

# Time Series Modelling of Road Traffic Accidents in West Arsi, Ethiopia

*The aim of this research is time series modeling of road traffic accidents in the west Arsi zone, Ethiopia, it focused on monthly traffic accidents from January 2016 to December 2020. The goal of this study was to explore the number of traffic accidents to fit a time series model for the monthly number of road traffic accidents and to forecast a tow year ahead of the number of road traffic accidents. The analysis was done by using statistical software packages using this software and knowledge of time series analysis, trend, ACF, PACF, and Box-Jenkins analysis were computed. From the trend plot, the road traffic accidents was fluctuate from month to month as well as from year to year. There was total accident fluctuation from month to month (not stationary) a total of 1010 RTA were observed. The mean of 16.83 and standard deviation of 5.764 for the total accident was served the minimum and maximum record of road traffic accidents is 4 and 33 respectively. By differencing data one time, (2, 1, 3) model was fitted for making a two-year ahead forecast. Proper model adequacy checking was done. Two-year ahead forecasts showed that October, January, and April 2021 are the months with the most prominent values. Even if the trend in total accidents was decreasing there is still a need to pay more attention in order to prevent the occurrence of accidents related to road traffic accidents.*

**Keywords:** RTA, Trend, Seasonal, Stationary, ARIMA, Forecast

## Introduction

Road traffic accidents constitute a major public health and development crisis. Over 1.35 million people die each year on the world's roads, with 30-50 more sustaining serious injuries and living with long-term adverse health consequences. Globally, road traffic crashes are a leading cause of death among young people and the main cause of death among those aged 15–44 years. However, low-income countries have fatality rates more than double those in high-income countries and there are a disproportionate number of deaths relative to these countries' level of motorization: 93% of road traffic deaths occur in low-and middle-income countries, yet these countries have just 60% of the world's vehicles [11].

Road traffic injuries place a heavy burden on national economies as well as on households. In low and middle-income countries particularly affect the economically active age group, or those set to contribute to family and the workforce in general. Many families are driven deeper into poverty by the loss of a breadwinner, or by the expenses of prolonged medical care, or the added burden of caring for a family member who is disabled as a result of road traffic injury. The economic costs also strike hard at a national level, imposing a significant burden on health, insurance, and legal systems. This is particularly true in countries struggling with other development needs, where investment in road safety is not commensurate with the scale of the problem. Data suggest that

road traffic deaths and injuries cause economic losses up to 3% of gross domestic product (GDP) globally whereas in low- and middle-income countries they are estimated to be 5% of GDP [2].

There are large disparities in road traffic death rates between regions. The risk of dying as a result of a road traffic injury is highest in the African Region (26.6 per 100,000 population), and lowest in the European Region (9.3 per 100,000) [25].

The African Region continues to have the highest road traffic death rates. The situation of road traffic accidents (RTA) is most severe in Sub-Saharan Africa where the lives of millions are lost and a significant amount of property is damaged (WHO, 2020). In Ethiopia, the situation has been worsened as the number of vehicles has increased and consequently due to increased traffic flow and conflicts between vehicles and pedestrians. Despite government efforts in the road development, road crashes remain to be one of the critical problems of the road transport sector in Ethiopia [22]. Every year many lives are lost and much property is destroyed due to road traffic accidents in the country. The country has experienced average annual road accidents of 8115 during 2000/01-2010/11 (CSA, April, [8].

In Ethiopia, the number of deaths due to traffic accidents is reported to be among the highest in the world. According to the latest WHO data published in 2018, Road Traffic Accidents Deaths in Ethiopia reached 29,386 or 4.81% of total deaths. The age adjusted Death Rate is 36.78 per 100,000 of population ranks Ethiopia 24<sup>th</sup> in the world.

Ethiopia is one among many low-income countries in Africa and, like other low-income countries it has a high rate of road traffic injury and death (Persson. A, [18]). Road accident in Ethiopia is one of the worst accident records in the world, as expressed per 10,000 vehicles. Moreover, road accidents are concentrated in Addis Ababa, which is the capital city of Ethiopia, and Oromia region accounting for 58% of all fatalities and two-third of all injuries (Getu, [11]). During 2006-2015 the average of road traffic injuries increased in Ethiopia, where Oromia regional state is the major contributor of total fatalities occurred and Addis Ababa city administration is a major contributor of serious and slight injuries as well as property damages (Amanuel, [3]). A better understanding of the trends of road traffic accident help to find the highest and the lowest point of road traffic accident in terms of month which also guide the concerned body to analysis the future accident related to the number of road traffic accident. In addition to this forecasting of road traffic accident help different organization whose work are related to accident in making their plan.

To date, there have been little types of research on road traffic accidents using pooled data in West Arsi Zone. The aim of the current study is therefore to model the trend of road traffic accidents in the west Arsi zone using time series modelling and forecast for the future.

#### *Statement of the Problem*

Road transport is one of the service sectors that contributed to human death,

injuries as well as related property damage. A road traffic crash usually occurs on vehicles travelling on public road or highway. Commonly known types of crashes include collisions when a vehicle collides with another vehicle, pedestrian, animal, road debris or other stationary obstruction, such as tree, pole, or building (Wikipedia).

In Ethiopia, road traffic accident has been one of the top causes of death. For example, in 2013, the number of people killed by road traffic accidents was equivalent to those who died due to malaria (which is the 9<sup>th</sup> cause of death) throughout the country (The Centre for Disease Control and Prevention), [21]. Road traffic death and injuries have therefore been the key public health and development challenges of the country and will continue to adversely affect the livelihood of the community and the economy of the country unless effective measures are taken to control the problem Fisseha & Sileshi, [9].

Road traffic accidents not only adversely affect the livelihood of individuals but also their family members, as it can lead households into poverty via the enduring effects of the episodes: the costs of medical care, treatment, and loss of family's income generators (Persson, [18]) Road traffic accidents have also a gigantic impact on the national economy by consuming the already inadequate resources, damaging invaluable property, and killing and disabling the productive age group of the community. In general, the severity of the problem is becoming horrific shockingly and reaching a catastrophic level (Fisseha & Sileshi, [9]).

Some research have done related to road traffic accident, but those research focused on analysing causes and determinants of RTA like “ Prevalence and Factors Associated with Road traffic Crash Among Bajaj and Motor Bicycle Drivers in Negelle Arsi Town Oromia Regional State, South East Ethiopia” (Yirgu & Wakwaya, [27]), “Statistical Modeling of numbers of human deaths per road traffic accident in the Oromia region, Ethiopia” (Aga, Woldeamanuel, & Tadesse, [2]), “Prevalence and determinants of post-traumatic stress disorder among road traffic accident survivors” (Bedaso, et al., [6]). However, this research focused on modelling and forecasting future road traffic accident in addition to the existing prevalence and what other literatures considered in west Arsi zone.

#### *Research Question*

- 1) What will be the number of traffic accidents in the next 2 years in our Zone?
- 2) In which month does a large number the traffic accident will occur in next two years in the west Arsi zone?
- 3) What is the pattern or trend of road traffic accident in the years of 2016 – 2020 in west Arsi zone looks like?
- 4) What type of trend will follow the road traffic accident in west Arsi zone?

## *Objective of the Study*

### General Objective

The main objective of this research work is to model road traffic accident by using time series techniques.

### Specific Objective

The specific objectives of this study include:

Assessing the trend of road traffic accident in West Arsi Zone, Ethiopia.

Forecasting future Number of road traffic accident in West Arsi zone, Ethiopia.

Studying problems and obstacles related to reducing traffic accidents.

## *Significance of the Study*

This study is purposeful to know how much a road traffic accident was serious. It is important in reducing the accident, by giving recommendation to the concerned bodies to solve the problem. The study is also important to make awareness for the society by fitting an appropriate model. The result could serve as reference material for other researchers in related topics or problems. In addition, it is important to forecast the future trends.

