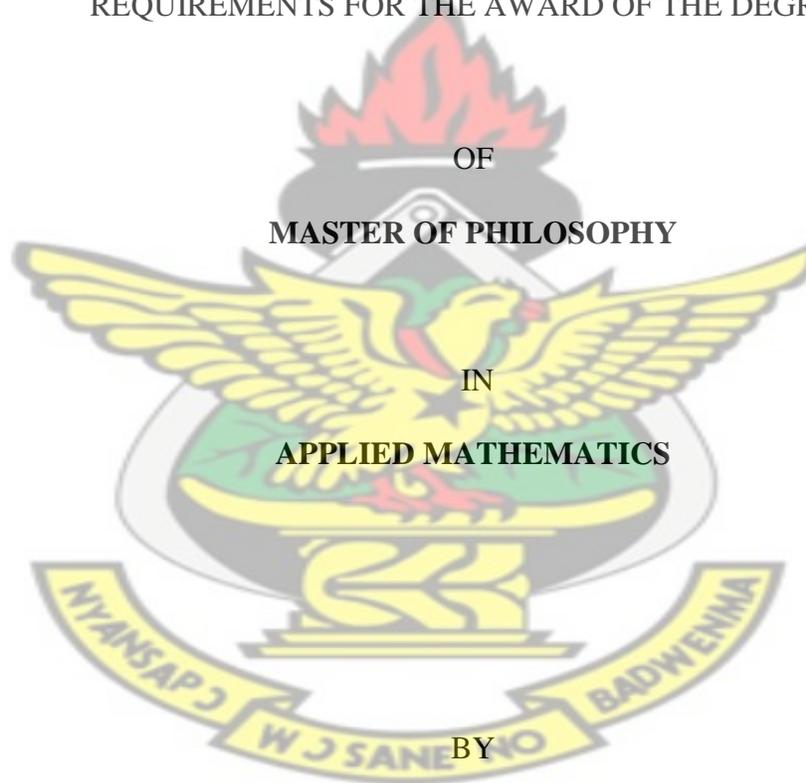


**ANALYSIS AND MODELING OF PREVALENCE OF MEASLES IN THE
ASHANTI REGION OF GHANA**

A THESIS SUBMITTED TO THE GRADUATE SCHOOL BOARD,
KWAME NKRUMAH UNIVERSITY OF SCIENCE AND
TECHNOLOGY, KUMASI, GHANA IN PARTIAL FULFILMENT OF THE
REQUIREMENTS FOR THE AWARD OF THE DEGREE



KWAME ASARE GYASI-AGYEI

JUNE, 2012

DECLARATION

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

KNUST

Kwame Asare Gyasi-Agyei

(PG 5070510)

Student Name and ID

Signature

Date

Rev. W. Obeng Denteh

Supervisor's Name

Signature

Date

Mr. Kwaku F. Darkwa

Head of Dept. Name

Signature

Date



DEDICATION

I dedicate this thesis to my mother, Madam Adwoa Twumwaa Gyasi-Agyei, my wife, Philomena Gyasi-Agyei and my children, Nicholas Gyasi-Agyei, Adelaide Gyasi-Agyei, Kingsley Gyasi-Agyei and Jessica Gyasi-Agyei.

KNUST



ACKNOWLEDGEMENTS

My deepest gratitude goes to God the Father and the Lord Jesus Christ for His guidance and protection that have enabled me to complete this thesis.

I most heartily wish to express my warm appreciation to Nana Kena Frimpong and Rev. W. Obeng Denteh, my supervisors, for their detailed painstaking attention to read through this thesis and the consequent feedback which improved the thesis' quality.

I wish to acknowledge my particular indebtedness to all my lecturers at Kwame Nkrumah University of Science and Technology for their support and the good education that they provided to us.

Special thanks go to my headmaster, Very Rev. Isaac Osei Boadi, of Juaben Senior High School and all the teachers at Juaben Senior High School for their advice and support.

