(2)})}{\left[\sum_{t=1}^{N-1}(x_t-\bar{x}_{(1)})^2\right]^{1/2}\left[\sum_{t=1}^{N-1}(x_t-\bar{x}_{(2)})^2\right]^{1/2}} \qquad (7)$$

Where $\bar{x}_{(1)}$ is the mean of the first N – 1 observations and $\bar{x}_{(2)}$ is the mean of the last N -1 observation. As the correlation coefficient given above measure correlation between successive observations it is called the autocorrelation coefficient or serial correlation coefficient.

For N reasonably large, the difference between the sub-period means $\bar{x}_{(1)}$ and $\bar{x}_{(2)}$ can be ignored and r_1 can be approximated as by

$$r_1 = \frac{\sum_{t=1}^{N-1}(x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^{N}(x_t - \bar{x})^2} \tag{8}$$

Equation (8) can be generalized to give the correlation between observations separated by k years:

$$r_k = \frac{\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{N}(x_t - \bar{x})^2} \tag{9}$$

The quantity r_k is called the autocorrelation coefficient at lag k. The plot of the autocorrelation function as a function of lag is also called the correlogram.

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...
- Confidence bands

The confidence band uses the following formula if the autocorrelation plot is used to check for randomness in the data.

$$\frac{\pm z_{1-\alpha/2}}{\sqrt{N}} \tag{10}$$

Where *N* is the sample size, *z* is the percent point function of the standard normal distribution and α is the significance level.

If autocorrelation plots are also used in the model identification stage for fitting ARIMA models, the confidence band uses the following formula:

$$\pm z_{1-\alpha/2}\sqrt{\frac{1}{N}\left(1+2\sum_{i=1}^{k} y_i^2\right)} \tag{11}$$

Where k is the lag, N is the sample size; z is the percent point function of the standard normal distribution and α is the significance level.

3.5.0 PARTIAL AUTOCORELATION FUNCTION

Partial autocorrelation function measures the degree of association between Y_t and Y_{t+k} when the effect of other time lags on Y are held constant. The partial autocorrelation function PACF denoted by $\{r_{kk}: k = 1,2,3,\ldots\}$. The set of partial autocorrelations at various lags k are defined by

$$r_{kk} = \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j}\, r_{k-j}}{1-\sum_{j=1}^{k-1} r_{k-1,j}\ \ r_j} \tag{12}$$

where $r_{k,\ j} = r_{k-1,j} - r_{kk} r_{k-1,k-j}, \quad j = 1,2,\ldots,k-1$

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. The approximate 95% confidence interval for the partial autocorrelations is at $\pm 2/\sqrt{N}$.

Partial autocorrelation plots are formed by:

- Vertical axis: Partial autocorrelation coefficient 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.

3.6.0 TIME SERIES MODELS

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.

3.6.1 AUTOREGRESSIVE (AR) MODELS

A model in which future values are forecast purely on the basis of past values of the time series is called an Autoregressive (AR) process.

An autoregressive model of order p, denoted by AR(p) with mean zero is generally given by the equation:

$$y_t = \phi_1 y_{(t-1)} + \phi_2 y_{(t-2)} + \phi_3 y_{(t-3)} + \cdots + \phi_p y_{(t-p)} + \varepsilon_t \tag{13}$$

Or

$$y_t = (\phi_1 L + \phi_2 L^2 + \phi_3 L^3 + \cdots + \phi_P L^P) y_t + \varepsilon_t \tag{14}$$

$$\phi(L) y_t = \varepsilon_t \tag{15}$$

$$\phi(L) = (1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \cdots - \phi_P L^P) \tag{16}$$

Where:

L, is the lag operator

$\phi_1, \phi_2, \phi_3, \ldots \phi_p (\phi_p \neq 0)$ are the autoregressive model parameters and

ε_t is the random shock or white noise process, with mean zero and variance σ_ε^2.

The mean of y_t is zero.

If the mean μ, of y_t is not zero, replace y_t by $y_t - \mu$. That is

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \cdots + \phi_p(y_{t-p} - \mu) + \varepsilon_t \tag{17}$$

Or write

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \tag{18}$$

Where $\alpha = \mu(1 - \phi_1 - \phi_2 - \cdots - \phi_p)$. (19)

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.

A necessary condition for stationary is that $r_k = 0$ as $k \to \infty$.

3.6.2 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:

$$y_t = \theta_1 \varepsilon_{(t-1)} + \theta_2 \varepsilon_{(t-2)} + \theta_3 \varepsilon_{(t-3)} + \cdots + \theta_p \varepsilon_{(t-p)} + \varepsilon_t \tag{20}$$

Or

$$y_t = (\theta_1 L + \theta_2 L^2 + \theta_3 L^3 + \cdots + \theta_q L^q)\varepsilon_t \quad (21)$$

$$y_t = \theta(L)\varepsilon_t \quad (22)$$

$$\theta(L) = (\theta_1 L + \theta_2 L^2 + \theta_3 L^3 + \cdots + \theta_q L^q) \quad (22)$$

Where:

L and ε_t as defined above.

$\theta_1, \theta_2, \theta_3, \ \ldots, \theta_p$ are the moving average model parameters,

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 $\theta(L)$ can be inverted, in other words if it can be expressed as an AR. An MA (q) is invertible if the roots of $\theta(L) = 0$ all lie outside the unit circle. A finite AR is always invertible.

3.6.3 AUTOREGRESSIVE MOVING AVERAGE (ARMA) MODELS

Autoregressive and Moving Average processes can be combined to obtain a very flexible class of univariate processes (proposed by Box and Jenkins), known as ARMA processes.

The time series y_t is an ARMA (p, q) process, if it is stationary and

$$y_t = \phi_1 y_{(t-1)} + \cdots + \phi_p x_{(t-p)} + \varepsilon_t + \theta_1 \varepsilon_{(t-1)} + \cdots + \theta_q \varepsilon_{(t-q)} \quad (24)$$

$$y_t = \phi(L) y_{t-1} + \theta(L)\varepsilon_t \quad (25)$$

Or

$$\phi(L)y_t = \theta(L)\varepsilon_t \tag{26}$$

Where ϕ, θ, ε_t and L as defined above with $\theta_p \neq 0$ and $\phi_p \neq 0$.

An ARMA process is stationary if the roots of $\phi(L)$ all lie outside the unit circle and invertible if the roots of $\theta(L)$ all lie outside the unit circle

3.6.4 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.

This time series method was used to model the road traffic accident data collected along the Accra -Tema from 1989-2010.

3.7.0 ARIMA MODEL BUILDING PROCESS

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. The diagram in figure 3.1 below shows ARIMA or Box-Jenkins model building process.

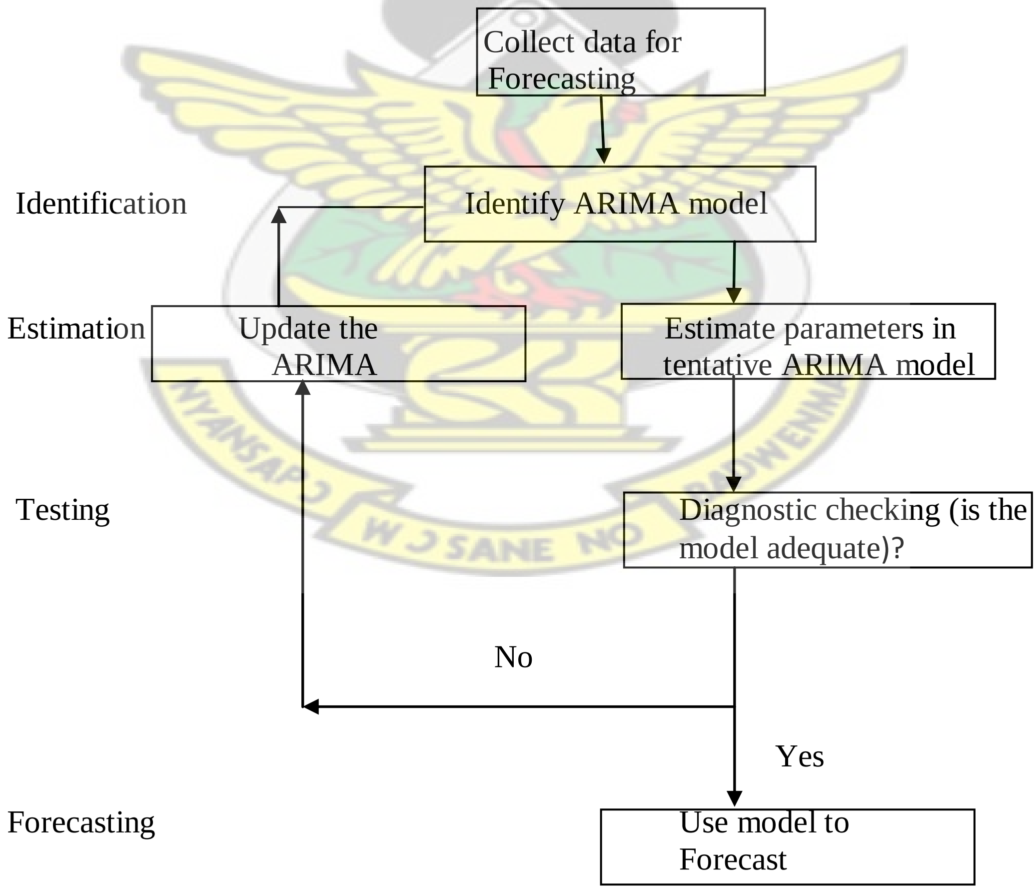

Figure 3.1: ARIMA model building process

3.7.1 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. The time series plot should show constant location and scale. It can also be detected from an autocorrelation plot. Specifically, non-stationarity is often indicated by an autocorrelation plot with very slow decay.

If the model is found to be non-stationary, stationarity could be achieved mostly by differencing the series, or go for Dickey Fuller test. Stationarity could also be achieved by some modes of transformation say log transformation.

Thus if Y_t denotes the original series, the non-seasonal difference of first order is

$$\text{If } X_t = Y_t - Y_{t-1} \tag{27}$$

where $t = 1,2,3,\ldots,T$

Once stationarity has been addressed, the next step is to identify the order (i.e., the p and q) of the autoregressive and moving average terms.

The primary tools for doing this are the autocorrelation plot and the partial autocorrelation plot. The sample autocorrelation plot and the sample partial autocorrelation plot are compared to the theoretical behavior of these plots when the order is known. Table 3.1 summarizes how we use the sample ACF/PACF for model identification.

Table 3.1: Theoretical behavior of the ACF and PACF for model identification.

PROCESS	ACF	PACF
AR(p)	Tails off	Cut off after the order p of the process
MA(q)	Cut off after the order q of the process	Tails off
ARMA(p,q)	Tails off	Tails off

3.7.1.1 Other tools for model identification

Akaike's Information Criteria (AIC) – The AIC which was proposed by Akaike uses the maximum likelihood method. In the implementation of the approach, a range of potential ARMA models is estimated by maximum likelihood methods, and for each, the AIC is calculated, given by:

$$\mathrm{AIC}(p,q) = \frac{-2\ln(\text{maximum likelihood}) + 2r}{n} \tag{28}$$

$$\mathrm{AIC}(p,q) = \ln(\hat{\sigma}_e^2) + r\frac{2}{n} + \text{constant} \tag{29}$$

Where;

N is the sample size or the number of observations in the historical time series data,

$\hat{\sigma}_e^2$ is the maximum liklihood estimate of σ_e^2 which is the residual or shock variance and

r=p+q+1, denotes the number of parameters estimated in the model.

Given two or more competing models the one with the smaller AIC value will be selected.

The AIC_C

The AIC is biased estimator and the bias can be appreciable for large parameter per data ratios. Hurvich and Tsai (1989) showed that the bias can be approximately eliminated by adding another non – static penalty term to the AIC, resulting in the corrected AIC, denoted by AIC_C and defined by the formula:

$$AIC_C = AIC + \frac{2(r+1)(r+2)}{n-r-2} \quad (30)$$

All parameters remained defined as above.

Schwarz's Bayesian Criterion (BIC)

Schwarz's BIC like AIC uses the maximum likelihood method. It is given by

$$BIC(p, q) = \ln(\hat{\sigma}_e^2) + r\frac{\ln(n)}{n}$$

(31)

Where all parameters remained defined as above.

The BIC imposes a greater penalty for the number of estimated model parameters than does than AIC. The use of minimum BIC for model selection results in a chosen model whose number of parameters is less than that chosen under AIC.