## *Scope of the Study*

This study was focuses on road traffic accident in west Arsi zone to show over all problems of road traffic accident related to time in which the accident is caused. The study was included that the number of road traffic accidents recorded in west Arsi zone, Ethiopia.

## **Literature Review**

### *Theoretical Literature Review*

### Definitions and Concepts

Road traffic accident (RTA) is defined as a collision or incident involving at least one road vehicle in motion. It can be said to be an unplanned occurrence of auto crash that may result in injuries, loss of lives and properties. RTA can be a collision among vehicles, between vehicles and pedestrians, between vehicles and animals, or between vehicles and geographical or architectural obstacles (Goswami A, Sonowal R, [12]).

Road traffic injuries cause considerable economic losses to individuals, their families, and to nations as a whole. These losses arise from the cost of treatment as well as lost productivity for those killed or disabled by their injuries, and for family members who need to take time off work or school to care for the injured. Road traffic crashes cost most countries 3% of their gross domestic

product (WHO, [23-26]).

### Measures of Car Accidents

Addis Ababa Police Commission report of the road crashes were summarized into three levels of injuries (fatal injury, serious injury and minor injury) and property damage.

According to the definitions given by the police commission:

#### **(i) Fatal Injury**

It is any confirmed road crash related death report (either on the spot or any time after the sustained injury) by Minilik II Referral Hospital.

#### **(ii) Serious Injury**

Defined as when one or more road user (s) (pedestrians, passengers and drivers) suffered severe cuts, bleeding, breaks, and other damages which required them to receive medical treatment as an “in-patient” in hospital.

#### **(iii) Minor Injury**

On the other hand, is defined as when one or more road users sustain (s) minor/slight cuts, scratches, and other minor damage which require the road user (s) to be treated as an outpatient without requiring hospitalization, and not resulting in fatality/death.

#### **(iv) Property Damage**

According to the commission, traffic police officers estimate the property damage in terms of financial costs. (Abdi, [8])

There is a relationship between season; weather and time factor in road traffic accident occurrence. Fatal accidents mostly occur during winter season. A study done by Kong et al has revealed that most of the accidents occur during the night, weekends and during months of October to December (Kong, et al., [6]).

The human factor or error contributes to the majority of road traffic accidents. In Kenya reported that human factors were responsible for 85% of all causes (Odero, [17]).

A study done by (Barreto et al. [5]) observed that exposure to high intensity noise at work place tends to be associated with occupationally acquired hearing deficits. These deficits increase the risk of motor vehicle injury to pedestrian workers (Barreto, Swerdlow, Smith, & Higgins, [5]).

According to the study by a member of the Swedish Medical University of Lund, in adequate communication to immediately inform officials and hospital emergency services about traffic accidents in rural areas is a problem, implying that many accidents and the number of victims cannot be registered. Poor emergency medical services and the absence of compulsory liability insurance laws are among reasons contributing to the high fatality rates. In the urban areas, although traffic police and hospitals are available, standers who have neither the necessary skills nor equipment in pre-hospital care usually evaluate accident victims. (*Google/literatures on car accident*)

Over 3000 Kenyans are killed on our roads every year, most of them between the ages of 15 and 44 years. The cost to our economy from these accidents is in excess of US\$ 50 million exclusive of the actual loss of life. The Kenyan government appreciates that road traffic injuries are a major public

health problem amenable to prevention (Odero, [17]).

In 2003, the newly formed Government of the National Alliance Rainbow Coalition took up the road safety challenge. It is focusing on specific measures to curtail the prevalent disregard of traffic regulations and mandating speed limiters in public service vehicles. Along with the above measures the Government has also launched a six-month Road Safety Campaign and declared war on corruption, which contributes directly and indirectly to the country's unacceptably high levels of road traffic accidents. Driving is a complex system in which a large number of variables are interacting with each other but also with varying degree of dependence. Accident may be due to judgement errors, ignorance, incompetence, rule violation, lapses or carelessness, all of which are human errors (Lemming 1969). Behavior is an intrinsic part of people human behavior approaches A good control of the vehicles on the road depends very much on the behaviours (which is very complex) and skill on the driver (Rundmo & Iversen, [20]).

(Jorgensen and Abane, [14]). Notes that, concerning road traffic behaviour, one can distinguish between driving skills (Knowledge and training) and driving style which reflects attitudes and traffic risk perception (Jørgensen & Abane, [14]).

Training of drivers increases their driver's skills. Study done by Asogwa in Nigeria has revealed that a sizeable proportion of drivers who possesses driving licenses never showed up in any driving school or went through a driving test but simply bought their licenses. Untrained drivers, not unexpectedly, often result in high accident rates (Asongwa, [4]).

(Leon, [16]) have also shown through their various studies that young drivers are more frequent involved in accidents caused by inappropriate speed and loss of control of the vehicle compared to other age group of drivers With regards to gender, it appears that males are more involved in motor accidents than females.

(Rivara & Barber, [19]) have also reported that among the drivers of motor vehicles that struck victims, 69% of them were males and 31% females, controlled for gender exposure level.

## Materials and Methods

### *Study Area and Study Design*

#### Study Area

**West Arsi** (Oromo: *Arsii Lixaa/Dhihaa*) is one of the zones of the Oromia Region in Ethiopia. This zone is named after a subgroup of the Oromo, who inhabit it. West Arsi was formed of Districts that included Arsi, Bale and East Shewa zones. (Wikipedia)

Based on the 2014 Census conducted by the Central Statistical Agency of Ethiopia (CSA), this Zone has a total population of 1,964,038, of whom 973,743 are men and 990,295 women.

1 **Figure 1. Map of West Arsi Zone, Ethiopia**



2  
3  
4 272,084 or 13.85% of population are urban inhabitants. 387,143 households  
5 were counted in this Zone, which results in an average of 5.01 persons to a  
6 household, and 369,533 housing units. The two largest ethnic groups reported in  
7 West Arsi were the Oromo (88.52%) and the Amhara (3.98%); all other ethnic  
8 groups made up 7.5% of the population. Oromo was spoken as a first language  
9 by 87.34% and 6.46% spoke Amharic; the remaining 6.2% spoke all other  
10 primary languages reported. The majority of the inhabitants were Muslim, with  
11 80.34% of the population having reported they practiced that belief, while 11.04%  
12 of the population professed Ethiopian Orthodox Christianity, Catholic  
13 Christianity 4.5% and 7.02% of the population professed Protestantism.  
14 (Wikipedia)

#### 15 Study Design

16 The study used a retrospective study design. All records from September  
17 2016 up to August 2020 were considered in the study. Secondary data from the  
18 West Arsi Zone traffic transport authority records were taken.  
19

#### 20 Methods of Data Collection

21 The study has been conducted using secondary data collection method. It  
22 has been taken from record of west Arsi zone transport authority. The data  
23 contains Monthly recorded road traffic accident of west Arsi zone in the period  
24 of 2009-2013.  
25

#### 26 Variables in the Study

##### 27 Dependent Variable

28 It is the variable that affected by the independent variables. In this case, the  
29 number of road traffic accident recorded monthly was the dependent variable.  
30

##### 31 Independent Variable

32 Is a variable that affects the dependent variable. In this study time is  
33  
34  
35

considered, as an independent variable e.

### *Methods of Data Analysis*

The method of data analysis is determined by the nature of data you are going to the use. Accordingly, the appropriate method of this study was time series model.

### Definition of Time Series

A time series is a series or sequence of data points measured typically at successive times. These data points are commonly equally spaced in time (Chatfield, [7]).

A time series containing records of a single variable is termed as univariate whereas a time series containing records of more than one variable is termed as multivariate.

A time series can be discrete or continuous. In a continuous time series observations are measured at every instance of time (continuously through time), whereas a discrete time series contains observations measured at discrete points of time.