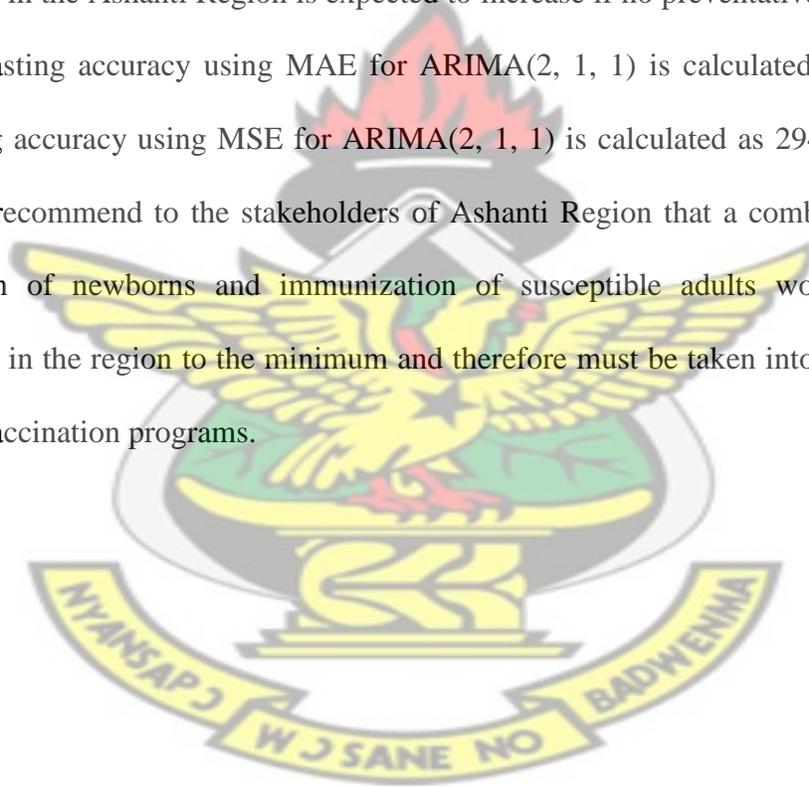
I am also grateful to all the authors whose works I have directly or indirectly made use of.

My profound thanks go to my brothers and sisters for their moral and financial support which have made this thesis possible, especially Dr. N. Gyasi-Agyei, Prof. Y. Gyasi-Agyei and Prof. A. Gyasi-Agyei. Also deserving mention are Mr. Omane Adjepong Maurice, Mr. Kofi Ababio Kicupson, Mr. James K. Agyen, Mr. George Agblenowo, of Juaben Senior High and the rest of my friends who have contributed positively to this thesis. Many people have helped to make this thesis possible, and to all of them, I say a sincere thank you.

Indeed it has been sleepless nights, a long and tedious road to the publication of this thesis, but to the glory of God, it has come out to see the light. It is my fervent prayer and hope that people will find it just as useful and stimulating as it is intended to be.

ABSTRACT

In this thesis, autoregressive integrated moving average (ARIMA) model is used to predict the prevalence and incidence of measles in the Ashanti Region of Ghana. The Mean Absolute Error (MAE) and the Mean Square Error (MSE) are used to compare the in-sample forecasting performance of four selected candidates' models. The working data from the Ashanti Health Services spans from 2001 to 2011. It is evident from the analysis that measles data in the Ashanti Region of Ghana could best be modelled with ARIMA(2, 1, 1) and that measles prevalence in the Ashanti Region is expected to increase if no preventative measures are taken. The forecasting accuracy using MAE for ARIMA(2, 1, 1) is calculated as 28.1141 and the forecasting accuracy using MSE for ARIMA(2, 1, 1) is calculated as 2947.15. The results of the study recommend to the stakeholders of Ashanti Region that a combination of increased vaccination of newborns and immunization of susceptible adults would reduce measles prevalence in the region to the minimum and therefore must be taken into consideration in the region's vaccination programs.