One disadvantage of the information criteria approach is the enormous work involved in computing maximum likelihood estimates of several models which is time consuming and expensive. However this problem has been overcome by the introduction of computers since there are softwares like R which can compute several of these information criteria.

3.7.2 Estimation Stage – Once a model is identified the next stage of the ARIMA model building process is to estimate the parameters. Estimating the parameters for the ARIMA (Box- Jenkins) models is a quite complicated non-linear estimation problem. For this reason, the parameter estimation should be left to a high quality software program that fits Box-Jenkins models.

Two approaches are used in the estimation, these include non-linear least squares and maximum likelihood estimation. In this study the estimation of the parameters was done using a statistical package called the R.

3.7.3 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:

3.7.3.1 Test of Significance of the Coefficients - In R, p-values are not given.

For each coefficient, $$t = \frac{\text{estimated coefficient}}{\text{standard error}} \qquad (32)$$

If $|t| \geq 2$, the estimated coefficient is significantly different from 0, then the model coefficient is statistically significant. If not, the model should probably be simplified, say, by reducing the model order. For example, an AR (2) model for which the second-order coefficient is not significantly different from zero might be discarded in favor of an AR (1) model.

3.7.3.2 Diagnostics of Residuals

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

Time Plot of the Residuals – Time plot of the standardized residuals should not show any structure. It must indicate no trend in the residuals, no outliers and in general case no changing variance across time.

Plot of Residual ACF – Once the appropriate ARIMA model has been fitted, one can examine the goodness of fit by means of plotting the ACF of residuals of the fitted model. If most of the sample autocorrelation coefficients of the residuals are within the limits $\pm 1.96/\sqrt{N}$ where N is the number of observations upon which the model is based then the residuals are white noise indicating that the model is a good fit.

The Normal Q-Q Plot - Another diagnostic check on the residuals is to determine whether it follows the normal distribution. This is done by using the normal Q – Q plot. Q-Q plot is a normal probability plot. It is a plot of the quantiles of two distributions against each other, or a plot based on estimates of the quantiles. The normal Q-Q plots is used to compare the distribution of a sample to a theoretical distribution. The normal Q – Q plots provide a quick way to visually inspect to what extent the pattern of data follows a normal distribution.

Testing the Model for Adequacy – After identifying an appropriate model for a time series data it is very important to check that the model is adequate. The error terms ε_t are examined and for the model to be adequate the errors should be random. If the error terms are statistically different from zero, the model is not considered adequate.

The test statistic used is the Ljung-Box statistic, also called the modified Box-Pierce statistic, is a function of the accumulated sample autocorrelations, r_j, up to any specified time lag m. As a function of m, it is determined as:

$$Q(m) = n(n+2)\sum_{j=1}^{m}\frac{r_j^2}{n-j} \quad (33)$$

Which is approximately distributed as a χ^2 with $df = n - p - q$,where p and q are orders of AR and MA respectively and n = number of usable data points after any differencing operations. This statistic can be used to examine residuals from a time series model in order to see if all underlying population autocorrelations for the errors may be 0 (up to a specified point). If the calculated value of Q is greater than χ^2 for $df = n - p - q$, then the model is considered inadequate and adequate if Q is less than χ^2 for $df = n - p - q$.

A p-value is calculated as the probability past Q (m) in the relevant distribution. A small p-value (for instance, p-value < .05) indicates the possibility of non-zero autocorrelation within the first m lags which makes the model inadequate and then an alternative model needs to be selected.

3.7.4 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.

The autoregressive representation

$$y_t = \alpha + \sum_{u=1}^{\infty}\phi_u * y_{(t-u)} + \varepsilon_t \quad (34)$$

suggests predicting the next observation beyond $y_1, \dots, y_T$ using

$$\hat{y}_{T+1} = \alpha + \sum_{u=1}^{\infty}\hat{\phi}_u * y_{(T+1-u)} + \varepsilon_t \quad (35)$$

Where the $\widehat{\phi}_u$ are obtained by substituting the estimated parameters in place of the theoretical ones. Once a forecast is obtained for y_{T+1} we can use it to obtain a forecast for y_{T+2} and then use these two forecasts to generate a forecast for y_{T+3}. The process can be continued to obtain forecasts out to any point in the future. In this study 10 years forecasts were made into the future.

3.7.5 METHODOLOGY

A secondary data which involved thirty one years of annual RTA cases, injuries and deaths from the period 1980-2010 were collected from the national MTTU. R statistical software was used in the analysis. In order to tentatively identify Box-Jenkins model, we must first determine whether the time series we wish to analyze is stationary.

Time plots of each of the data analyzed in R were found to have an increasing trend and this was verified by the slow decay of the ACF plots. Thus the RTA data is

non – stationary. Descriptive analysis of the RTA data was done using the time plots

The data was made stationary by removing the trend and this was done by differencing it once. Time plots were subsequently produced to verify that the data was now stationary.

Once the RTA data was made stationary, we used the sample ACF and PACF plots produced from the R output in order to identify various Box- Jenkins models for each of the three accident data.

Estimates of the models' parameters were produced and their statistical significance tested. Residuals from the models were also checked in order to identify if the residuals are white noise.

Suitable models were each selected and fitted for each differenced data based on their AIC, $AIC_{C,}$ and BIC values and finally

The models selected were used in making forecasts for the values of the RTA cases, injuries and deaths.

CHAPTER 4

RESULTS AND ANALYSIS

This chapter presents the analysis of accident data collected along the Accra – Tema motorway from the years 1980-2010 which involves number of accident cases, persons' injured and persons' killed. The R statistical software was used for the analysis and various tentative time series ARIMA models developed were fitted to each data and the suitable models were selected based on diagnostics of the residuals of each model and other criteria.

Ten years' forecasts were estimated using the best models for each data.

4.1.0 DATA PRESENTATION

Accident data on the Accra – Tema Motorway spanning the period 1980-2010 were compiled from the National MTTU, Accra which involves number of accident cases, number of injured and number of persons' killed. See appendix for the data presentation.

4.2.0 ANALYSIS OF THE MOTORWAY ACCIDENT CASES

4.2.1 DESCRIPTIVE ANALYSIS OF THE ACCIDENT CASES DATA

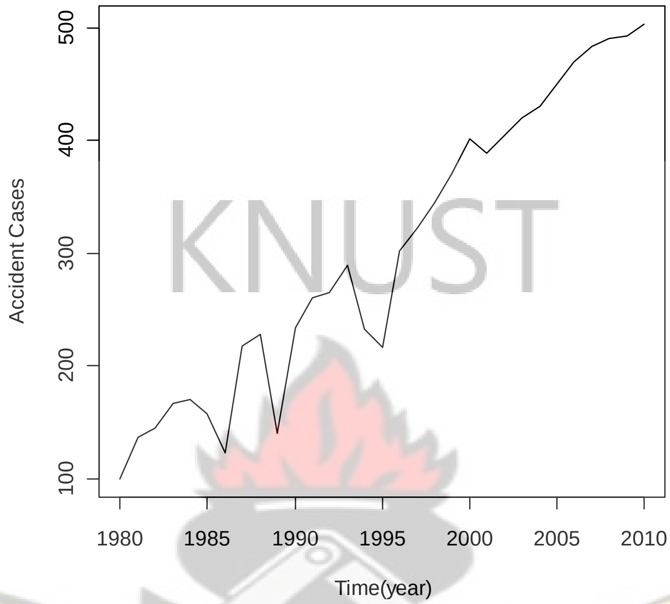

Figure 4.1: Time plot of Motorway accident data cases from 1980-2010

Figure 4.1 shows the time plot of the Accra – Tema accident data cases from 1980 – 2010. There is a systematic change in the time plot in Figure 4.1 which is known as the trend. Accident cases increased from 1980 to 1984 followed by a decrease from 1985 to 1987. An irregular pattern was observed from 1988 to 1990 followed by an increase to 1993 followed again by a decrease to 1995.

A severe increase was observed in 1996 and this continued to 2001 followed by a decrease in 2002. The accident cases increased sharply from 2003 to 2010.

In general, the trend in accident cases along the Accra – Tema Motorway is increasing but not always the case.

The annual motorway accident cases in Figure 4.1 do not exhibit seasonal variation and it is not stationary due to the trend component.

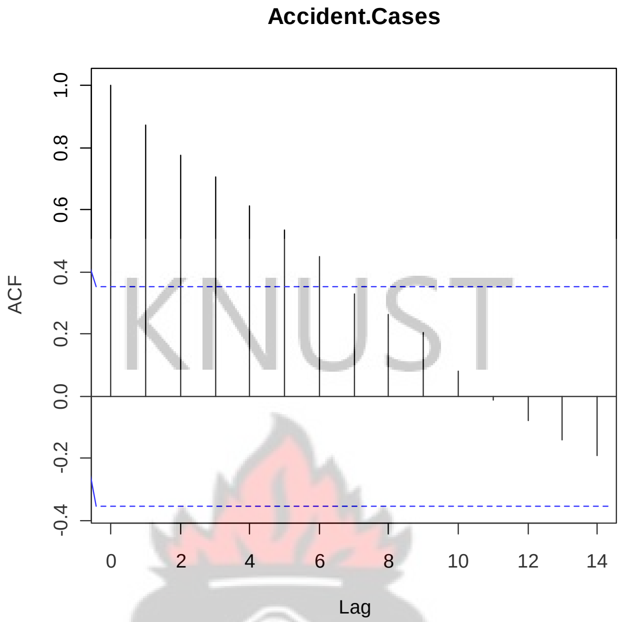

Figure 4.2: Autocorrelation function of Accra-Tema motorway accident cases.

The autocorrelation function of motorway accident cases is shown in Figure 4.2 which describes the correlation between values of the motorway accident cases at different points in time, as a function of the two times or the time difference.

The first several autocorrelations are persistently large and trailed off to zero rather slowly. A trend exists and this time series is non-stationary (it does not vary about a fixed level).

4.2.2 TREND DIFFERENCING OF THE ACCIDENT DATA CASES

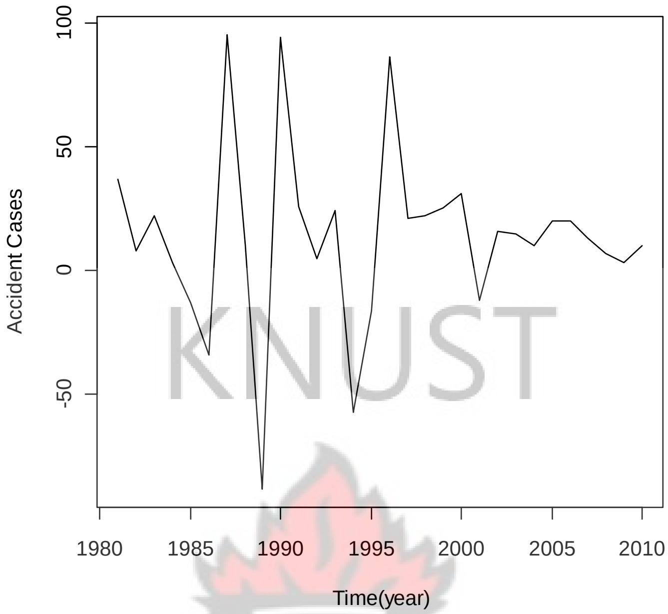

Figure 4.3: First difference of the motorway accident cases.

A transformation of the motorway accident cases data using the first differencing method is performed to remove the trend component in the original accident data cases which is shown in Figure 4.3. The observations move irregularly but revert to its mean value and the variability is also approximately constant. The motorway data now looks to be approximately stable.

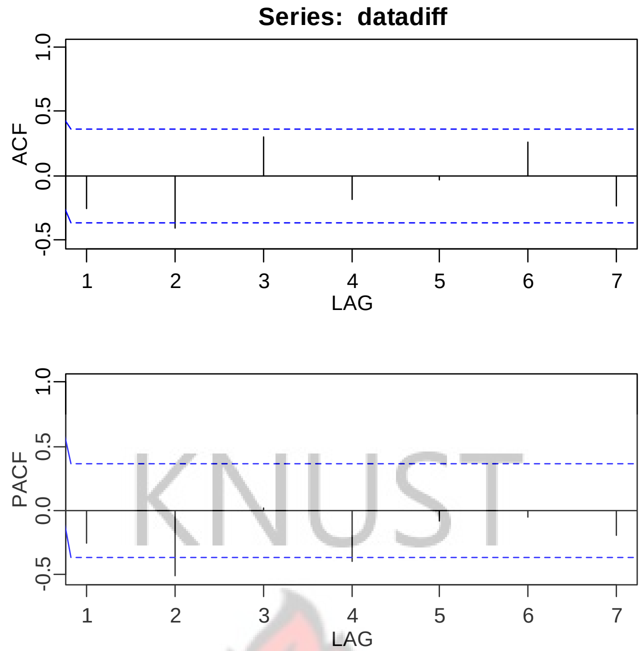

Figure 4.4: ACF and PACF plots of the first differencing of the accident data cases.