*Time series analysis used for:*

- 1) The analysis of time series is of great significance not only to the economists and business man but also to the scientists, gastronomists, geologists, sociologists, biologists, research worker etc. for the reason below:-
- 2) It helps in understanding past behavior: -By observing data over a period of time one can easily understand what changes have taken place in the past. Such analysis will be extremely helpful in predicting the future behavior.
- 3) It helps in planning future operations: -The major use of time series analysis is in the theory of forecasting. The analysis of the past behavior enables to forecast the future. Time series forecasts are useful in planning, allocating budgets in different sectors of economy.
- 4) It facilitates comparison: -Different time series are often compared and important conclusions are drawn there from.

### Components of Time Series

#### *Trend Component*

Trend is the underlying long-term movement over time in the value of the data recorded. This shifting or trend is usually the result of long-term factors such as change in the population, demographic characteristics of the population, technology and consumer preferences. There are no proven automatic techniques to identify trend components in time series data.

The form of the trend pattern may be linear or non-linear. Linear trend pattern is not only the simplest, but also the most commonly encountered trend



1 pattern.

2

### 3 *Seasonal Variation*

4 Seasonal variation is a *regular*, relatively short-term repetitive up-and-  
5 down fluctuations of the values of the data recorded. The short-term fluctuations  
6 in recorded values are due to different circumstances, which affect results at  
7 different times of the year, on different days of the week, at different times of the  
8 day, etc.

9

### 10 *Cyclical Variation*

11 Cyclical variations are medium term changes in the series, caused by  
12 circumstances, which repeat in cycles. The duration of a cycle extends over long  
13 period of times, usually two or more years. Most of the economic and financial  
14 time series show some kind of cyclical variation. For example, a business cycle  
15 consists of four phases, prosperity, decline, depression and recovery. Cyclical  
16 variations are longer term than seasonal variations.

17

### 18 *Irregular (Error) Component*

19 Irregular variations in a time series are caused by unpredictable influences,  
20 which are not regular and also do not repeat in a particular pattern. These  
21 variations are caused by incidences such as war, strike, earthquake, flood,  
22 revolution, etc. There is no defined statistical technique for measuring random  
23 fluctuations in a time series.

24

### 25 *Model of Time Series*

26

#### 27 A. Additive Model

28

29 If the model in which the four components of time series are given in the  
30 form of summation. We compute symbolically

$$Y_t = T_t + S_t + C_t + I_t$$

31 Where,  $Y_t$  = Observations at time  $t$

32  $T_t$  = Trend component at time  $t$

33  $S_t$  = Seasonal component time  $t$

34  $C_t$  = Cyclical component at time  $t$

35  $I_t$  = Irregular component at time  $t$

36

37 It is an appropriate model if we assume that all components are independent  
38 of one another and the magnitude of the seasonal fluctuations does not vary with  
39 the level of the series.

40

#### 41 B. Multiplicative model

42

43 Multiplicative we use this model when the size of the seasonal pattern in  
44 data depends on the level of the data and our assumption increase so does the  
45 seasonal pattern. Most time series exhibit such a pattern,

In this model, the time series data given in the form of the product of four

1 components.

$$Y_t = T_t * S_t * C_t * I_t$$

2

3 Where  $T_t$ ,  $S_t$ ,  $C_t$  and  $I_t$  are trend, seasonal, cyclical and irregular component.  
4  $Y_t$  is the time series value at time  $t$ .

5

## 6 Data Analysis in Time Series

7 To do analysis of any time series data, first we should have to check the  
8 stationary of the data. Stationary means there is no growth (decline in the data) to  
9 form forecasting most of the probability theory of time series is concerned and  
10 for these reason time series analyses often lagged one to turn a nonstationary  
11 series into stationary. To say a time series data  $Y_t$ - stationary

12

- 13 1) Means of  $Y_t$  and Variance of  $Y_t$  is constant for all time periods
- 14 2) Covariance:  $\gamma_k = E [(Y_t - \mu)(Y_{t-k} - \mu)]$  Covariance between  $Y_t$  and  
15  $Y_{t+i}$  is constant for time series and for fixed  $i$  where  $i=1, 2, \dots, k$ .

16

## 17 *Examining Stationary of Time Series Data*

18

### 19 *Time series plot*

20 Time series plot is the most frequently used from of graphic design to show  
21 no obvious change in the variance over time and then we say the series is  
22 constant variance and the mean or no evidence of change in the mean over time.  
23 It is also used to examining stationary of time series data.

24

- 25 1) If a time series is plotted and if there is no evidence of a change in a  
26 mean over time then we say the series is stationary on the mean.
- 27 2) If the plotted series shows no obvious change in the variance time then  
28 we say the series is constant variance.

29

### 30 *Autocorrelation function:*

31 Let  $\{ Y_t \}$  be a stochastic process with mean  $\mu$  and variance, auto covariance  
32 function, then we define the autocorrelation function (acf) denoted by as  
33  $\rho_k = \frac{\gamma_k}{\gamma_0}$ .

34 The autocorrelation function is the preferred measure; this is because it is a  
35 unit less measure.

36 The autocorrelation function has two interesting properties. These are:

37

- 38 1. The acf is an even function of the lag in that,  $\rho_k = \rho_{-k}$
- 39 2.  $|\rho_k| \leq 1$

40

### 41 *Partial autocorrelation function:*

42 This is the correlation between  $Y_t$  and  $Y_{t-k}$  after removing the effect of the  
43 intervening variables  $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$  for any stationary series at lag  $k$  and  
44 is denoted by  $\phi_{kk}$ . Thus

45

$$\phi_{kk} = \text{Corr}(Y_t, Y_{t-k} | Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1})$$

The partial autocorrelation function is used to determine the autoregressive part for an ARMA model.

The autocorrelation of stationary data drops to zero relatively quickly, while for nonstationary data they are significantly different from zero for several and PACF would have a long spike close to 1 at lag 1.

### *Differencing*

Differencing is the process of changing a nonstationary time series into a stationary time series. Regularly differencing is taking successive differences of the data. The method of taking first difference of data is simply to subtract the values of two adjacent observations on time series. If the original data has  $n$  observations ( $Y_1, Y_2, \dots, Y_n$ ), the first differenced data will be  $n-1$  observations ( $X_2, X_3, \dots, X_n$ ).

Where  $X_2 = Y_2 - Y_1, X_3 = Y_3 - Y_2, \dots, X_n = Y_n - Y_{n-1}$

Generally  $X_t = \Delta Y_t = Y_t - Y_{t-1}$

$$\begin{aligned} Z_t = \Delta X_t &= \Delta^2 Y_t = \Delta(\Delta Y_t) = \Delta(Y_t - Y_{t-1}) = \Delta Y_t - \Delta Y_{t-1} \\ &= (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) \end{aligned}$$

Download package called **urca** and **tseries** are used to check or detect stationarity on R software.

### *Augmented Dickey-Fuller (ADF) Test*

It is a common statistical test used to test whether a given time series is stationary or not. It is one of the most commonly used statistical test when it comes to analysing the stationarity of a series.

Dickey-fuller test using R software to check the stationarity of data. The procedure is as follow.

$H_0$  = The data is Stationary

$H_1$  = Not  $H_0$

If the Augmented p-value is greater than alpha value reject  $H_0$  and take the remedial action.

### *Test of Randomness*

It is test to determine whether the given data are random or not. There are different tests of randomness such as rank test, phase length test and difference sign test here in our case we use Bartels test and defined as follow.

### *Bartels test for randomness*

Bartels [5] has proposed a rank version of von Neumann's ratio test one of non-parametric test which is used for test randomness of a data but we can simply use R studio by the command: bartels.test(DY) # where DY is the data we want to check the randomness.