CONTENTS

Title Page	i
Declaration.....	ii
Dedication.....	iii
Acknowledgements.....	iv
Abstract.....	v

KNUST

CHAPTER 1

INTRODUCTION.....	1
1.1 Biological Background of the Thesis.....	1
1.1-1 Definitions of Measles	1
1.1-2 Cause of Measles	1
1.1-3 Measles Virus Pictures and Images	2
1.1-4 Transmission of Measles	3
1.1-5 Symptoms, Incubation Period and Rashes of Measles Attack	3
1.1-6 Prevention from Measles	4
1.1-7 Complications and Risk Factors of Measles	5
1.1-8 Treatment of Measles	6
1.1-9 Vaccination of Measles	6
1.1-10 Recovery after measles	6
1.2 The Studied Geographical Area: Ashanti Region of Ghana	7
1.3 Motivation for the Thesis.....	9
1.4 Objective of the Study	9

1.5 Significance of the Study	10
1.6 Materials and Methodology	10
1.7 Structure of the Study	11

CHAPTER 2

REVIEW OF RELATED LITERATURE 12

2.0 Introduction	12
2.1 Sample Abstracts Relevant to this Thesis	12

CHAPTER 3

METHODOLOGY 29

3.0 Introduction	29
3.1 Definitions of Time Series	29
3.2 Objective of Time Series Analysis	30
3.2-1 Description	30
3.2-2 Explanation	31
3.2-3 Prediction	31
3.2-4 Control	31
3.3 Components of a Time Series	32
3.3-1 Secular Trend	32
3.3-2 Seasonal Variation	32
3.3-3 Cyclical Variations	33
3.3-4 Irregular Variation	33
3.4 Stationary Time Series	33

3.4-1 Achieving Stationarity (Differencing)	34
3.5 Auto-correlation and Partial Auto-correlation Functions	35
3.5-1 Auto-correlation Function (ACF)	35
3.5-2 The Sample Distribution of Autocorrelation Function Coefficients	35
3.5-3 Partial Autocorrelation Function (PACF)	36
3.5-4 White Noise	36
3.6 Autoregressive Integrated Moving Average Models (ARIMA)	37
3.6-1 An Autoregressive Model of Order p , AR (p)	37
3.6-2 Autoregressive Process of order 2 denoted by AR (2)	38
3.6-3 Moving Average Model of order q , MA (q)	38
3.6-4 Autoregressive Moving Average Model (ARMA)	39
3.6-5 ARMA (2, 1) Model	39
3.7 The AR and MA polynomials	40
3.8 Causality and Invertibility of an ARMA (p, q) Model	40
3.9 Estimating the Parameters of an ARMA Model	41
3.10 The Autoregressive Integrated Moving Average Model	42
3.10-1 ARIMA (1, 1, 1) Process	43
3.11 Seasonal Autoregressive Integrated Moving Average Model	43
3.11-1 Purely Seasonal Model	44
3.11-2 Seasonal Moving Average Models	45
3.11-3 Mixed SAR and SMA Models	45
3.12 The Box–Jenkins Methodology for ARIMA Models	46
3.13 Other Specification Methods	47
3.13-1 The Akaike’s Information Criteria (AIC)	48
3.13-2 The Schwartz’s Bayesian Information Criteria (BIC)	49

3.13-3 Estimation of the Parameters of the Model Identified	50
3.13-4 Testing the Model for Adequacy (Portmanteau Test)	50
3.14 Measuring Forecasting Accuracy (Error Metrics).....	51
3.15 The Diebold-Mariano Statistic for Comparing Predictive Accuracy.....	52