The top part of Figure 4.4 shows the autocorrelation function of the first differencing of the motorway accident data at various lags and the bottom part is the partial autocorrelation function of the first differencing of the motorway accident data also at different lags.

Comparing the autocorrelations with their error limits, the only significant autocorrelation is at lag 2, indicating an MA (2) behavior. Similarly, only the lags 2 and 4 partial autocorrelations are significant, indicating an AR (2) or AR (4) but applying the principle of parsimony we use AR (2). The following models are suggested;

- ARIMA(2,1,0)
- ARIMA(0,1,2)
- ARIMA(2,1,2)

To select the best model for forecasting into the future, each model is assessed based on its parameter estimates, the corresponding diagnostics of the residuals and the AIC, BIC and AIC_C.

4.2.3 MODEL SELECTION FOR THE ACCIDENT CASES DATA

4.2.3.1 PARAMETER ESTIMATES AND DIAGNOSTICS OF ARIMA (2, 1, 0) MODEL

Coefficients:

	ar1	ar2	constant
	-0.7060	-0.6288	-0.6263
s.e.	0.1372	0.1304	3.2325
t – value	5.1458	4.8221	0.1936

Sigma^2 estimated as 1556: log likelihood = -148.33, aic = 304.66

$AIC

[1] 8.549788

$AICc

[1] 8.669788

$BIC

[1] 7.689908

The parameters based on the t – value estimates are statistically significant since the t – values are each greater than 2 in absolute except the constant term.

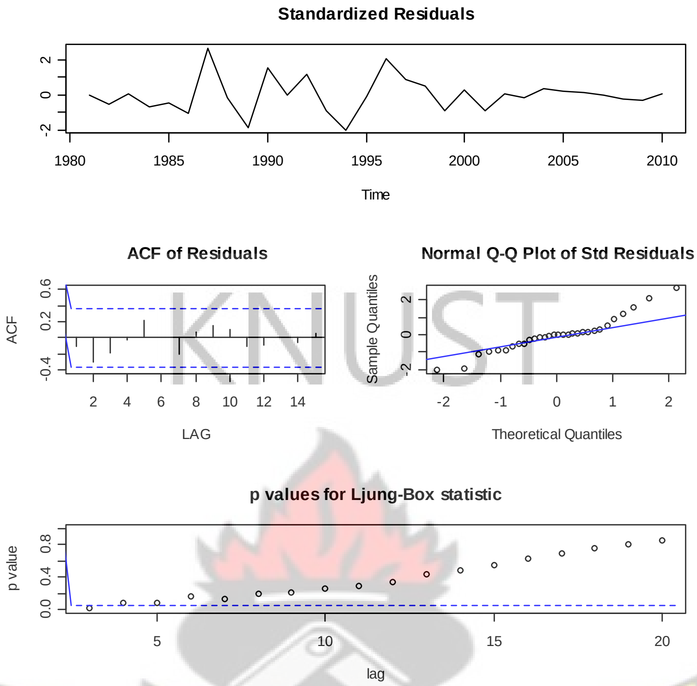

Figure 4.5: Diagnostics of ARIMA (2, 1, 0)

Diagnostics of the residuals from ARIMA (2, 1, 0) is shown in Figure 4.5 above.

a) The top part is the time plot of the standardized residuals of ARIMA (2, 1, 0). The standardized residuals plot shows no obvious pattern and looks like an i.i.d. of mean zero with few outliers.

b) The middle part of the diagnostics is the plot of the ACF of the residuals. There is no evidence of significant correlation in the residuals at any positive lag.

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 are located on the straight line except some few residuals deviating from the normality. Therefore the normality assumption looks to be satisfied and so the residuals appear to be normally distributed.

d) The bottom part of the diagnostics is the time plot of the Ljung – Box statistics. It is observed that one of the p – values of the Ljung – Box statistics plot is significant. This does not provide evidence for a significant at any positive lag.

In general the model fits well and it is adequate.

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

Coefficients:

	ar1	ar2	ma1	ma2	constant
	-0.7159	-0.5242	-0.4936	-0.5064	0.0543
s.e.	0.2563	0.1481	0.3270	0.3164	0.4562
t – value	2.7932	3.5395	1.5095	1.6005	0.1190

sigma^2 estimated as 937.7: log likelihood = -142.8, aic = 297.59

$AIC

[1] 8.176784

$AICc

[1] 8.36519

$BIC

[1] 7.410317

The parameters based on the t – value estimates of the MA coefficients are statistically not significant since the t – values are each less than 2 in absolute value.

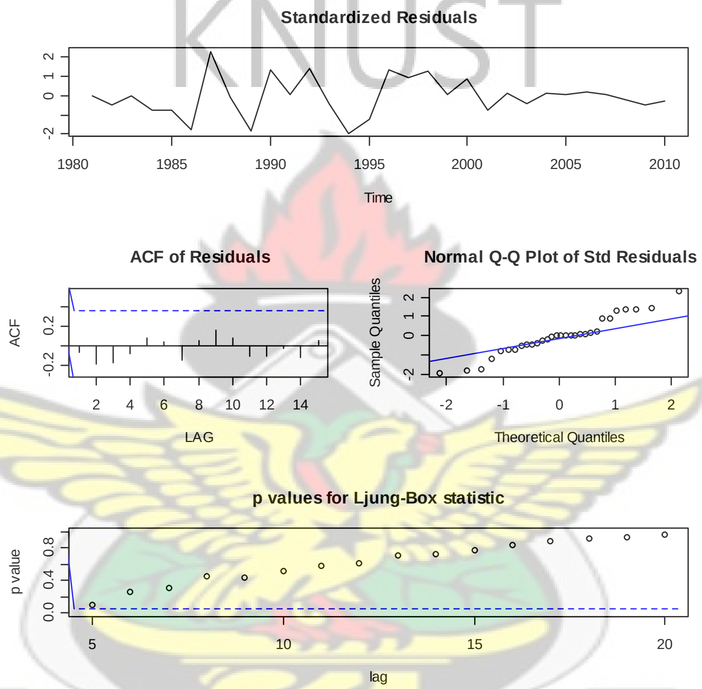

Figure 4.6: Diagnostics of ARIMA (2, 1, 2)

Diagnostics of the residuals from ARIMA (2, 1, 2) is shown in Figure 4.6 above.

a) The standardized residuals plot shows no obvious pattern and looks like an i.i.d. of mean zero with few outliers.

b) The middle part of the diagnostics is the plot of the ACF of the residuals. There is no evidence of significant correlation in the residuals at any positive lag.

c) The Q-Q plot is a normal probability plot. It doesn't look too bad, so the assumption of normally distributed residuals looks okay.

d) It is also observed that the p – values for the Ljung-Box statistics plot is not significant at any positive lag. That is all p – values are greater than 0.05.

4.2.3.3 PARAMETER ESTIMATES AND DIAGNOSTICS OF ARIMA (0, 1, 2) MODEL

Coefficients:

	ma1	ma2	constant
	-1.9940	0.9973	0.2576
s.e	0.4204	0.4205	0.1410
t – value	4.7431	2.3717	1.8270

Sigma^2 estimated as 833.3: log likelihood = -144.05, aic = 296.1

$AIC

[1] 7.925451

$AICc

[1] 8.045451

$BIC

[1] 7.065571

The parameters based on the t – value estimates are statistically significant since the t – values are each greater than 2 in absolute value except the constant term.

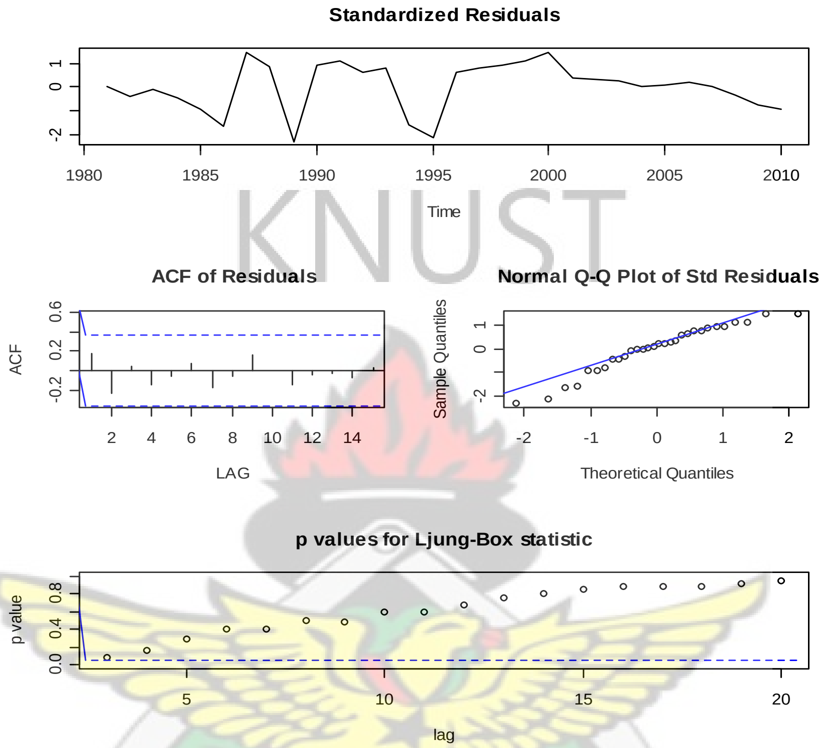

Figure 4.7: Diagnostics of ARIMA (0, 1, 2)

Diagnostics of the residuals from ARIMA (0, 1, 2) is shown in Figure 4.7 above.

a) The standardized residuals plot shows no obvious pattern and looks like an i.i.d. of mean zero with few outliers.

b) The ACF of residuals plot shows no significant residual autocorrelation for the ARIMA (0, 1, 2) model.

c) The normal Q-Q plot of the residuals doesn't look too bad, so the assumption of normally distributed residuals look okay.

d) The p – values for the Ljung-Box statistics plot is not significant at any positive lag. That is all p – values are greater than 0.05.

In general the model fits well and it is adequate.

Table 4.1: Parameter estimates and diagnostics for models selection for RTA cases

Model	**Test on Parameter estimates**				
	Parameter	Estimate	S.E	t- value	Sig. if t≥|2|
ARIMA(2,1,0)	ar 1	-0.7060	0.1372	5.1458	Sig
	ar 2	0.6288	0.1304	4.8221	Sig
	Constant	-0.6263	3.2325	0.1938	Non-sig
ARIMA(2,1,2)	ar 1	-0.7159	0.2563	2.7932	Sig
	ar 2	-0.5242	0.1481	3.5395	Sig
	ma 1	-0.4936	0.3270	1.5095	Non- sig
	ma 2	-0.5064	0.3164	1.6005	Non-sig
	Constant	0.0543	0.4562	0.1190	Non-sig
ARIMA(0,1,2)	ma 1	-1.9940	0.4204	4.7431	Sig
	ma 2	0.9973	0.4205	2.3717	Sig
	Constant	0.2576	0.1410	1.8270	Non Sig

DIAGNOSTICS			
	ARIMA(2,1,0)	**ARIMA(2,1,2)**	**ARIMA(0,1,2)**
AIC	8.549788	8.176784	7.925451
AIC_C	8.669788	8.36519	8.045451
BIC	7.689908	7.410317	7.065571

4.2.4 SELECTION OF BEST MODEL FOR FORECASTING RTA CASES

The standardized residuals plots of all the models are independently and identically distributed with mean zero and some few outliers. There is 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 are not significant at any positive lag for all the models except ARIMA (2, 1, 0) which has one of the p – values of the Ljung – Box statistics less than 0.05.

From table 4.1 above all the parameters in the MA coefficients of ARIMA(2,1,2) model are not significant at 5% level of significance while the parameters in the ARIMA(2,1,0) and ARIMA(0,1,2) models are significant except the constant terms.