### Turning point test

If  $y_1, \dots, y_n$  is a sequence of observations, we say that there is a turning point at time  $i$ ,  $1 < i < n$  if  $y_{i-1} < y_i$  and  $y_i > y_{i+1}$  or if  $y_{i-1} > y_i$  and  $y_i < y_{i+1}$ . If  $T$  is the number of turning points of an iid sequence of length  $n$ , then, since the probability of a turning point at time  $i$  is  $\frac{2}{3}$ , the expected value of

$T$  is

Hypothesis Test

$H_0: Y_t, t = 1, 2 \dots n$  Observations are random.

$H_1$ : observations are not random (time independent)

Where  $Y_t$ : The observation at time  $t$

Let  $T$  is the number of turning point for the set of observation

$$\mu_T = E(T) = \frac{2(n-2)}{3}$$

$$\sigma_T^2 = Var(T) = \frac{(16n-29)}{90} \quad T \text{ is approximately } N(\mu_T, \sigma_T^2)$$

$$\text{Test statistic, } Z_T = \frac{T - \mu_T}{\sigma_T}$$

$$\text{Critical value } Z_{\frac{\alpha}{2}}$$

Decision rule, Reject  $H_0$  if  $|Z_{cal}| > \frac{\alpha}{2}$  That means the time series is not independently identically distributed.

### Trend Analysis

Trend is general tendency to increase or decrease during a long period. There are several methods to detect trend. From these methods, graphical method, semi-average method, moving average method, and least square method are used for determine trend analysis of road traffic accident. In order to measure trend, we will try to eliminate seasonal, cyclical, and irregular components from time series data. Trend analysis fits the general model to time series data and to estimate the constant mean model and also provide forecasts among the linear or nonlinear trend model. The above listed methods are suitable for manual calculation. We can also use non- parametric statistical tests.

### Mann-Kendall trend test

$H_0$  = the series not follow a trend

$H_1$  = Not  $H_0$  (follow a trend)

For this test we use mann-kendall trend test (source: help package Trend)

In test statistics if we have insignificant p-value, we can reject the null hypothesis and conclude that is not follow trend.

### Remedial for trend analysis

Differencing is used for remedial action of trend analyses and stationarity.

### Seasonality

There are also several tests for seasonality such as the Friedman test and the Kruskal-Wallis test. Both test are available in R and apply to stable additive seasonality. The command for the Kruskal-Wallis test is `kruskal.test()` from the preinstalled stats package. The command for the Friedman test is `friedman.test()` which is also in the stats package.

### *Friedman test*

The Friedman test is a test which is used to detect differences in treatment across multiple test attempts. It can detect differences in the mean between  $\geq 2$  samples. Here is an explanation why the Friedman test is useful for seasonality: Stable seasonality test (also called an F-test, Friedman test) is a test for the presence of seasonality based on a one-way analysis of variance on the SI ratios. Thus the test is performed on the determined time series adjusted for prior factors. Generally, the test compares the periods (months or quarters) variance with the residual variance. Under the assumption that the first one is caused by seasonal factors while the second one derives from irregular movements the test checks if the first type of variations are repetitive and regular enough to be reliably identified as the seasonal movements.

### Kruskal-Wallis test

The Kruskal-Wallis test is similar to the Friedman test. It also tests whether  $\geq 2$  samples have different means. The Difference to the Friedman test is that the Kruskal-Wallis test is based on an Analysis of Variance (ANOVA) between the samples.

Kruskal-Wallis test is a non-parametric test used for comparing samples from two or more groups. This test does not make assumptions about normality. However, it assumes that the observations in each group come from populations with the same shape of distribution. The null hypothesis states that all months (or quarters, respectively) have the same mean.

Source (google /Stack exchange/detecting seasonality)

## Box-Jenkins Approach

Statisticians George Box and Jenkins developed a practical approach to build ARIMA model, precisely concerning about how to select an appropriate model that can produce accurate forecast based on a description of historical pattern in the data and how to determine the optimal model orders which best fit to a given time series and also satisfying the parsimony principle. The Box-Jenkins methodology does not assume any particular pattern in the historical data of the series to be forecasted. Rather, it uses a three-step iterative approach of model identification, parameter estimation and diagnostic checking to determine the best parsimonious model from a general class of ARIMA model. This three-step process is repeated several times until a satisfactory model is finally selected.

The basis of Box-Jenkins approach to modelling time series consists of

three faces

- 1) Model selection/identification.
- 2) Parameter estimation.
- 3) Model checking.

### *Autoregressive model*

A time series is said to be autoregressive process if the current time series is a linear aggregate of a finite number of previous value plus random shock, that is,

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t$$

Where  $Y_t$  - Current Value

$Y_{t-p}$  - a value at lag p

$a_t$  White noise error

$\phi_1, \phi_2 \dots \phi_p$  Parameter of the model which is estimated from the data.

### *Moving Average model*

Time series is said to be in moving average (MA) process if the current time series is linear combination of current and finite number of previous shock. The qth order moving average process can be expressed as

$$Y_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q}$$

Where

$a_t$  White noise error

$a_{t-q}$  White noise error at lag of q

$\theta_1, \theta_2, \dots, \theta_q$  Moving average parameter which describe the effect of the past error on  $Y_t$ .

### *ARMA/ARIMA Models*

Auto Regressive Moving Average (ARMA): process that are formed as combination of Auto Regressive (AR) and Moving Average (MA) process. ARMA process of order (pq) has the form:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q}$$

Auto Regressive Integrated Moving Average (ARIMA): it is general model capable of representing wide range of nonstationary series that can be reduced to a stationary series with a degree of differencing and having p auto regressive q moving average. ARIMA process of order (p, d, q).

Where: p- is the number of auto regressive terms

d- Is the number of non –stationary differencing process.

q –is the number of moving average terms (lagged errors in the equation)

1) *Autoregressive (p)*. The number of autoregressive orders in the model.

Autoregressive orders specify which previous values from the series are used to predict current values.

2) *Difference (d)*. Specifies the order of differencing applied to the series before estimating models. Differencing is necessary when trends are present (series with trends are typically nonstationary and ARIMA

modelling assumes stationary) and is used to remove their effect. The order of differencing corresponds to the degree of series trend--first-order differencing accounts for linear trends, second-order differencing accounts for quadratic trend, and so on.

- 3) Over differencing introduces unnecessary correlations into a series and will complicate the modelling process source (Jonathan D. Cryer, [13]).
- 4) *Moving Average (q)*. The number of moving average orders in the model. Moving average orders specify how deviations from the series mean for previous values are used to predict current values.

#### *Method of ARIMA Model Identification*

In order to identify ARIMA model the study requires Box –Jenkins modelling method rather than applying stationary models directly because the data used in this study is non-stationary or not. In Box-Jenkins modelling, the general approach is to difference an observed time series data until it appears to come from a stationary process. The basic steps in the Box –Jenkins procedures to identify ARIMA models are: studying auto correlation coefficient (ACC), studying partial auto correlation (PAC) and studying Akaike Information Criterion (AIC).

- 1) Significant pacf is responsible for order of autoregressive model: source: (Jonathan D. Cryer, 2008) page 113
- 2) The choice of the model order  $q$  then relies on the analysis of the sample ACF. (Jonathan D. Cryer, [13]).

Generally, for model selection we focus on the following table

**Table1.** Behavior of ACF and PACF for ARMA ( $p, q$ )

Model	ACF	PACF
AR( $p$ )	Exponential decay	Cut off after lag $p$
MA( $q$ )	Cut off after lag $q$	Exponential decay
ARMA( $p, q$ )	Exponential decay	Exponential decay

#### *AIC (Akaike information criterion)*

The Akaike information criterion (AIC) is a measure of the relative quality of a statistical model for a given set of data. As such, AIC provides a means for model selection. AIC deals with the trade-off between the goodness of fit of the model and the complexity of the model. It is founded on information theory: it offers a relative estimate of the information lost when a given model is used to represent the process that generates the data. For any statistical model, the AIC value is

$$AIC = 2K - 2\ln(L)$$

Where  $k$  is the number of parameters in the model, and  $L$  is the maximized value of the likelihood function for the model. Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value.