CHAPTER 4

ANALYSIS AND RESULTS 53

4.0 Introduction	53
4.1 Time Plot of Prevalence of Measles in the Ashanti Region of Ghana.....	53
4.2 Stationarity Checks Using the ACF, PACF, KPSS and Dickey-Fuller.....	54
4.3 First-Order Differencing of Prevalence of Measles in the Ashanti Region from January, 2001 to November, 2011.....	57
4.4 Objective Test for Stationarity for the First Order Differenced Series.....	58
4.5 Selecting Computing Models Using ACF and PACF of the First Order Differencing of Measles Prevalence	58
4.6 Estimation of Tentative Models.....	60
4.6.1 Parameter estimate and diagnostics of ARIMA (0, 1, 2) model.....	60
4.6.2 Parameter estimate and diagnostics of ARIMA (2, 1, 0) model.....	62
4.6.3 Parameter estimate and diagnostics of ARIMA (3, 1, 1) model.....	64
4.6.4 Parameter estimate and diagnostics of ARIMA (2, 1, 1) model.....	66
4.7 Forecasting From ARIMA(2, 1, 1).....	68
4.8 The Error Metrics.....	69
4.9 Diebold – Mariano (DM) Test.....	70
4.10 Time plot of actual measles data and the fitted Models.....	71

CHAPTER 5

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS 72

5.1 Conclusion 72

5.2 Recommendations 73

REFERENCES..... 74

APPENDIX..... 78

KNUST



CHAPTER 1

INTRODUCTION

1.1 Background of the Thesis

Measles is one of the leading causes of death among young children even though a safe and cost-effective anti-measles vaccine is available. In the year 2008 alone 164 000 measles related deaths were reported globally – nearly 450 deaths every day or 18 deaths every hour. More than 95% of measles deaths occur in low-income countries with weak health infrastructure. Measles is a human disease and it is not known to occur in animals (WHO, 2011).

1.1-1 Definitions of Measles

Measles is one of the most contagious but vaccine-preventable diseases which is caused by the measles virus. It is one of the most readily communicable diseases and probably the best known, and most deadly of all childhood rash/fever illnesses. It is a childhood disease that rarely occurs in adults (Wikipedia, 2008).

Measles is a respiratory disease caused by a virus. As noted above, the disease of measles and the virus that causes it share the same name. Measles virus normally grows in the cells that line the back of the throat and lungs (CDC, 2011).

1.1-2 Causes of Measles

Measles virus is spread through the respiratory route. This virus is contained in the millions of tiny droplets that come out of the nose and mouth when a measles carrier coughs or sneezes. One can catch measles by breathing in these droplets or, if the

droplets have settled on a surface, by touching the surface and then placing the hands near the nose or mouth. The measles virus can survive on surfaces for a few hours. In fact, the virus is one of the most contagious viruses known to man. As a result, it can spread rapidly in a susceptible population. Infected people carry the virus in their respiratory tract before they get sick, so they can spread the disease without being aware of it (Nettleman, 2008).

If people are immune to the virus (either through vaccination or by having had measles in the past), they cannot get the disease caused by that virus. For example, someone who had measles as a child would not be able to get the disease again.

1.1-3 Measles Virus Pictures and Images

Figure 1.1 depicts a selection of measles viruses; these are the way the measles virus looks under the microscope. Sources: (CDC, 2011), (Schoenstadt, 2006)

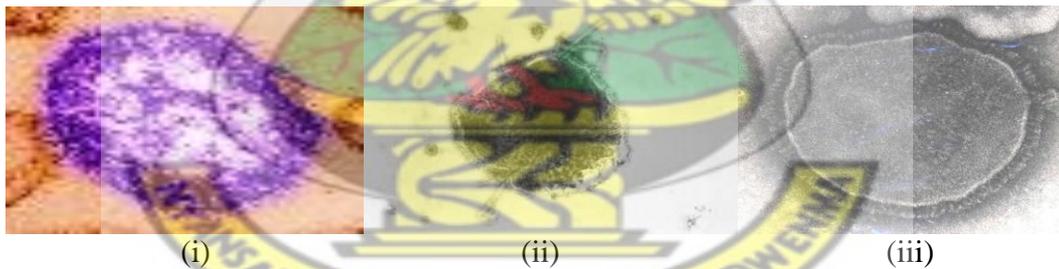


Figure 1.1: *Types of Measles Virus.*

Measles virus resides in the mucus in the nose and throat of the infected person. The virus only infects humans. The virus is rapidly inactivated by heat, light, acidic pH, ether, and trypsin (an enzyme). It has a short survival time (under 2 hours) in the air or on

objects and surfaces. The majority of people infected with the virus recover, but measles complications can be dangerous (Schoenstadt, 2006).