The AIC, AIC_C and BIC are good for all the models but they favor ARIMA (0, 1, 2) model.

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

4.2.5 FITTING THE ACCIDENT CASES MODEL

ARIMA (0, 1, 2) model is the best model for forecasting the motorway accident cases. This is a non – seasonal integrated moving average with one level of differencing without AR terms. The model in terms of the differenced series x_t is given as:

$$x_t = \alpha + \varepsilon_t - \theta_1 \varepsilon_{(t-1)} - \theta_2 \varepsilon_{(t-2)} \tag{36}$$

In terms of the observed series the model becomes,

$$y_t - y_{t-1} = \alpha + \varepsilon_t - \theta_1 \varepsilon_{(t-1)} - \theta_2 \varepsilon_{(t-2)} \tag{37}$$

Explicit representation of the above series in terms of the white noise process, ε_t is more difficult than in the stationary case (Cryer and Chan,2008).

and we write;

$$y_t = \alpha + y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{(t-1)} - \theta_2 \varepsilon_{(t-2)} \quad (38)$$

The point estimate of each parameter of ARIMA (0, 1, 2) from table 4.1 are as follows:

$\hat{\alpha} = 0.2576, \quad \hat{\theta}_1 = -1.9940, \quad \hat{\theta}_2 = 0.9973$

All the estimates are significant except the constant hence it is dropped.

The fitted ARIMA (0, 1, 2) model for the motorway accident cases from 1980 – 2010 is given by;

$$\hat{y}_t = y_{t-1} + \varepsilon_t + 1.9940\varepsilon_{t-1} - 0.9973\varepsilon_{t-2} \quad (39)$$

where ε_t has an estimated variance of 833.3.

4.2.6 FORECASTING MOTORWAY ACCIDENT CASES

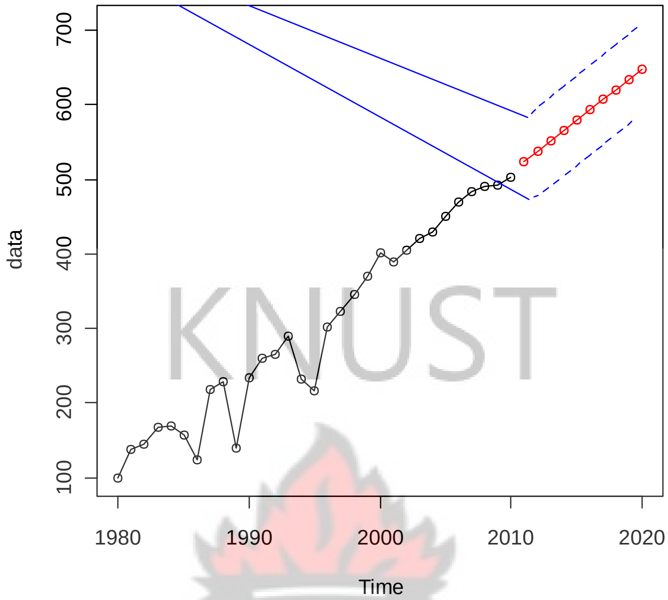

Figure 4.8: Graph of the accident cases, its forecasts and confidence intervals

Figure 4.8 gives the visual representation of the original motorway 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, the Accra – Tema motorway accident cases will continue to increase in the next 10 years.

10 steps prediction into the future;

$pred

Time Series:

Start = 2011

End = 2020

Frequency = 1

[1] 521.0033 537.0175 550.8698 564.7220 578.5743 592.4265 606.2788 620.1311

[9] 633.9833 647.8356

$se

Time Series:

Start = 2011

End = 2020

Frequency = 1

[1] 28.06339 33.27209 33.27209 33.27209 33.27209 33.27209 33.27209 33.27209

[9] 33.27209 33.27209

4.3.0 ANALYSIS OF THE MORTORWAY ACCIDENT INJURY DATA

4.3.1 DESCRIPTIVE ANALYSIS OF THE ACCIDENT INJURY DATA

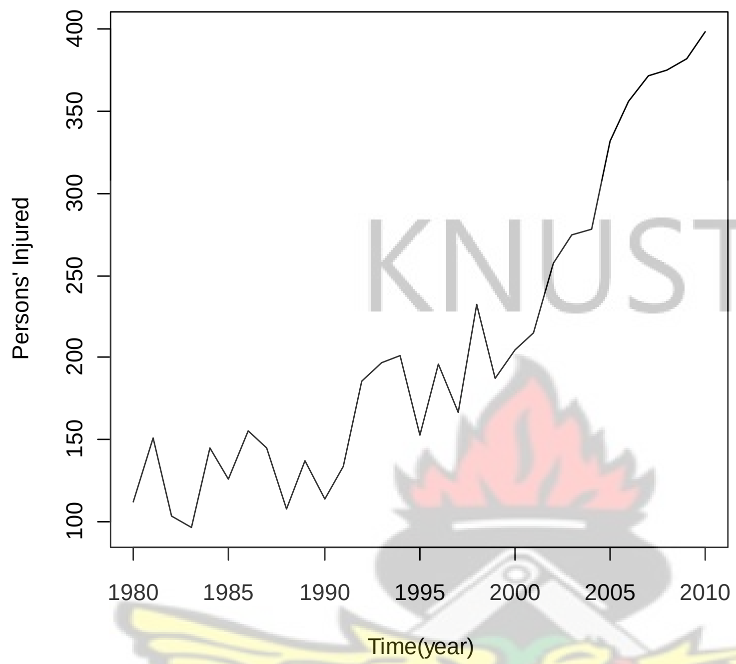

Figure 4.9: Time plot of Accra – Tema accident injury data from 1980-2010

Figure 4.9 shows the time plot of the motorway accident injury data from 1980 – 2010. There is a systematic change in the time plot in Figure 4.9, which is not periodic known as the trend. Accident injuries increased from 1980 to 1981 followed by a decrease to 1983. An irregular pattern was observed from 1984 to 2000. The accident injuries were observed to have increased from 2001 to 2010.

In general, the trend in the motorway accident injuries is increasing but not always the case.

The annual motorway accident injuries time plot in Figure 4.9 does not exhibit seasonal variation and it is not stationary due to the trend component.

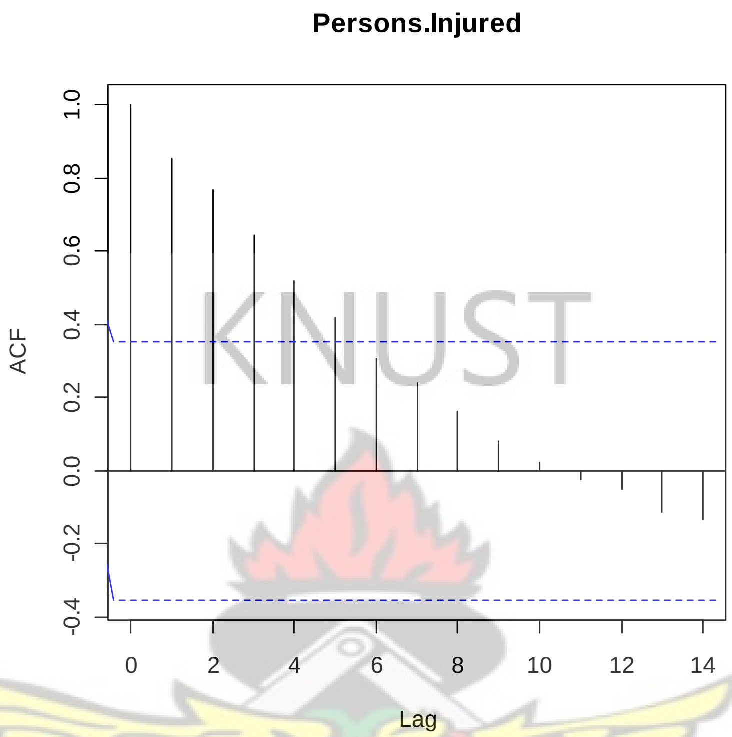

Figure 4.10: Autocorrelation function of the motorway accident injuries

The autocorrelation function of the motorway accident injuries is shown in Figure 4.10 which shows that the autocorrelation function is increasing gradually and that there is a trend in the motorway accident injury data.

4.3.2 TREND DIFFERENCING OF THE ACCIDENT INJURY DATA

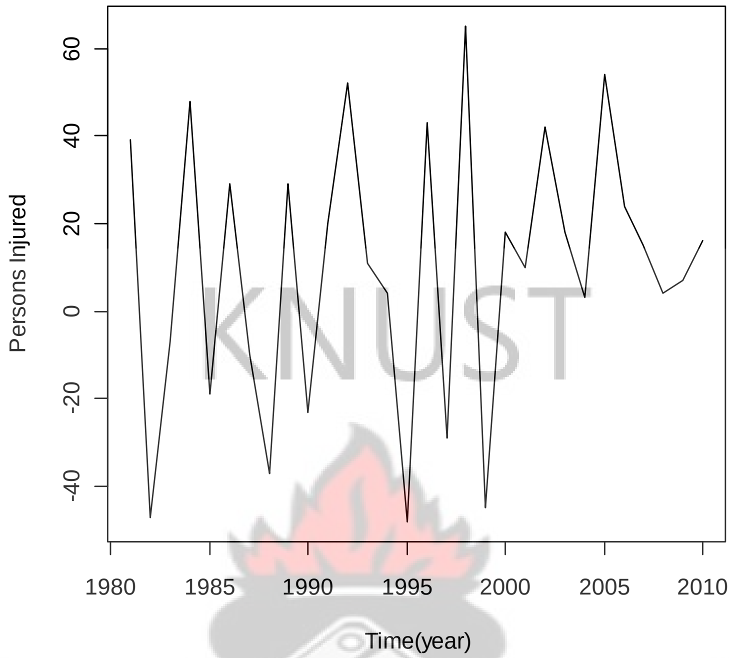

Figure 4.11: First difference of the motorway accident injury data.

A transformation of the motorway accident injury data using the first differencing method is performed to remove the trend component in the original motorway accident injury data which is shown in Figure 4.11. The observations move irregularly but revert to its mean value and the variability is also approximately stable.

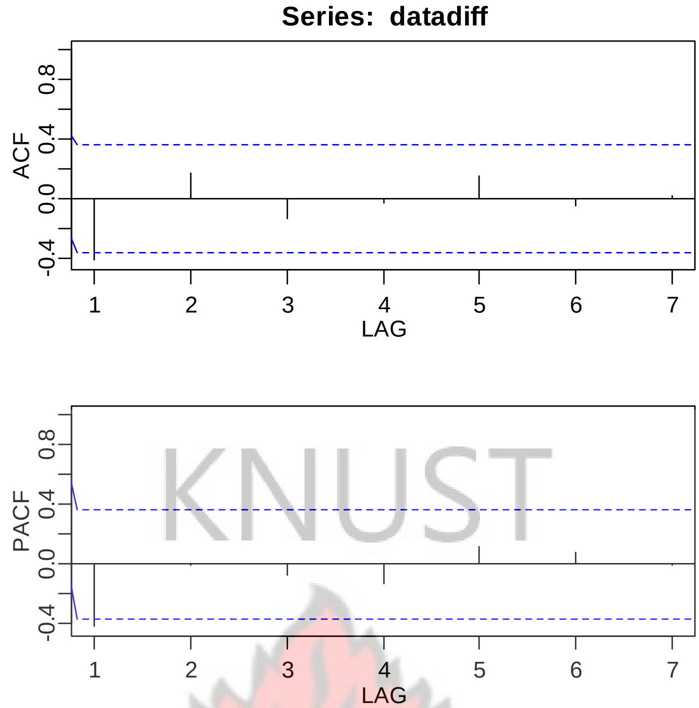

Figure 4.12 ACF and PACF of the first differencing of motorway accident injury data

The Figure 4.12 above shows the ACF and the PACF of the first differencing of motorway accident data. It can be observed that both the ACF and the PACF cut off after lag one, indicating MA (1) and AR (1) respectively. The following models are suggested;

- ARIMA (1,1,0)
- ARIMA (0,1,1)
- ARIMA (1,1,1)

To select the best model for forecasting into the future, each model is assessed based on its parameter estimates, the corresponding diagnostics of the residuals and the AIC, AIC_C and BIC values

4.3.3 MODEL SELECTION FOR MOTORWAY ACCIDENT INJURY DATA

4.3.3.1 PARAMETER ESTIMATES AND DIAGNOSTICS OF ARIMA (1, 1, 0) MODEL

Coefficients:

	ar1	constant
	-0.7237	0.3391
s.e.	0.1290	4.0035
t – value	5.6101	0.0847

Sigma^2 estimated as 1340: log likelihood = -145.93, aic = 297.85

$AIC

[1] 8.33369

$AICc

[1] 8.431126

$BIC

[1] 7.427104

The parameter based on the t – value test of the model above is statistically significant since it is greater than 2 in absolute value except the constant.