#### *BIC (Bayesian's Information Criteria)*

The Bayesian information criterion (BIC), proposed by Schwarz and hence also referred to as the Schwarz information criterion and Schwarz Bayesian information criterion, is another model selection criterion based on information theory but set within a Bayesian context. The difference between the BIC and the AIC is the greater penalty imposed for the number of parameters by the former than the latter. Burnham and Anderson provide theoretical arguments in favor of the AIC, particularly the AICc over the BIC.<sup>10</sup> Moreover, in the case of multivariate regression analysis, Yang explains why AIC is better than BIC in model selection.

The BIC is computed as follows:

$$BIC = -2 \log(L) + k \log(n)$$

Where the terms are the same as the terms described in our description of the AIC. The best model is the one that provides the minimum BIC.

### *Residual Analysis*

Before a model can be used for inference the assumptions of the model should be assessed using residuals. Recall from regression analysis, residuals are given by:

$$Residual = Actual\ Value - Predicted\ Value$$

### *Test of Independence*

A test of independence can be performed by:

- 1) Examining ACF: compute the sample ACF of the residual. Residuals are independent if they do not form any pattern and are statistically insignificant, that is, they are within  $Z_{\alpha/2}$  standard deviation.
- 2) Using runs test: Randomness (independence) in the residuals can be tested in several ways. The runs test examines the residuals in sequence to look for patterns. For example, patterns will occur if the residuals are correlated.

### *Test of Normality*

Test of normality can be performed by:

- 1) Constructing Histogram: Gross normality can be assessed by plotting histogram of the residuals. Histogram of normally distributed residuals should approximately be symmetric and bell shaped.
- 2) Plot Residuals against Normal Score: Normality of residuals can be checked more carefully by plotting the normal scores.

### *Test of Constant Variance*

Test of constant variance can be inspected by plotting the residuals over time. If the model is adequate, we expect the plot to suggest a rectangular scatter around a zero horizontal level with no trend whatsoever.



## Forecasting

Forecasting may represent a prediction as to what might happen to one such as road traffic accident next year or in five-year time or it may be a prediction as to the future of a much more complex entity such as the economy.

Forecasting refers to the using of knowledge use have at one moment of time to estimate what will happen at another moment of time. The forecasting problem is created by the interval of time between the moments.

Road traffic accident forecasting refers to the statistical analysis of the past and current movement in a given time series to as to obtain clues about the future pattern of the movements.

### Forecast Accuracy

#### (i). *Measure of Accuracy*

A measure of accuracy refers to goodness of fit. The three measure of accuracy are MAPE, MAD and MSD for each the forecasting and smoothing models. From those three measure the smaller value; the model is better fit. The concept associated with measuring forecast accuracy is forecast error defined as

$$\text{Forecast Error} = \text{Actual Value} - \text{Forecast}$$

#### *Mean absolute percentage error (MAPE)*

1) It measures the accuracy of fitted time series model.

2) It expresses accuracy as a percentage.

$$\text{MAPE} = \sum_{t=0}^n \frac{\left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|}{n} * 100$$

Where:  $Y_t$  = the actual value

$\hat{Y}_t$  = Forecasted value

$n$  = the number of observations.

#### *Mean absolute deviation (MAD)*

It measures the accuracy of fitted time series data value. It expresses the accuracy in some unit as the data which helps to conceptualize the amount of the error.

$$\text{MAD} = \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{n}$$

#### *Mean square deviations (MSD)*

It is very similar to MSE (mean square error) which is a commonly used to measure the accuracy of fitted time series values. It always compared using some denominator 'n' regardless of the models. So we can compare MSD values across models.

$$\text{MSD} = \sum_{t=1}^n \frac{(Y_t - \hat{Y}_t)^2}{n}$$

## *Ethical Clearance*

Ethical clearance will be obtained from the University.

## **Results and Discussions**

### *Descriptive Statistics*

In this chapter, results obtained by using different methods that described in chapter three and the results of the RTA model specifications used for forecasting will be presented. This study is based on RTA collected monthly from July 2016 up to June 2020. The total number of observations is 60.

**Table 2.** *Descriptive Statistics of Total Accident*

<b>Descriptive Statistics</b>						
	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Sum</b>	<b>Mean</b>	<b>Std. Deviation</b>
Total Accident	60	4	33	1010	16.83	5.764
Valid N (listwise)	60					

The total observation which is collected from Transport Authority of West Arsi Zone was five years' data and the summary of this monthly data (like mean, Sum and standard deviation) are given above in (Table 2). the total sum RTA recorded was 1010 this means from 2016 up to 2020, 1010 accidents was occurred in west Arsi zone, the Minimum and Maximum accidents occurred in past 60 months are 4 and 33 respectively which is recorded in Transport Authority of West Arsi Zone. A Mean of total accident in past 60 months are 16.83 from this we understand that for last 5 years in average approximately 17 accidents was occurred monthly and standard deviation 5.764 tell us the varieties of accident in months of last 5 years is approximately 6.

As we try to describe the measures of RTA accident in literature review the RTA has four measures those are Fatal Injury, Series Injury, Minor Injury and Property Damage. In this study we include only the descriptive part for these measures, because of the short time we have. We focus only on the total accident for the inferential parts.

**Table 3.** *Descriptive Statistics for Fatal, Series and Minor Injury*

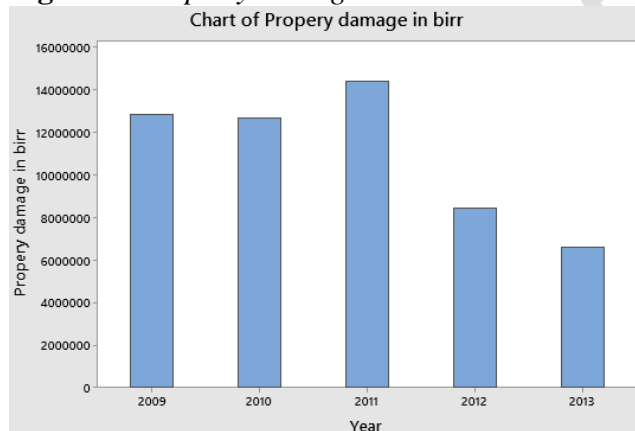
<b>Descriptive Statistics for Fatal, Series and Minor Injury</b>						
	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Sum</b>	<b>Mean</b>	<b>Std. Deviation</b>
Fatal Injury	60	0	16	509	8.48	3.539
Series Injury	60	0	12	250	4.17	2.865
Minor Injury	60	0	20	251	4.18	3.387
Valid N (listwise)	60					

The summary of monthly data for measures of RTA (like mean, Sum and standard deviation) are given above in (Table 2) and the total sum of Fatal Injury, Serious Injury and Minor Injury recorded was 509, 250 and 251 respectively, the Minimum for all measures is 0 and Maximum Fatal Injury, Serious Injury and Minor Injury 16, 12 and 20 respectively. A Mean of Fatal Injury, Serious Injury and Minor Injury in a year is 8.48, 4.17 and 4.18 respectively and a standard deviation of Fatal Injury, Serious Injury and Minor Injury is 3.539, 2.865 and 3.387 respectively.

**Table 4. Property Damage**

Year	Property damage in birr
2009	12,825,600
2010	12,653,065
2011	14,385,480
2012	8,411,700
2013	6,587,250
Total	54,863,095

**Figure 2. Property damage in birr**



As we can see from above (Table 3) and (Figure 1)\_the maximum and minimum property damage in birr is 14,385,480 in 2011 and 6,587,250 in 2013 respectively and generally our county Ethiopia losses 54,863,095 birr in last 5 years only in west Arsi zone.