1.1-4 Transmission of Measles

The measles virus resides in the mucus in the nose and throat of an infected person, so transmission typically occurs through coughing and sneezing. Measles is an illness that is spread through the coughs and sneezes of infected people. After being infected with the virus, a person does not become sick immediately; it takes several days for symptoms to appear. Transmission of measles occurs so easily that anyone who is not immunized will probably get the disease eventually (Schoenstadt, 2006).

1.1-5 Symptoms, Incubation Period and Rashes of Measles Attack

Symptoms do not appear immediately when a person becomes infected with the measles virus. When a person becomes infected with the measles virus, it begins to multiply within the cells that line the lungs and the back of the throat. The virus can also spread to the lymph glands, bone marrow, liver, eyes, thymus, tonsils, spleen, skin and brain.

The symptoms typically appear ten to fourteen days after a person is infected with the measles virus. The period between measles transmission and the start of symptoms is called the incubation period. During this period, the virus is multiplying. It includes fever, sore throat, cough, sore eyes, red watery eyes, vomiting, runny nose, loss of appetite and fatigue. *Koplik's spot* is the characteristic symptom of measles. It appears on the inside of the mouth. Koplik's spots are small white areas that may have bluish-colored centers. The skin rash appears within three to five days of the onset of symptoms.

The rash often begins on the face and spreads downward all over the body. A very high fever may develop with the rash. The rash starts to disappear after a few days, and the fever resolves. It is better to leave the rashes alone as scratching leaves the patient in worse condition (Nettleman, 2008).

Figure 1.2 shows a few pictures of people under measles attack. As noted below it is known that the appearance of measles can differ from the pictures shown, they are just for information purpose.

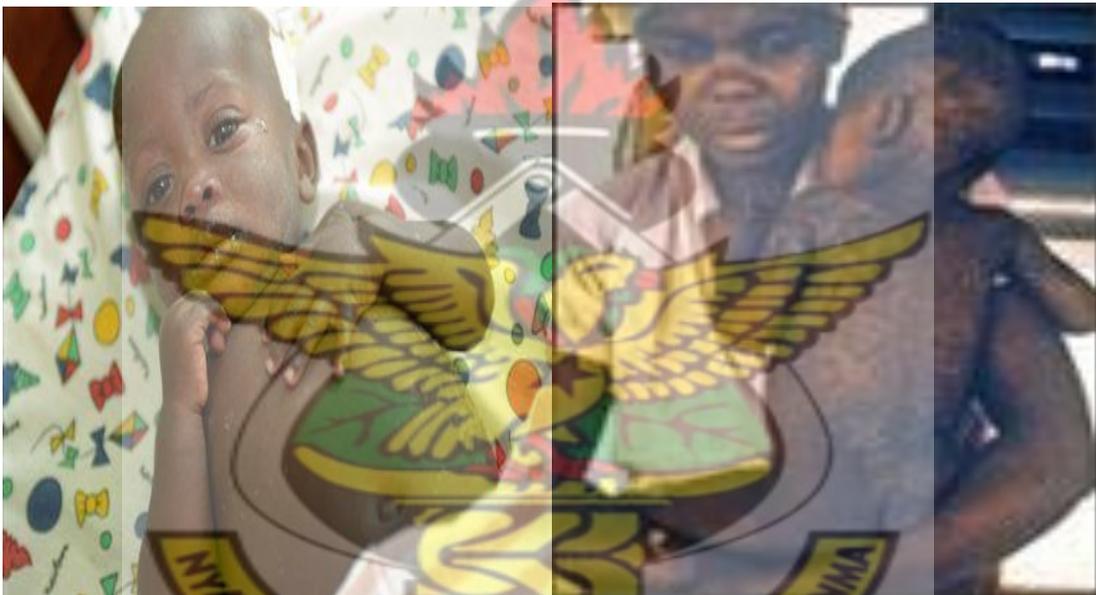


Figure 1.2: *Pictures of Measles Images*

1.1-6 Prevention from Measles

Generally two doses of live measles vaccine are recommended; one shot at 15 months of age, and the second shot before entering either kindergarten or class one. Measles vaccine is administered to children between 12 and 15 months, but can be done from 6 months