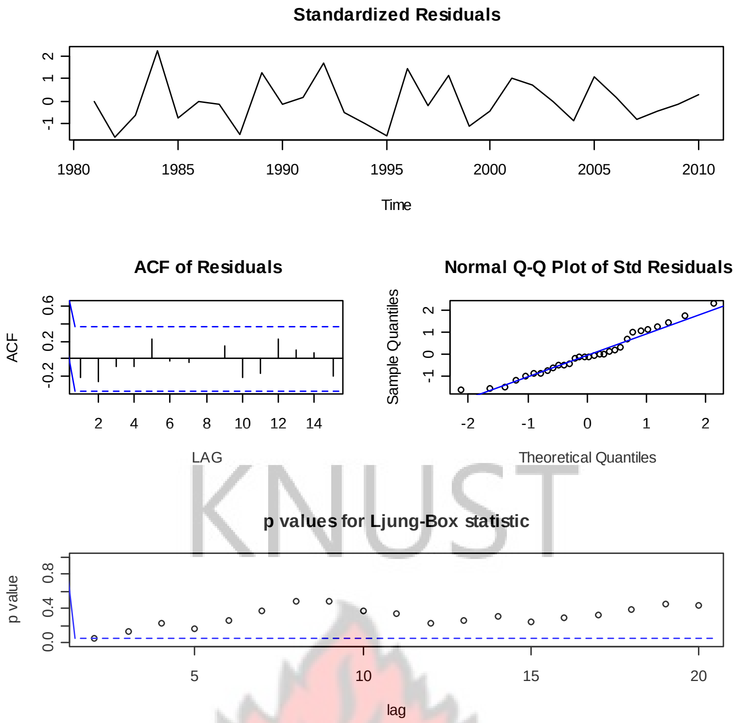

Figure 4.13: Diagnostics of ARIMA (1,1,0)

Diagnostics of the residuals from ARIMA (1, 1, 0) is shown in Figure 4.13 above. The time plot of the standardized residuals shows no obvious pattern and look like an i.i.d. sequence of mean zero with some few outliers.

The plot of the ACF of the residuals of the Figure 4.13 shows no evidence of significant correlation at any positive lag.

The normal Q – Q plot of the standardized residuals of the above figure indicates that most of the residuals are located on the straight line except some few residuals deviating from normality. Therefore the normality assumption looks to be satisfied and so the residuals appear to be normally distributed.

All the p – values of the Ljung – Box statistic are not significant at any positive lag. That is no p – value is less than 0.05.

4.3.3.2 PARAMETER ESTIMATES AND DIAGNOSTICS OF ARIMA (0, 1, 1) MODEL

Coefficients:

	ma1	constant
	-0.9999	0.6293
s.e.	0.0873	0.6502
t – value	11.4536	0.9679

Sigma^2 estimated as 950.2: log likelihood = -142.27, aic = 290.54

$AIC

[1] 7.989966

$AICc

[1] 8.087402

$BIC

[1] 7.083379

The parameter based on the t – value is statistically significant since it is greater than 2 in absolute value except the constant.

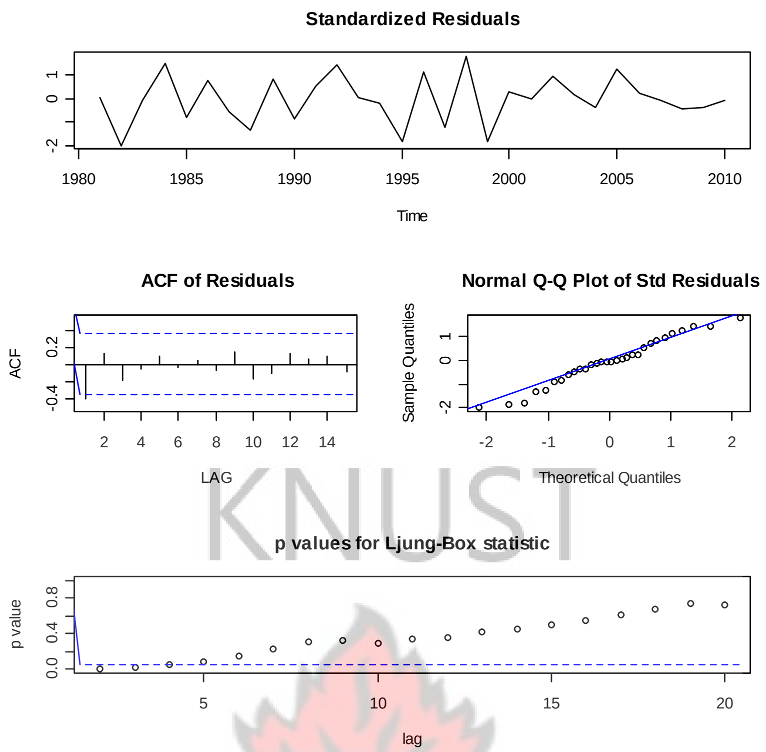

Figure 4.14: Diagnostics of ARIMA (0,1,1)

Diagnostics of the residuals from ARIMA (0, 1, 1) is shown in Figure 4.14 above. The time plot of the standardized residuals shows no obvious pattern and look like an i.i.d. sequence of mean zero with some few outliers.

The plot of the ACF of the residuals of the Figure 4.14 shows an evidence of significant correlation in the residuals at lag one which we do not want. All the autocorrelations of the residuals must be non- significant.

The normal Q – Q plot of the standardized residuals indicates that most of the residuals are located on the straight line except some few residuals deviating from normality. Therefore the normality assumption looks to be satisfied and so the residuals appear to be normally distributed.

The diagnostics of the time plot of the Ljung – Box statistics indicate some amount of significance at positive lags, we want non – significance p - values.

4.3.3.3 PARAMETER ESTIMATES AND DIAGNOSTICS OF ARIMA (1, 1, 1) MODEL

Coefficients:

	ar1	ma1	constant
	-0.4663	-1.0000	0.7237
s.e.	0.1646	0.0942	0.4013
t – value	2.8329	10.6157	1.8034

sigma^2 estimated as 727: log likelihood = -138.88, aic = 285.77

$AIC

[1] 7.78893

$AICc

[1] 7.90893

$BIC

[1] 6.92905

The parameters based of the AR coefficients are statistically significant since they are each greater than 2 in absolute value except the constant.

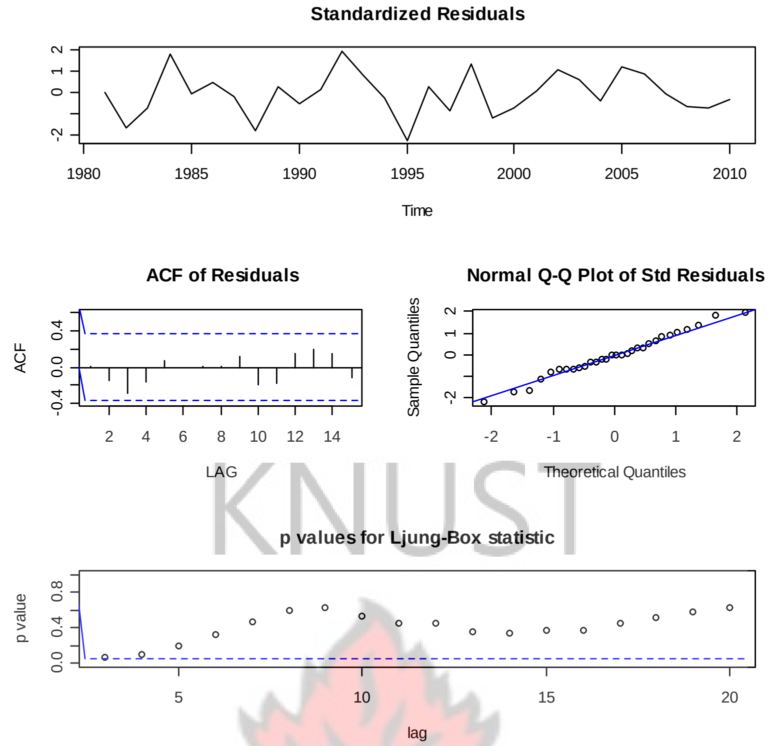

Figure 4.15: Diagnostics of ARIMA (1, 1, 1) model.

Diagnostics of the residuals from ARIMA (1, 1, 1) is shown in Figure 4.15 above. The time plot of the standardized residuals shows no obvious pattern and look like an i.i.d. sequence of mean zero with some few outliers.

The plot of the ACF of the residuals of the Figure 4.15 shows no evidence of significant correlation at any positive lag.

The normal Q – Q plot of the standardized residuals of the above figure indicates that most of the residuals are located on the straight line except some few residuals deviating from normality. Therefore the normality assumption looks to be satisfied and so the residuals appear to be normally distributed.

All the p – values of the Ljung – Box statistic are not significant at any positive lag. That is no p – value is less than 0.05.

Table 4.2: Parameter estimates and diagnostics of models selected for RTA injury data

Model	**Test on Parameters**				
	Parameter	**Estimate**	**S.E**	**t- value**	**Sig. if t≥|2|**
ARIMA(1,1,0)	ar 1	-0.7237	0.1290	5.6101	Sig
	Constant	0.3391	0.40035	0.0847	Non –sig
ARIMA(1,1,1)	ar 1	-0.4663	0.1646	2.8329	Sig
	ma 1	-1.0000	0.0942	10.6157	Sig
	Constant	0.7237	0.4013	1.8034	Non sig
ARIMA(0,1,1)	ma 1	-0.9999	0.0873	11.4536	Sig
	constant	0.6293	0.6502	0.9679	Non-sig

DIAGNOSTICS			
	ARIMA(1,1,0)	**ARIMA(1,1,1)**	**ARIMA(0,1,1)**
AIC	8.33369	7.78893	7.989966
AIC_C	8.431126	7.90893	8.087402
BIC	7.427104	6.92905	7.083379

4.3.4 SELECTION OF BEST MODEL FOR FORECASTING THE ACCIDENT INJURY

The standardized residuals plots of all the models are independently and identically distributed with mean zero and some few outliers. There is no evidence of significance in the autocorrelation functions of the residuals of all the models except ARIMA (0,1,1) which has some evidence of significance in the ACF of the residuals. The residuals appear to be normally distributed in all the models. The Ljung – Box statistics are not significant at any positive lag for all the models except ARIMA (0, 1, 1) which has some of the p – values of the Ljung – Box statistics less than 0.05.

From table 4.2 above all the parameters in the coefficients of ARIMA(1,1,0), ARIMA (0,1,1) and ARIMA (1,1,1) models are statistically significant at 5% level of significance except their constants.

The AIC, AIC_C and BIC are good for all the models but they favor ARIMA (1, 1, 1) model.

From the discussion above it is clear that ARIMA (1, 1, 1) model is the best model for forecasting the motorway accident injury.

4.3.5 FITTING THE ACCIDENT INJURY MODEL

ARIMA (1, 1, 1) model is the best model for forecasting the motorway accident injury, this is a non- seasonal model with one AR term and one MA term. The model in terms of the differenced series x_t is given by:

$$x_t = \alpha + \phi_1 x_{(t-1)} + \varepsilon_t - \theta_1 \varepsilon_{(t-1)} \tag{40}$$

In terms of the observed series the model becomes

$$y_t - y_{t-1} = \alpha + \phi_1 (y_{(t-1)} - y_{(t-2)}) + \varepsilon_t - \theta_1 \varepsilon_{(t-1)} \tag{41}$$

and we may write as;

$$y_t = \alpha + (1 + \phi_1)\, y_{t-1} - \phi_1 y_{(t-2)} + \varepsilon_t - \theta_1 \varepsilon_{(t-1)} \tag{42}$$

This is called the difference equation form of the model and appears to be ARMA (2,1) process, (Cryer and Chan, 2008).

The point estimate of each parameter of ARIMA (1,1, 1) from table 4.2 are as follows:

$\hat{\alpha} = 0.7237,\ \hat{\theta}_1 = -1.0000,\ \hat{\phi}_1 = -0.4663$

All the estimates are significant except the constant hence it is dropped.