#### Test of Randomness

To test the randomness of the data we use bartels test

$H_0$ : The series is random

$H_0$ : The series is not random

The R output of Bartels test of the differenced data

Bartels. Test (DifferencedTotalAccident)

Bartels's test for randomness

Data: DifferencedTotalAccident

RVN = 2.7573, p-value = 0.9987

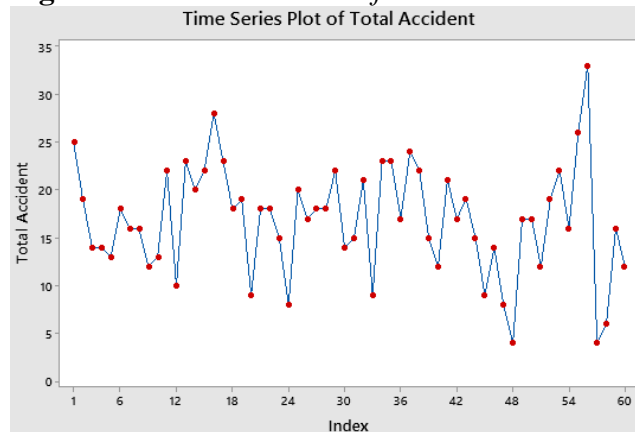
Alternative hypothesis: The series is significantly different from randomness

*Discussion:* - Since the  $p$ -value is greater than the alpha value so we fail to reject the null hypothesis that says the data is random. So, our data (Differenced Total Accident) is random.

*Stationary*

### Time Series Plot

**Figure 3.** Time Series Plot of Total Accident



The time series plot of our data below indicates some decrement of the total accident over time. It is meaning the mean of the data over time is not constant and so that it is not stationary. To make it stationary we employed some technical remedial action, usually differencing is common in time series analysis.

### Augmented Dickey-Fuller (ADF)

We have seen that the original data is not stationary by time series plot. The Augmented Dickey-Fuller test is one of the formal tests to test the stationarity of time series data.

Hypothesis for ADF test

$H_0$  = The data is not stationary

$H_1$  = The data is stationary

The R output of Augmented Dickey-Fuller test of the original data

adf. Test (TotalAccident, k = 2)

Augmented Dickey-Fuller Test

Data: TotalAccident

Dickey-Fuller = -3.1678, Lag order = 2, p-value = 0.1028

Alternative hypothesis: stationary as we see from the R output for augmented dickey-fuller test for total accident at  $\alpha = 5\%$  and lag order = 2 the  $p$  - value = 0.1028

*Discussion:* - Since the  $p$ -value is greater than the alpha value we fail to

*reject the null hypothesis that says the data is not stationary. So, our data (Total Accident) is not stationary we need to take some remedial action to make stationary. The best remedial action for nonstationary time series data is differencing.*

### Differencing

Differencing is the usual technique to change certain nonstationary time series data in to stationary. The following plots are the time series plot of the differenced data. As we can observe from the plot the series fluctuates around 0 which implies that the mean of the data is constant over time. This is simply meaning the data is stationary by time series plot of the differenced total accident data.

**Figure 4.** Time Series Plot of Differenced Total Accident



*The R output of Augmented Dickey-Fuller test of the differenced data*

adf. test (Differenced Total Accident, k = 2)

Augmented Dickey-Fuller Test

Data: Differenced Total Accident

Dickey-Fuller = -5.8948, Lag order = 2, p-value = 0.01

Alternative hypothesis: stationary as we see from the R output for augmented dickey-fuller test for differenced total accident at  $\alpha = 5\%$  and lag order = 2 the  $p - value = 0.01$

*Discussion:* - Since the  $p$ -value is less than the alpha value so we reject the null hypothesis that says the data is not stationary. So, our data (Differenced Total Accident) is stationary.

### *Trend Analysis*

Analysis of the trend shows whether the total accident was increase or not in the study area. We use two methods for this test first statistical tests and graphical observations.

### *Statistical test*

H0 = The series not follow a trend

H1 = Not H0 (follow a trend)

For this test we use mann-kendall trend test

*The R output of mann-kendall test of the differenced data*

mk. test (TotalAccident, continuity = TRUE)

Mann-Kendall trend test

Data: TotalAccident

$z = -1.1253$ ,  $n = 60$ ,  $p\text{-value} = 0.2605$

Alternative hypothesis: true  $S$  is not equal to 0

Sample estimates:

$S: -177.0000000$   $\text{varS}: 24461.0000000$   $\text{tau}: -0.1023092$

*Discussion:*  $p\text{-value}$  is greater than alpha value so we do not reject the null hypothesis so the total accident does not follow trend.

*Now we are going to test the differenced data by same test.*

mk. test (DifferencedTotalAccident)

Mann-Kendall trend test

Data: DifferencedTotalAccident

$z = 0.57023$ ,  $n = 59$ ,  $p\text{-value} = 0.5685$

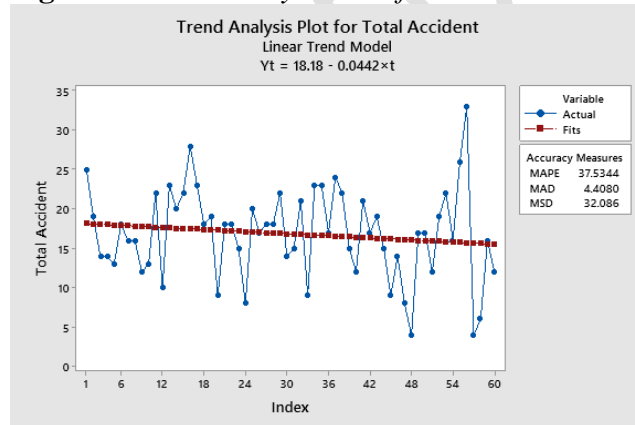
Alternative hypothesis: true  $S$  is not equal to 0

Sample estimates:

*Discussion:*  $p\text{-value}$  is greater than alpha value so we do not reject the null hypothesis so the differenced total accident does not follow trend.

*The following graphs are the graphs of Trend analysis of the total accident.*

**Figure 5. Trend Analysis Plot for Total Accident**



The Linear Trend Model written at the top of the graph is:

$$Y_t = 18.18 - 0.0442 \times t$$

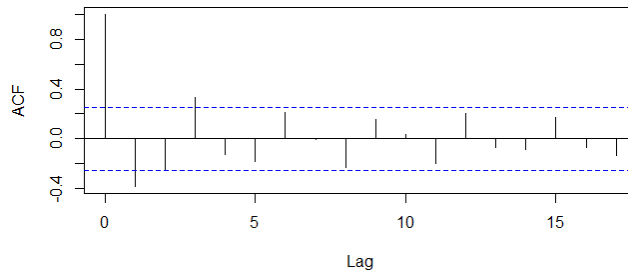
which is meaning that the total accident was decreasing by 0.0442 as the time is changed by one month.

From the fluctuation of the data on (Figure ) there is no evidence of seasonal component and so that we cannot estimate the seasonal component of the time series data of RTA we used in this study.

### Box-Jenkins Approach

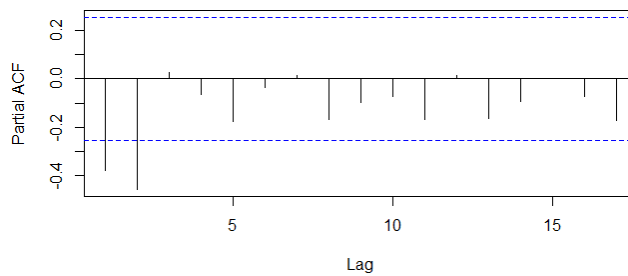
The basic steps of Box-Jenkins procedures are analyzed in the following ways. Reason for selecting ARIMA model is both ACF and PACF exponentially decay.