during an epidemic of measles. Children and adults exposed to measles virus, which have not developed immunity to the disease, can be vaccinated within 3 days after exposure. The vaccine is prescribed for pregnant women and children less than one year. Instead, for these categories of people it is preferred to use immunoglobulin (antibodies), administered within 2 days after exposure to the virus. Immunization is recommended one time for all persons born after 1956 who lack evidence of immunity to measles. A second dose of measles vaccine is recommended for young adults in settings where individuals congregate (WHO, 2011). The best way to prevent measles is to get the measles vaccine (Schoenstadt, 2006).

1.1-7 Complications and Risk Factors of Measles

Complications of measles include the following: ear infections, diarrhoea, pneumonia, seizures and encephalitis (inflammation of the brain) – this is rare, but can cause permanent brain damage or death. Up to 30 percent of people with measles will develop complications – usually children under five and adults over the age of 20. Measles during pregnancy increases the risk of miscarriage, premature labour and low birth-weight babies (WHO, 2011).

Unvaccinated young children are at highest risk of contracting measles and its complications, including death. Any non-immune person (who has not been vaccinated or previously recovered from the disease) can become infected. The risk factors include the following: lack of immunization with the measles vaccine, travel to, or residence in, a country where measles is still prevalent, vitamin A deficiency (WHO, 2011).

1.1-8 Treatment of Measles

There is no treatment that can kill the measles virus, so treatment focuses on supportive care, or the relief of symptoms. Supportive care can include the following:

- Get plenty of rest.
- Sponge baths with warm water may reduce discomfort due to fever.
- Medications to control fever or pain, and antibiotics to treat secondary infections.
- Drink plenty of fluids to help avoid dehydration.
- A humidifier or vaporizer may ease the cough.

Remember never to give *aspirin* to children or teenagers because it may cause a disease known as *Reye syndrome* (Nettleman, 2008).

1.1-9 Vaccination of Measles

The measles vaccine and the combined Measles Mumps Rubella (MMR) vaccine are very safe and effective and generally have few side effects. Mild reactions such as fever, redness or swelling at the injection site have been reported. As with any medicine, there is a slim chance that serious problems could occur after getting the vaccine. However, the potential risks associated with measles disease are much greater than the potential risks associated with the measles vaccine. MMR vaccine should not be given to persons who are pregnant or severely immuno suppressed patients (WHO, 2011).

1.1-10 Recovery after measles

As in any viral infection, measles-infected patients should be encouraged to drink plenty of water, fruit juices, tea or lemonade. These will be used to replace fluids lost through sweat and heat during febrile episodes. Feed measles patients with mashed vegetable,

soup, pilaf, mashed meat, perris and food prepared using steam cooking. You may also give him yogurt, bananas, apples and carrots lynx. Children with measles need more rest to recover. Usually the child can safely return to school after 7-10 days after the fever and the rash have disappeared. However, a medical consultation may be necessary to diagnose the exact state of the episode (WHO, 2011).

1.2 The Studied Geographical Area: Ashanti Region of Ghana

The Ashanti Region is the third largest of 10 administrative regions in Ghana, occupying a total land surface of 24,389 square kilometers or 10.2 per cent of the total land area of Ghana. In terms of population, however, it is the most populated region with a population of 3,612,950 in 2000, accounting for 19.1 per cent of Ghana's total population. The Ashanti region also harbors the capital city of Kumasi. The Ashanti region is centrally located in the middle belt of Ghana. It lies between longitudes 0.15W and 2.25W, and latitudes 5.50N and 7.46N. The region shares boundaries with four of the ten political regions, Brong - Ahafo Region in the north, Eastern region in the east, Central region in the south and Western region in the South west. The region is divided into 27 districts, each headed by a district chief executive (Wikipedia, 2011).