Hence the fitted ARIMA (1, 1, 1) model for the motorway accident injury data from 1980 – 2010 is given by;

$$\hat{y}_t = (1 - 0.4663) y_{t-1} + 0.4663 y_{t-2} + \varepsilon_{t-1}$$

and we write

$$\hat{y}_t = 0.5337 y_{t-1} + 0.4663 y_{t-2} + \varepsilon_{t-1} \qquad (43)$$

Where ε_t has an estimated variance of 727.

4.3.6 FORECASTING MOTORWAY ACCIDENT INJURY

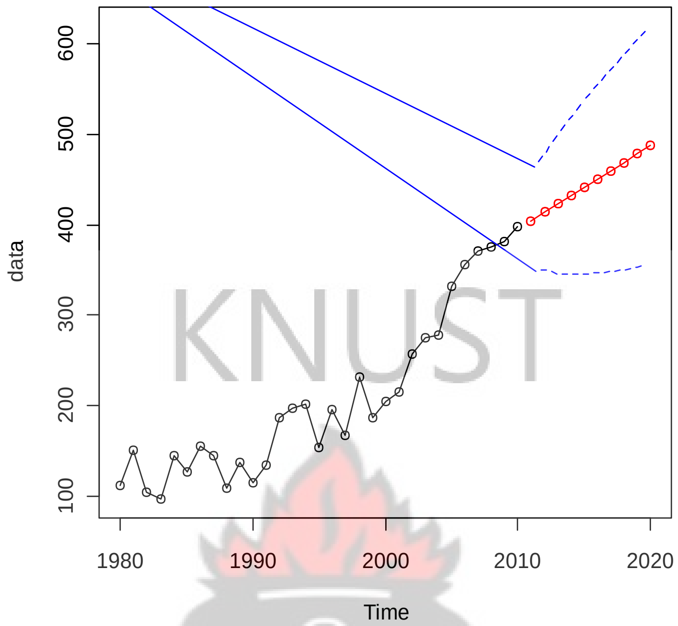

Figure 4.16: Graph of the accident injury data, its forecasts and confidence interval.

Figure 4.16 gives the visual representation of the original motorway accident injury 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, the Accra – Tema motorway accident injury will continue to increase in the next 10 years.

10 steps prediction into the future with the model is shown below;

$pred

Time Series:

Start = 2011

End = 2020

Frequency = 1

[1] 404.3524 414.8146 423.3716 432.8117 441.8425 451.0631 460.1956 469.3690

[9] 478.5234 487.6866

$se

Time Series:

Start = 2011

End = 2020

Frequency = 1

[1] 27.93275 32.51603 39.19084 43.75270 48.34963 52.34400 56.14136 59.65967

[9] 62.99841 66.16153

4.4.0 ANALYSIS OF THE MOTORWAY ACCIDENT MORTALITY DATA

4.4.1 DESCRIPTIVE ANALYSIS OF THE ACCIDENT MORTALITY DATA

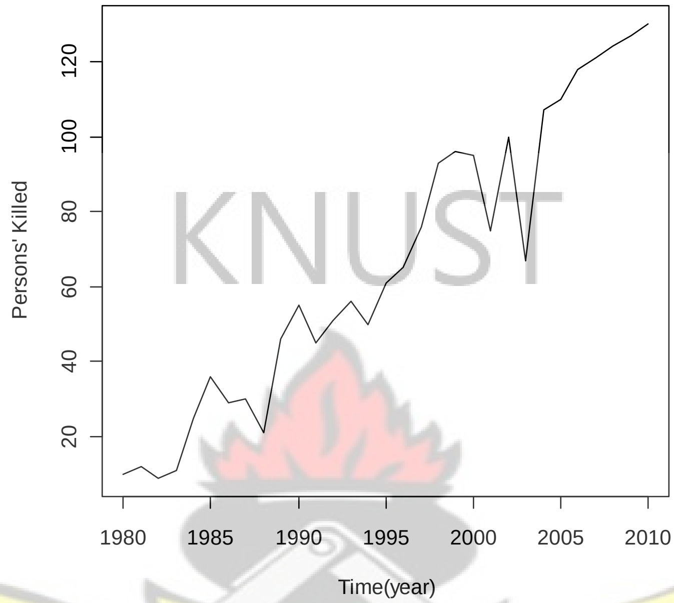

Figure 4.17: Time plot of the accident mortality data from 1980 – 2010

Figure 4.17 shows the time plot of the motorway accident mortality data from 1989 – 2010. There is a systematic change in the time plot in the figure above which is known as trend. Mortality increased by small amount from 1980 to 1981 followed by an increasing and decreasing movement of the mortality data. There was a sharp increase from 2004 and this increase continued to 2010.

In general, the trend in the motorway accident mortality is increasing but not always the case. The annual motorway accident mortality time plot in the Figure 4.17 is not stationary due to the trend component.

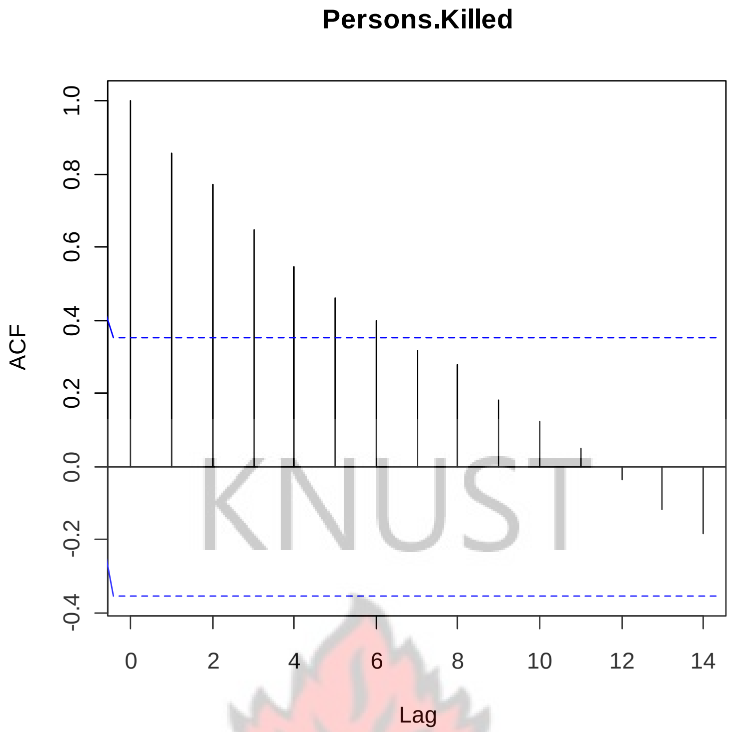

Figure 4.18: Autocorrelation function of the motorway accident mortality data.

The autocorrelation function of the motorway accident mortality is shown in Figure 4.18.The autocorrelation function is decreasing gradually and that shows that there is a trend in the motorway accident mortality data.

4.4.2 TREND DIFFERENCING OF THE ACCIDENT MORTALITY DATA

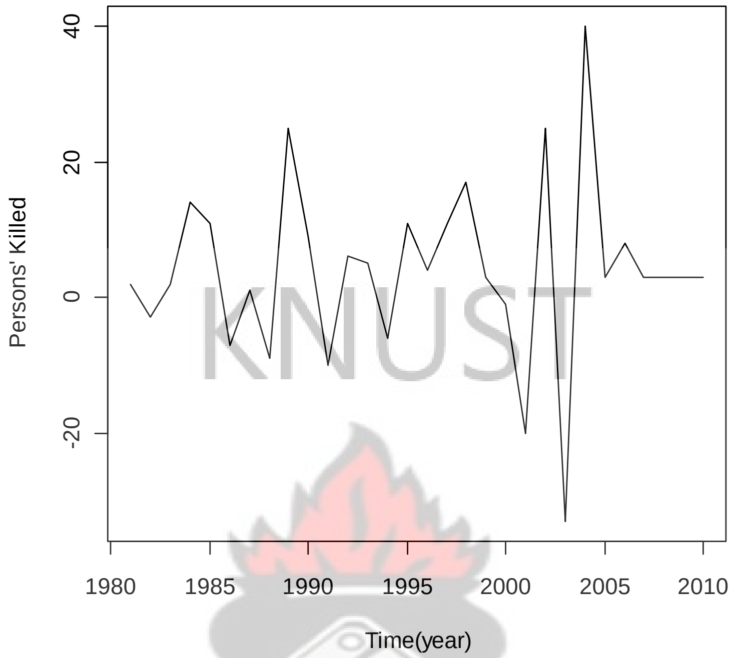

Figure 4.19: First difference of the motorway accident mortality data

First differencing method is performed to transform the motorway accident data by removing the trend component as shown in Figure 4.19 above. The observations move irregularly but revert to its mean value. The motorway accident mortality data now looks to be approximately stable.

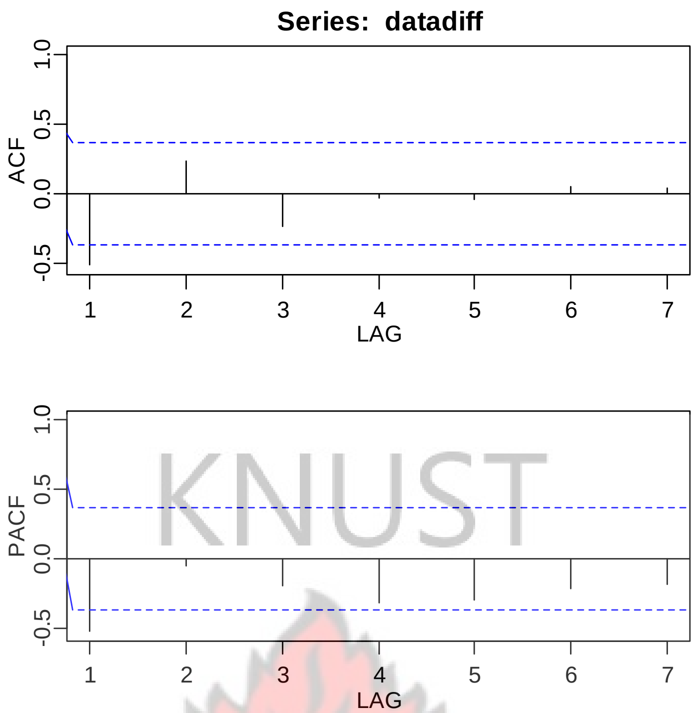

Figure 4.20: ACF and PACF of the first differencing of the accident mortality data

Comparing the autocorrelations with their error limits of the Figure 4.20 above, the only significant autocorrelation is at lag 1, indicating an MA (1) behavior. Similarly, only the lag 1 partial autocorrelations are significant. The following models are suggested;

- ARIMA(1,1,0)
- ARIMA(0,1,1)
- ARIMA(1,1,1)

To select the best model for forecasting into the future, each model is assessed based on its parameter estimates, the corresponding diagnostics of the residuals and the AIC, BIC and AIC_C values.

4.4.3 MODEL SELECTION FOR THE ACCIDENT MORTALITY DATA

4.4.3.1 PARAMETER ESTIMATES AND DIAGNOSTICS OF ARIMA (1,1,1) MODEL

Coefficients:

	ar1	constant
	-0.7251	0.1103
s.e.	0.1178	1.7035
t – value	6.1553	0.0647

sigma^2 estimated as 243.2: log likelihood = -121.18, aic = 248.36

$AIC

[1] 6.627091

$AICc

[1] 6.724527

$BIC

[1] 5.720504

The parameter based on the t – value test is statistically significant except the constant.