**Figure 6. ACF Plot of Differenced Total Accident**  
Series DifferencedTotalAccident



ACF is significant on lag 1 and 3

**Figure 7. PACF Plot of Differenced Total Accident**  
Series DifferencedTotalAccident



PACF is significant on lag 1 and 2

### ARIMA Model

#### Model Selection Criteria

ARIMA is the abbreviation of Autoregressive Integrated Moving Average. To fit the ARIMA model we need to find the appropriate order of the Autoregressive and Moving Average. We know that we made differencing of lag 1 before, to make the data stationary which is meaning that the order of the difference is 1 ( $d = 1$ ). To select the appropriate order of AR and MA different selection mechanisms have been employed. The possible significant lags are selected based on ACF and PACF. Then we find the appropriate ARIMA model by exchanging the possible significant lags. ARIMA model is said to be the best if it has the lowest AIC and BIC lowest volatility and highest log-likelihood. From (

Figure 6.) of ACF 1 and 3 lags are slightly cuts off the 95% confidence interval. This indicates that the maximum lag of Moving Average component is 3 (MA (3)). The PACF (

Figure ) of differenced total accident indicates 1 and 2 lags are significant and so that the maximum order of AR is 2 (AR (2)). The possible arima orders for differenced total accident are (1,1,1), (1,1,3), (2,1,1) and (2,1,3). The following data is taken from the output of R for the corresponding arima order to select the best fit model for our data.

**Table 5. Model Selection Criteria**

Model	AIC	AICc	BIC	$\sigma^2$ (Volatility)	Log-likelihood
ARIMA (1,1,1)	382.4	382.84	388.64	33.55	-188.2
ARIMA (1,1,3)	384.43	385.57	394.82	33.51	187.22
ARIMA (2,1,1)	387.79	388.53	396.1	38.1	-189.9
ARIMA (2,1,3)	379.49	381.11	391.96	28.47	-183.75

Arima (TotalAccident, order = c (2,1,3))  
 Series: TotalAccident  
 ARIMA (2,1,3)  
 Coefficients:

**Table 6. The ARIMA**

	ar1	ar2	ma1	ma2	ma3
	-0.9553	-0.9589	0.1270	-0.0869	-0.9810
s. e.	0.0419	0.0465	0.1331	0.1457	0.1342

sigma^2 estimated as 28.47: log-likelihood=-183.75  
 AIC=379.49 AICc=381.11 BIC=391.96

The above (

Table ) suggests that, the ARIMA (2,1,3) has the lowest AICc, lowest volatility and highest log-likelihood. Therefore, ARIMA (2,1,3) is the best model to fit our data of total accident. Hence, the model output will be: Coeftest (Arima (TotalAccident, order = c(2,1,3)))

z test of coefficients:

**Table 7. Nonreason ARIMA model for non-seasonal ARIMA**

	Estimate	Std. Error	z value	Pr(> z )
ar1	-0.955288	0.041861	-22.8203	< 2.2e-16 ***
ar2	-0.958867	0.046486	-20.6269	< 2.2e-16 ***
ma1	0.126991	0.133131	0.9539	0.3401
ma2	-0.086924	0.145696	-0.5966	0.5508
ma3	-0.980992	0.134227	-7.3085	2.702e-13 ***

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

The non seasonal ARIMA model for total accident is



$$Y_t = -0.9553Y_{t-1} - 0.9589Y_{t-2} + a_t + 0.1270a_{t-1} - 0.0869a_{t-2} - 0.9810a_{t-3}$$

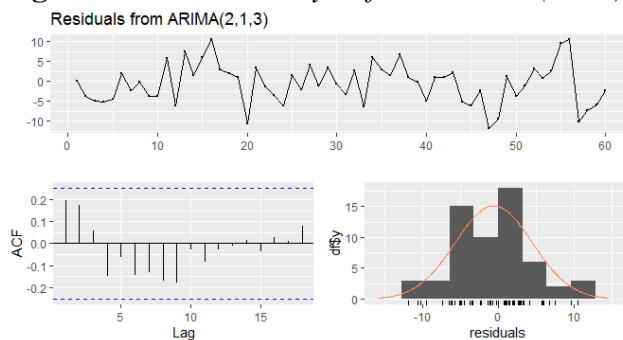
ONLY FOR REVIEW

### Residual Analysis

After we select the best fit model for our data the next step is making diagnosis about residuals of our model before we use it further.

The lower right plot is a histogram of the residuals. The histogram of the residuals is almost normally distributed. This means that the normality assumption of the error term is not violated. The lower right plot is acf of residual. As we can see from ACF plot of residual all the autocorrelations are insignificant or within the interval, thus we can say that the residuals are independent. The upper plot is a time series plot of residuals. The time series plot of the residuals indicates that there is no pattern being followed. This means that the residuals are random or constant variance. Therefore, this model can be used for prediction and policy making.

**Figure 8. Residual Analysis from ARIMA (2,1,3)**

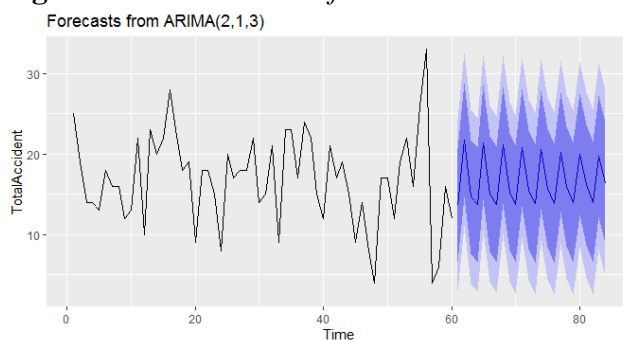


### Forecasting

After identification, estimation of parameters and diagnostic checking procedure, the next step is forecasting with appropriate model for future values. We have the appropriate identified ARIMA (2,1,3) model for total accident. We are going to forecast next 24 months from 2021-2022 which is years for the future.

Generally, the total road traffic accident in west Arsi zone for next two years i.e. 2021-2022. C using ARIMA (2,1,3) model will be look like this:

**Figure 9. Forecast Plot of Total Accident.**



From (Figure ), the first 60 observations are the original total accident data. The upper and lower shaded region is the 80 and 95 percent confidence region of the forecasting values and the center line is forecasting line. The line indicates that the forecasting values are almost decreasing. The point forecast and the corresponding 80% and 95% confidence intervals for the values are attached to this document at the end (APPENDIX). To test the accuracy of the forecasting values, the Box Ljung test of the errors is applied and the output is the following.

*Modified Box-Pierce (Ljung-Box) Chi-Square Statistic*

Here, from (

Table ) p-values is greater than 0.05, suggesting that there are no significant autocorrelations between successive forecasting errors. Therefore, the model used for forecasting and the forecast is accurate.

**Table 8. Modified Ljung-Box**

Lag	12	24	36	48
Chi-Square	5.96	13.53	26.58	39.14
DF	8	20	32	44
P-Value	0.651	0.853	0.737	0.680

*R output for measures of forecast*

Accuracy (Arima (TotalAccident, order = c (2,1,3)))

**Table 9. Training set**

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Trainin	-0.63030	5.06193	4.06870	-19.4121	34.6393	0.72523	0.19729
g set	17	2	9	4	3	81	29

*Discussion:* - we do have MAPE and RMSE value which are measures forecasting error which expected to small in measure of root mean square error our forecast accurate 94.94% similarly, in measure of mean percentage error our forecast is about 65.36%.

*Conclusion:* - Our forecast is approximately 65.39% accurate

*Discussion*

Road traffic accident have become a major world public health concern and economic burden claiming thousands of lives, resulting in thousands of physical injuries and costing the economy millions of dollars.

The purpose of this study was to model RTA in west Arsi zone between the 2016 and 2021 and forecast it for the next 2 years by using the time series analysis. Our study findings showed a decreasing trend over the past years. Also, results showed a higher percentage of fatalities in the years 2023 in comparison to all years. On the other hand, lower percentage of accident was seen in the year 2016.