The center of population of the Ashanti Region is located in the Kumasi Metropolitan District. According to the 2000 census, the region had a population of 3,612,950, making it the most populous region. However, its density (148.1 per square km) is lower than those of the Greater Accra (895.5/km²) and Central (162.2/km²) Region. Majority of the region's population are Ghanaians by birth (87.3%) with about five per cent naturalized

Ghanaians. A smaller proportion (5.8%) of the population originate from outside Ghana, made up of 3.7 per cent mainly from the five English-speaking countries of ECOWAS and 2.1 per cent from other African countries. The non-African population living in the region is 1.8 per cent of the total population. Akans are the predominant ethnic group in the region, representing 77.9% of Ghanaians by birth. A high proportion (78.9%) of the Akan population is the Asantes. The non-Akan population in the region comprises the Mole- Dagbon (9.0%), the Ewe (3.2%), the Grusi (2.4%), the Mande-Busanga (1.8%) and the Ga- Dangme (1.4%). The other smaller ethnic groups form about 1.3 per cent of the population of the region (Wikipedia, 2011).

Farming is the predominant occupation in the Ashanti Region. Crops cultivated include cocoa and foodstuffs; gold mining is also a major economic activity engaged in owing to the high amounts of gold deposits. Gold mining is said to be the oldest industry in the region as it stated long before the first European set foot on the Ashanti land. Timber haulage is done on a large scale. The region consists of deciduous and tropical rain forest. The extreme west of the region receives the highest amount of rainfall. It has tropical climate characterized by moderate temperatures all year round (Wikipedia, 2011).

Unfortunately, at the time of completion of this thesis, the population and housing census data of Ghana for 2010 was not made available so I used that of 2000.

1.3 Motivation for the Thesis (Problem Statement)

Measles is one of the leading causes of death among young children even though a safe and cost-effective vaccine is available. It is so serious that Dillner (2001) stated that in the developing world, mothers say, "never count your children until after the measles." (WHO, 2011).

Measles can also make a pregnant woman have a miscarriage, give birth prematurely and low birth-weight babies (WHO, 2011).

The major problem of measles is that Measles weakens the immune system and opens the door to secondary health problems, such as pneumonia, blindness, diarrhoea, encephalitis etc. (WHO, 2011).

When one person has measles, 90 percent of the people they come into close contact with will become infected, if they are not already immune to it (Schoenstadt, 2006).

1.4 Objective of the Thesis

The following were the objectives of the thesis:

- i. to observe the pattern of measles infections in the Ashanti Region of Ghana from January, 2001 to November, 2011.
- ii. to model the prevalence of measles in the Ashanti Region of Ghana using the Box-Jenkins ARIMA process.
- iii. to evaluate the impact of AIC on in-sample forecasting performance of the selected ARIMA models.

1.5 Significance of the Thesis

- i. This thesis provides a method for assessing the effectiveness of measles in the Ashanti Region of Ghana for the near future.
- ii. This thesis contributes to the research information on measles in the country, so that it can help in further work in the area of research to instigate the effects of the disease on Ghanaians.
- iii. The thesis attempts to present both application and theory at a level accessible to a wide variety of students and researchers.

1.6 Materials and Methodology

The data used for the modeling and analysis was obtained from the Ministry of Health in the Ashanti Region of Ghana. It was a secondary data. The data on monthly bases consist of the measles cases from various hospitals in the Ashanti Region for the period of January 2001 to December 2011. The model used to analyze the collected data is the Auto-Regressive Integrated Moving Average (ARIMA) which was developed by Box and Jenkins in 1970. The R software package has been used in addition to manual calculations to model the given data. The Internet and Kwame Nkrumah University of Science and Technology school library are among the resources exploited. Details of this section can be found in chapter 3.