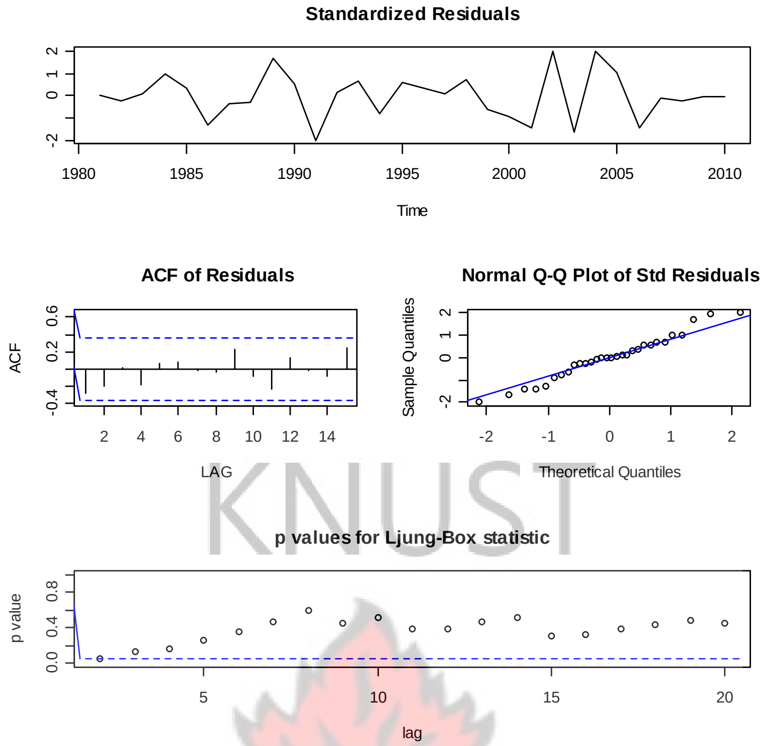

Figure 4.21: Diagnostics of ARIMA (1,1,0) model

Diagnostics of the residuals from ARIMA (1,1,0) is shown in Figure 4.21 above. The time plot of the standardized residuals shows no obvious pattern and look like an i.i.d. sequence of mean zero with some few outliers.

The plot of the ACF of the residuals of the Figure 4.21 shows no evidence of significant correlation at any positive lag.

The normal Q – Q plot of the standardized residuals of the above figure indicates that most of the residuals are located on the straight line except some few residuals deviating from normality. Therefore the normality assumption looks to be satisfied and so the residuals appear to be normally distributed.

All the p – values of the Ljung – Box statistic are not significant at any positive lag. That is no p – value is less than 0.05.

4.4.3.2 PARAMETER ESTIMATES AND DIAGNOSTICS OF ARIMA (0, 1, 1) MODEL

Coefficients:

	ma1	constant
	-1.0000	0.0529
s.e.	0.0861	0.2841
t – value	11.6144	0.1862

sigma^2 estimated as 181.4: log likelihood = -118.26, aic = 242.53

$AIC

[1] 6.334247

$AICc

[1] 6.431683

$BIC

[1] 5.42766

The parameter based on the t – value test is statistically significant. That is it is greater than 2 in absolute value. The constant term is however not significant.

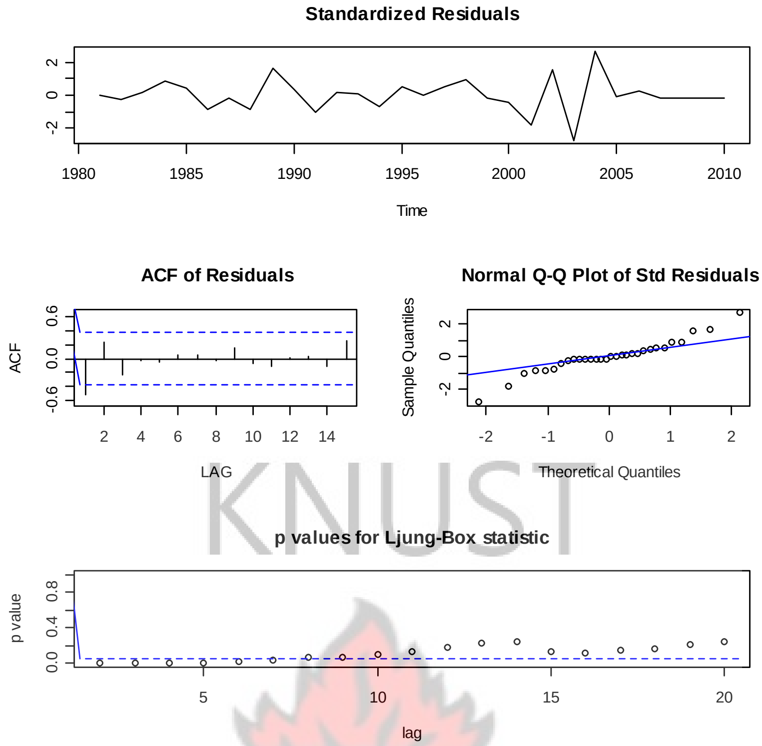

Figure 4.22: Diagnostics of ARIMA (0, 1, 1) model

Diagnostics of the residuals from ARIMA (0,1,1) is shown in Figure 4.22 above. The time plot of the standardized residuals shows no obvious pattern and looks like an i.i.d. sequence of mean zero with some few outliers.

The plot of the ACF of the residuals of the Figure 4.22 shows an evidence of significant correlation at lag one. We want non – significant autocorrelations.

The normal Q – Q plot of the standardized residuals of the above figure indicates that most of the residuals are located on the straight line except some few residuals deviating from normality. Therefore the normality assumption looks to be satisfied and so the residuals appear to be normally distributed.

Some p – values of the Ljung – Box statistic are significant at positive lags. That is some of the p – values are less than 0.05. We want non – significance p – values.

4.4.3.3 PARAMETER ESTIMATES AND DIAGNOSTICS OF ARIMA (1, 1, 1) MODEL

Coefficients:

	ar1	ma1	constant
	-0.4863	-1.0000	0.0542
s.e.	0.1560	0.0921	0.1691
t – value	3.1173	10.8578	0.3205

sigma^2 estimated as 132.8: log likelihood = -114.26, aic = 236.52

$AIC

[1] 6.088925

$AICc

[1] 6.208925

$BIC

[1] 5.229045

The parameter based on the t – value of each coefficient is statistically significant since they are each greater than 2 in absolute value except the constant.

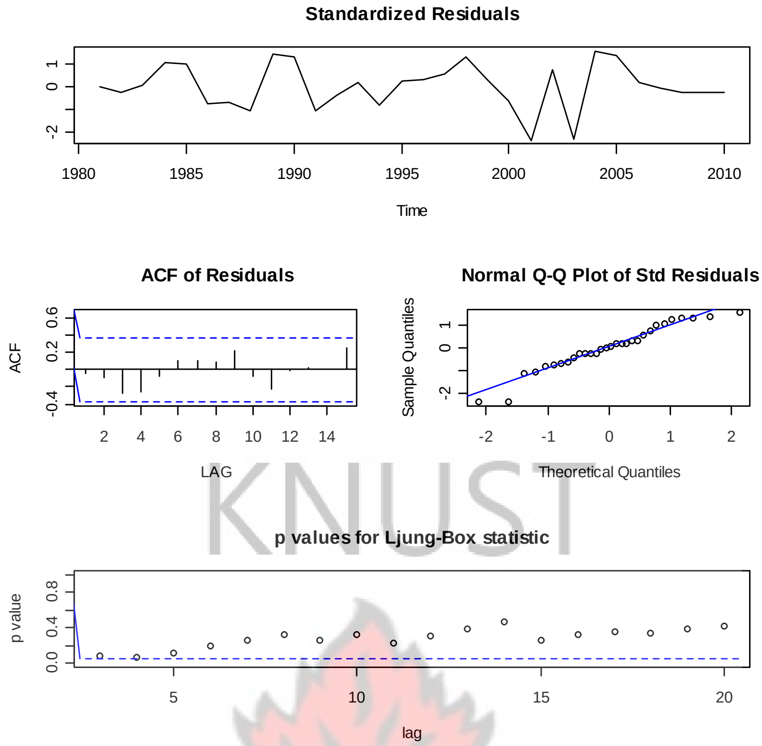

Figure 4.23: Diagnostics of ARIMA (1,1 1) model

Diagnostics of the residuals from ARIMA (1,1,1) is shown in Figure 4.23 above. The time plot of the standardized residuals shows no obvious pattern and look like an i.i.d. sequence of mean zero with some few outliers.

The plot of the ACF of the residuals of the Figure 4.23 shows no evidence of significant correlation at any positive lag.

The normal Q – Q plot of the standardized residuals of the above figure indicates that most of the residuals are located on the straight line except some few residuals deviating from normality. Therefore the normality assumption looks to be satisfied and so the residuals appear to be normally distributed.

The p – values of the Ljung – Box statistic are not significant at any positive lag. That is no p – value is less than 0.05.

Table 4.3: Parameter estimates and diagnostics of models selected for RTA deaths

Model	Test on Parameters				
	Parameter	**Estimate**	**S.E**	**t- value**	**Sig. if t≥|2|**
ARIMA(1,1,0)	ar 1	-0.7251	0.1178	6.1553	Sig
	Constant	0.1103	1.7035	0.0647	Non –sig
ARIMA(1,1,1)	ar 1	-0.4863	0.1560	3.1173	Sig
	ma 1	-1.0000	0.0921	10.8578	Sig
	Constant	0.0542	0.1691	0.3205	Non sig
ARIMA(0,1,1)	ma 1	-1.0000	0.0861	11.6144	Sig
	Constant	0.0529	0.2841	0.1862	Non-sig

DIAGNOSTICS

	ARIMA(1,1,0)	**ARIMA(1,1,1)**	**ARIMA(0,1,1)**
AIC	6.627091	6.088925	6.334247
AIC_C	6.724527	6.208925	6.431683
BIC	5.720504	5.229045	5.42766

4.4.4 SELECTION OF BEST MODEL FOR FORECASTING THE ACCIDENT DEATH

The standardized residuals plots of all the models are independently and identically distributed with mean zero and some few outliers. There is no evidence of significance in the autocorrelation functions of the residuals of all the models except ARIMA (0,1,1) which has some evidence of significance in the ACF of the residuals. The residuals appear to be normally distributed in all the models. The Ljung – Box statistics are not significant at any positive lag for all the models except ARIMA (0, 1, 1) which has some of the p – values of the Ljung – Box statistics less than 0.05.

From table 4.3 above all the parameters in the coefficients of ARIMA(1,1,0), ARIMA (0,1,1) and ARIMA (1,1,1) models are statistically significant at 5% level of significance except their constants.

The AIC, AIC_C and BIC are good for all the models but they favor ARIMA (1, 1, 1) model.

From the discussion above it is clear that ARIMA (1,1,1) model is the best model for forecasting the motorway accident mortality.

4.4.5 FITTING THE ACCIDENT DEATH/MORTALITY MODEL

ARIMA (1,1,1) model is the best model for forecasting the motorway accident death, this is a non- seasonal model with one AR term and one MA term. The model in terms of the differenced series x_t is given by:

$$x_t = \alpha + \phi_1 x_{(t-1)} + \varepsilon_t - \theta_1 \varepsilon_{(t-1)} \tag{44}$$

In terms of the observed series the model becomes

$$y_t - y_{t-1} = \alpha + \phi_1 (y_{(t-1)} - y_{(t-2)}) + \varepsilon_t - \theta_1 \varepsilon_{(t-1)} \tag{45}$$

and we may write as;

$$y_t = \alpha + (1 + \phi_1)\, y_{t-1} - \phi_1 y_{(t-2)} + \varepsilon_t - \theta_1 \varepsilon_{(t-1)} \tag{46}$$

This is called the difference equation form of the model and appears to be ARMA (2,1) process, (Cryer and Chan, 2008).

The point estimate of each parameter of ARIMA (1,1,1) from table 4.3 are as follows:

$\hat{\alpha} = 0.0542, \quad \hat{\theta}_1 = -1.0000, \hat{\phi}_1 = -0.4863$

All the estimates are significant except the constant hence it is dropped.

Hence the fitted ARIMA (1, 1, 1) model for the motorway accident death data from 1980 – 2010 is given by;

$$\hat{y}_t = (1 - 0.4863) y_{t-1} + 0.4863 y_{t-2} + \varepsilon_{t-1} \quad (47)$$

and we write

$$\hat{y}_t = 0.5137 y_{t-1} + 0.4863 y_{t-2} + \varepsilon_{t-1} \quad (48)$$

Where ε_t has an estimated variance of 132.8.