Our findings indicated decreasing trend in the frequency of total accident. Another study in Amhara region indicated the observed injury RTA cases show an overall decreasing trend whereas fatal RTA cases and total RTA cases have an overall increasing trend between September 2021 to May 2025 (Getahun, [10]). Another study in Zanzibar Province, Iran indicated there was a decreasing trend of the fatalities due to traffic accidents between 2016 and 2021 (Yousefzadeh-Chabok, Ranjbar-Taklimie, Malekpouri, & Razzaghi, [28]).

The overall findings of this study indicate that 24 months of forecasts were provided for total accidents would continue in a non-decreasing trend. The evidences from other studies also showed similar findings (Getahun, [10]).

## **Conclusion and Recommendations**

### *Conclusion*

The study is conducted by using the data from secondary source. Based on the analysis conducted in this research on the monthly road traffic accident, the following ideas were concluded. The time series plot indicates that the total road traffic accident is time dependent and it has a decreasing trend throughout the time. The plot didn't reveal any evidence of seasonal component and as a result the seasonal estimation is not made in this research. Before applying the Box-Jenkins methodology, the stationarity of the data has been checked by the time series plot and augmented ducky fuller test and found that the mean of the data is not constant and so that not stationary. Differencing of lag (1) is taken as a remedial action and the data becomes stationary after differencing. Then after, the ARIMA model is selected based on ACF and PACF. The assumption of parsimonious model is also used in this research not to overestimate the parameters. Each order of ARIMA model is evaluated to find out the order with the least AIC, lowest BIC, lowest volatility and highest log-likelihood and the parameter of each lag was also checked.

Accordingly, the ARIMA order (2, 1, and 3) for total accident is found to be the most appropriate model to fit the data of road traffic accident. After the model was fitted the diagnostic checking have been applied by using test of independency of residual, test of normality of residual and constant variance of residual, so that the model fitted is appropriate to forecast the total road traffic accident. Since the linear trend analysis implies decreasing in number of RTA there is decreasing performance of road traffic accident in west Arsi zone in general.

### *Recommendations*

The main reason the problem solver is not done is data collection. Just they save a data for only report case here the research is not done by report Example: As the accident occur if a data is recoded on computer any researcher simply extracts a data it can simply solve the required problem.

Based on the result obtained in this study, we need to recommend every concerned body to take an action.

To control the harmfulness of the road traffic accidents, every individual in the country should have to be informed about the disasters of car accident and also must have awareness of the accident.

Having a serious consideration of the problem, all the concerned bodies not only West Arsi Zone Traffic Office but also the society as the whole should have to take care of the occurrence of the traffic accident since it has economic and social effects in the countries.

## References

- Abdi, T. A. (2017). Road Crashes in Addis Ababa, Ethiopia. Tariku Ayana, Abdi.
- Aga, M. A., Woldeamanuel, B. T., & Tadesse, M. (2021). Statistical Modeling of numbers of human deaths per road traffic accident in the Oromia region, Ethiopia. PLOS ONE.
- Amanuel, K. (2017, February 3). Trends, Causes, and Costs of Road Traffic Accidents in Ethiopia.
- Asongwa, S. (1992). Road traffic accidents in Nigeria: A review and a reappraisal, Accident Analysis and Prevention.
- Barreto, S. M., Swerdlow, A., Smith, P. g., & Higgins, C. D. (1997). Risk of death from motor-vehicle injury in Brazilian steelworkers: A nested case-control study.
- Bedaso, A., Kediro, G., Ebrahim, J., Tadesse, F., Mokennen, S., Gobena, N., & Gebrehana, E. (2020). Prevalence and determinants of post-traumatic stress disorder among road traffic accident survivors. BMC Emergency Medicine.
- Chatfield. (2004). Time Series Analysis and Forecasting.
- CSA. (April, 2012). Population and Housing Census of Ethiopia: Administrative Report By Central Statistical Authority,. Central Statics Agency.
- Fisseha, M. H., & Sileshi, T. (2014). The neglected health problem in Amhara National Regional State, Ethiopia. Road traffic accident.
- Getahun, K. A. (2021). Time series modeling of road traffic accident. Journal of Big Data, 8-13.
- Getu, S. (2007). Causes of Road Traffic Accidents and possible counter measures on Addis Ababa- Shashemne Roads. MSc Thesis Addis Ababa University department of Civil Engineering.
- Goswami A, Sonowal R. (2009). A statistical analysis of road traffic accidents in Dibrugarh City, Assam, India. Division of Epidemiology and Nutrition, Regional Medical Research Centre.
- Jonathan D. Cryer, K.-S. (2008). Time Series with Application in R. Iowa City, Iowa: springer.
- Jørgensen, S. H., & Abane, A. M. (1999). A comparative study of urban traffic accidents in developing and developed countries:. 123-128.
- Kong, L., Lekawa, M., Navarro, R., McGrath, M., Margulies, & Hich J. R. (1996). Risk Factors and Road Traffic Accidents in Tanzania:.
- Leon, S. (1996). Reducing death on the Road. The effects of minimum safety standards Unpublicised crash test, seat belts and alcohol..
- Odero, W. (1995). Road traffic accidents in Kenya: an epidemiological appraisal.
- Persson. A. (2008). Road traffic accidents in Ethiopia: Magnitude, causes and possible interventions. Advances in Transportation Studies. Road traffic accidents in Ethiopia:

- magnitude, causes and possible interventions, pp. 5-16.
- Rivara, F., & Barber, M. (1985). Demographic analysis of childhood pedestrian.
- Rundmo, T., & Iversen, H. H. (2004). Risk Perception and Driving Behaviour among Adolescents in Two Norwegian Counties before and after a Traffic Safety Campaign. The Centres for Disease Control and Prevention. (2013). Fact Sheet.
- UNECA. (2009). Case Study: Road Safety in Ethiopia Final Report. United Nations. Economic Commission for Africa. Retrieved from United Nations. Economic Commission for Africa.
- WHO. (2013). Road Safety Status. Retrieved from World Health Organization: [https://www.who.int/violence\\_injury\\_prevention/road\\_safety\\_status/2013/report/factsheet\\_afro.pdf](https://www.who.int/violence_injury_prevention/road_safety_status/2013/report/factsheet_afro.pdf)
- WHO. (2015). Global status report. Retrieved from World Health Organization: [https://www.who.int/violence\\_injury\\_prevention/road\\_safety\\_status/2015/Executive\\_summary\\_GSRRS2015.pdf](https://www.who.int/violence_injury_prevention/road_safety_status/2015/Executive_summary_GSRRS2015.pdf)
- WHO. (2018, June 17). Global Status Report on Road Safety. Retrieved from World Health Organization: [https://www.who.int/violence\\_injury\\_prevention/road\\_safety\\_status/2018/GSRRS2018\\_Summary\\_EN.pdf](https://www.who.int/violence_injury_prevention/road_safety_status/2018/GSRRS2018_Summary_EN.pdf)
- WHO. (2021, June 21). WHO. Retrieved from World Health Organization: <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>
- Yirgu, A. N., & Wakwaya, S. B. (2020). Prevalence and Factors Associated with Road traffic Crash Among Bajaj and Motor Bicycle Drivers in Negelle ArsiTown Oromia Regional State, South East Ethiopia. International Journal of Advanced Research (IJAR).
- Yousefzadeh-Chabok, S., Ranjbar-Taklimie, F., Malekpouri, R., & Razzaghi, A. (2016). A Time Series Model for Assessing the Trend and Forecasting the Road Traffic Accident mortality. 3-4.
- Desta Sodano Sheiso, Mekashew Ali Mohye. Euler's Method for Solving Logistic Growth Model Using MATLAB. International Journal of Systems Science and Applied Mathematics. Vol. 7, No. 3, 2022, pp. 60-65. doi: 10.11648/j.ijssam.20220703.13