4.4.6 FORECASTING MOTORWAY ACCIDENT MORTALITY/DEATH

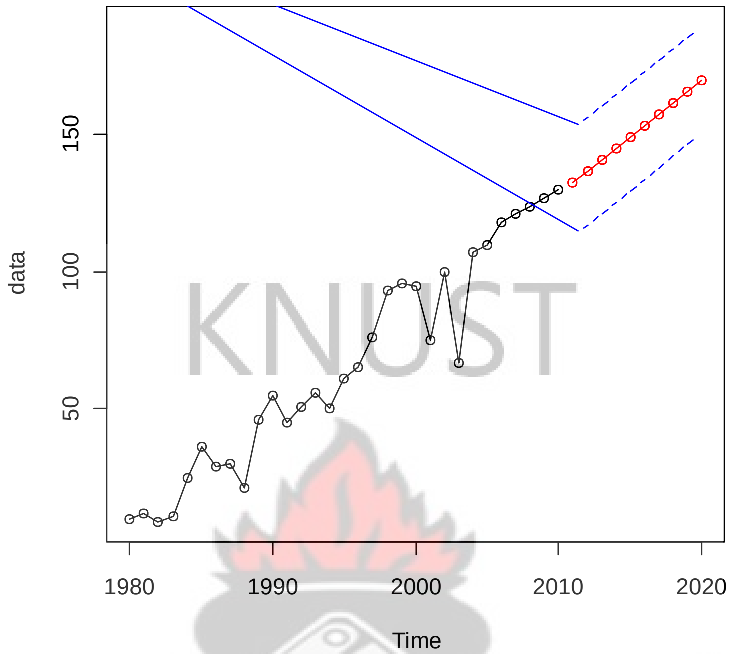

Figure 4.24: Graph of the accident mortality data, its forecasts and confidence interval

Figure 4.24 gives the visual representation of the original motorway accident mortality 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, the Accra – Tema motorway accident mortality will continue to increase in the next 10 years.

10 steps prediction into the future with the model is shown below

$pred

Time Series:

Start = 2011

End = 2020

Frequency = 1

[1] 132.6851 136.7417 140.8861 145.0361 149.1865 153.3370 157.4874 161.6378

[9] 165.7882 169.9386

$se

Time Series:

Start = 2011

End = 2020

Frequency = 1

[1] 9.707042 9.746221 9.747616 9.747700 9.747706 9.747706 9.747706 9.747706

[9] 9.747706 9.747706

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1.0 CONCLUSIONS

Road traffic accident in Ghana is increasing at an alarming rate and has raised major concerns. The NRSC recognizes the contributions road safety researches makes to development of accident reduction initiatives. It is against this background that this thesis was carried out in order to identify the patterns of RTA cases, injuries and deaths and to develop a time series ARIMA models to predict 10 years RTA cases, injuries and deaths along the Accra - Tema motorway.

Time series analysis of the data from the years 1980 – 2010 showed that patterns of RTA cases, injuries and deaths are increasing along the Accra – Tema motorway.

ARIMA models were subsequently developed for the accident data cases, injuries and deaths over the period 1980 – 2010, after identifying various tentative models.

ARIMA (0,1, 2) was identified to be suitable model for forecasting in to the future of the accident cases whist ARIMA(1,1,1) was found to be suitable model for the accident injury and deaths cases along the Accra - Tema motorway.

The study also revealed that road traffic accident cases, injuries and deaths along the motorway would continue to increase over the next 10 years.

Despite its limitation, my knowledge of the study area makes me confident that this study has generated reliable information that could be useful for RTA prevention on the Accra – Tema motorway. This information is important for raising the level of awareness among stakeholders in road safety since the problem of RTA is becoming a growing epidemic in Ghana.

It is also useful in setting priorities when planning RTA interventions.

In general, road users, MTTU, NRSC, researchers and other stakeholders are expected to benefit.

5.2.0 RECOMMENDATIONS

The ARIMA(0,1,2) model is recommended for forecasting RTA cases whilst ARIMA(1,1,1) was recommended for RTA injuries and deaths along the Accra – Tema motorway but the following precautionary measures should be taken into consideration in order to prevent the increasing(large) forecast values of these models:

- The models should not be used to forecast long time ahead (preferably a maximum of 10 years). This is because long time periods could lead to arbitrary large forecasts values.
- Law on over speeding should be strictly enforced. The posted speed limit on the Accra – Tema motorway is 80kph but most drivers tend to exceed this limit which results in RTA.
- The use of seat belt. Research has proven the effectiveness of seat belt in reducing RTA. The compulsory seat belt law for all drivers and front seat passengers should be highly enforced.
- 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, with punishment being dealt with quickly and efficiently.

- Maintenance of the motorway. It is very important that the motorway be properly maintained in terms of the use of appropriate materials for patching pot holes, provision of street lights to aid visibility in the night, installation of traffic lights at new intersections created along the road.
- Education on over taking. It is very important that, drivers must be given proper education on when to overtake and on which lanes to overtake as wrongful overtaking has been found to be one of the major causes of accident on the road.

Finally, it is also recommended that further research should be conducted to look for more appropriate models that can take care of drastic government interventions.

REFERENCES

1. Aaron, J.E. and M.K. Strasser, (1990). Driver and Traffic Education. The Macmillan Company Inc., London, New York.
2. Abdel, Hamid Abbassi (2005). Stats Methods of Predicting Fatalities of Road Traffic Accidents in Kuwait. "Egyptian Population and Family Family Review. pp. 2 – 19. Accident Analysis and Prevention, 18(2):pp.119-127.
3. Aminuddin A. (1990). Road Safety Situation in Malaysia, Road Safety Seminar, Raintree, Kuala Lumpur.
4. Andreessen, D. (1985). Linking deaths with vehicles and population. Traffic Engineering and Control, 26(11):pp.101-109.
5. AUSTROADS (1994). Road Safety Audit. AUSTROADS Publication, Sydney Australia, ISBN: 085588455X.
6. Balogun, J.A. and O. K. Abereoje (1992). Pattern of road traffic accidents cases in a Nigerian university teaching hospital between 1987 and 1990. J. Trop. Med. Hyg.,95: 23-29.
7. Cejun, Liu and Chou – Lin Chen (2004). Time Series Analysis and Forecast of Annual Crash Fatalities. "National Centre for Statistics and Analysis".
8. Cryer, Jonathan D. and Chan, Kung – Sik (2008). Time Series Analysis with applications in R, 2nd Edition. New York: Springer.
9. Emenalo, M., Pustelli, A., Ciampi and Joshi, H. P. (1987). Analysis of road traffic accidents data in Zambia. Traffic Engineering and Control 28: pp. 635-640.
10. Evans, L. (2003). Estimating Accident Fatalities in Britain. Crowthonne, U.K:

"Transport Research Laboratory Report" 445; 2003: 1-35

11. Fieldwick, R. and Brown, R. J.(1987). The effect of speed limits on road casualties.Traffic Engineering and Control. 28: pp 635-640.
12. Hijar, M., G. Carrilo, M. Flores, R. Anaya and V. López (2000). Risk factors in highway traffic accidents: A case control study. Accident Analysis and Prevention, 32:703-709.
13. Hurvich, C. M. and Tsai, C, L.(1989). Regression and Time Series Model Selection in Small Samples.Biometrika, 76,2,297 – 307.
14. Jamal, R. M. Ameen and Jamil A. Naji (2001). Causal models for road accident fatalities in Yemen. "Accident Analysis & Prevention" 33: pp 547- 561.
15. Jha, N., D.K. Srinivasa, G. Roy and S. Jagdish (2004). Epidemiological Study of road traffic accident cases: A study from South India. India J. Community Med.29: 20-24.
16. Kardara, C., and Kondakis (1997). Road Safety Situation in Greece. "Traffic Engineering and Control. 6:509-12.
17. Kim, H. S., 1990. Statistical Models for Number of Fatalities, Proc. REAAA, vol. 2, Kuala Lumpur.
18. Lascala, M., Jadaan Kis and Homel R. (1996). Epidermology of Transportation Related Injuries in a Subsahara African.179-184.
19. Mekky Ali (1985). Effects of rapid increase in motorization levels on road fatalityrates in some rich developing countries. Accident Analysis and Prevention.16 (3): pp.211-222.
20. Minter, A. L (1987). Road casualties' improvement by learning processes. Traffic Engineering and Control. 28:pp. 74-79.

21. Mock, C., J. Amegashie and K. Darteh (1999). Role of commercial drivers in motor vehicles related injuries in Ghana. Injury Prevention, 5: 268-271.

22. Nantulya, V. M. and M. R. Reich (2002). The neglected epidemic: Road traffic injuries in developing countries. BM., 324:1139-1141.

23. O'Neil, B. and D. Mohan (2002). Reducing motor vehicle crash deaths and injuries in newly motorizing countries. BMJ., 324: 1142-1145.

24. OOIDA (2003). Testimony of Julie Anna Cirillo. Land Line Magazine, Antioch, TN., USA.

25. Oskam, J., J. Kingma and H.J. Klasen (1994). The Groningen trauma study. Injury patterns in a Dutch trauma centre. Eur. J. Emergency Med. 1: 167-172.

26. Partyka, Susan C. (1984). Simple models of fatality trends using employment and population data." Accident Analysis $ Prevention". 16(3): pp. 211-222.

27. Pocock, R.J., Pustelli, A and Joshi, H. P. (1981). Some Statistical aspects of road safety. "Jounal of Royal Society" . pp.1-24.

28. Pramada Villi, P. and Sarkar, P.K. (1997). Models for road accidents in India. "Highway Research Bulletin, Indian Roads Congestion". 56: pp. 1- 11.

29. Pramada Villi, P. and Sarkar, P.K. (1993). Variation in the pattern of road accidents in different states and union territories in India. Proceedings of the third national conference on transportation systems: Analysis and Policy pp. 1X-5 to 1X-9.

30. Rehan, M. K. (1995). A Macro Analysis of Road Accident Trends in Malaysia. "Journal of Eastearn Asia Society for Transportation Studies." Vol.1 No. 3

31. Rehan, M. K. (1995). A Macro Analysis of Road Accident Trends in Malaysia,

Journal of Eastern Asia Society for Transportation Studies, Vol.1, No. 3.

32. Smeed, R.J. (1949). Some Statistical Aspects of Road Safety Research, Journal of the Royal Statistical Society, 112, pp.1034.
33. Smeed, R.J., Jacobs, G.D. and C.A. Cutting (1986). Further Research on accident rates in developing countries.
34. Stutts, J.C. and W.W. Hunter,(1999). Motor vehicle and roadway factors in pedestrians and bicyclist injuries. Accident and prevention, 31: 505-515.
35. Tsauo, P. Elvik, R., Lopez, A. (1996). Research on road accidents in Taiwan. “Traffic engineering and Control”.18: 166-170.
36. WHO (2004). World Report on Road Traffic Injury Prevention: Summary. World Health Organisation, Geneva, Switzerland
37. WHO (2008). World Report on Road Traffic Injury Prevention: Summary. World Health Organisation, Geneva, Switzerland.

GLOSSARY

Road Traffic Accident: When a vehicle collides with another vehicle, pedestrian, animal or geographical or architectural obstacle.

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

Accident death/mortality: Loss of human life immediately or after road traffic accident

Accident Cases: Number of road traffic accidents 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 sequence of observations ordered in time.

Stationary time series: A time series whose value fluctuate around a constant mean with a constant variance.

Differenced Data: if a time series is not stationary it is differenced i.e. the differenced data contains one or less point than the original data.

Autoregressive (AR) model: A model in which future values are forecast purely on the basis of past values of the time series.

Moving Average (MA) model: A model in which future values are forecast based on linear combination of past forecast errors.

DALYs: Health gap that combines in one measure the time lived with disability and time lost due to premature death.

Confidence interval: The degree of certainty of obtaining the same results if the study were to be repeated e.g. 95% certainty that the true value of a variable such as a mean, proportion, or rate is contained within a specified range; the confidence interval.

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
DALY	Disability Adjusted Life Years.
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
RTI	Road Traffic Injury
s.e.	Standard error
WHO	World Health Organisation

APPENDICE

ACCRA – TEMA MOTORWAY ACCIDENT DATA FROM 1980 – 2010

YEAR	ACCIDENT CASES	PERSONS' INJURED	MORTALITY/PERSONS' KILLED
1980	100	112	10
1981	137	151	12
1982	145	104	9
1983	167	97	11
1984	170	145	25
1985	157	126	36
1986	123	155	29
1987	218	145	30
1988	228	108	21
1989	140	137	46
1990	234	114	55
1991	260	134	45
1992	265	186	51
1993	289	197	56
1994	232	201	50
1995	216	153	61
1996	302	196	65
1997	323	167	76
1998	345	232	93
1999	370	187	96
2000	401	205	95
2001	389	215	75
2002	405	257	100
2003	420	275	67
2004	430	278	107
2005	450	332	110
2006	470	356	118
2007	483	371	121
2008	490	375	124
2009	493	382	127
2010	503	398	130

Source: National MTTU, Ghana.