1.7 Structure of the Thesis

This thesis is organized as follows: Chapter 1 reviews the background of the subject matter of the thesis. Topics discussed include the biological background, problem statement, objectives of the thesis, as well as the significance and the structure of the thesis. Chapter 2 consists of the review of relevant literature governing the theory of prevalence of measles. Chapter 3 consists of the discussion of the methods used for the thesis. All the formulae used in this study are discussed thoroughly in this chapter. The Box–Jenkins method for identifying a plausible ARIMA model is given in this chapter along with techniques for parameter estimation and forecasting for these models. A partial theoretical justification of the use of ARMA models is also discussed in this chapter. Chapter 4 deals with data analysis, modeling and forecasting. In this chapter, **R** codes of modeling time series were applied; that is, all of the plots and numerical output displayed in this thesis have been produced with the **R** software. Most of the numerical outputs have been edited for additional clarity or for simplicity. Actual measles data drawn from various hospitals in the Ashanti Region of Ghana are used throughout in this chapter to illustrate the methodology explained in chapter 3. Chapter 5 contains the summary and findings, conclusions and the recommendations of the thesis. It also deals with the discussion of the results obtained from the **R** codes approach in chapter 4.

CHAPTER 2

REVIEW OF RELATED LITERATURE

2.0 Introduction

This chapter reviews the work of other researchers related to the objectives of this thesis.

2.1 Sample Abstracts Relevant to this Thesis

Many measles researchers have attempted to model epidemiology of measles epidemic in various countries. However, Mathematical models have not yet been used to study the prevalence of measles in the Ashanti Region of Ghana.

First of all, considering the topic “A discrete-time model with vaccination for measles epidemic” which is the work of Allen et al. (1991). They used a discrete-time, age-independent *SIR*-type epidemic model. The effects of vaccination were also included in the model. They verified three mathematically important properties for the model. Their solutions were non-negative, the population size was time-invariant, and the epidemic concluded with all individuals either remaining susceptible or becoming immune. They applied their model to measles epidemic on a university campus. The simulated results were in good agreement with the actual data. The results of the simulations indicated that a rate of immunity greater than 98% might be required to prevent an epidemic in a university population. Their model had applications to other contagious diseases of *SIR* type.

Allen et al. (1993) stated that an epidemic of rubella occurred on the campus of Texas Tech University in January, February and March of 1989. A vaccination programme was

initiated as soon as the epidemic was confirmed. Extensive case histories of all confirmed cases were collected by the Lubbock City Health Department and given an exhaustive statistical analysis by a group from the Department of Mathematics at Texas Tech University. The data and statistical analysis were used to formulate stochastic and deterministic models of the measles epidemic based on the standard SEIR model. The analysis and the simulations indicated that in order to prevent measles outbreak on a university campus a high rate of immunity above 98 per cent might be required.

Rhodes and Anderson (1996) presented a detailed analysis of the pattern of measles outbreaks in the small isolated community of the Faroe Islands. Measles outbreaks in that population was characterized by frequent fade-out of infection resulting in long intervals when the disease was absent from the islands. Using an analysis of the distribution of epidemic sizes and epidemic durations they proposed that the dynamical structure observed in the measles case returns reflected the existence of an underlying scaling mechanism. Consequently the dynamics were not as purely stochastic as is usually thought for epidemiological systems of that sort. They used a Lattice-Based Epidemic Model to provide a theoretical estimate of the scaling exponents and showed that a conventional compartmental *SEIR* model was unable to reproduce that result.

A Mathematical model of the dynamics of measles in New Zealand was developed in 1996. The model successfully predicted an epidemic in 1997 and was instrumental in the decision to carry out an intensive MMR (measles-mumps rubella) immunization campaign in that year. While the epidemic began some months earlier than anticipated, it

was rapidly brought under control, and its impact on the population was much reduced. In order to prevent the occurrence of further epidemics in New Zealand, an extended version of the model had since been developed and applied to the critical question of the optimal timing of *MMR* immunization (Tobias and Roberts, 2000).

Also deserving mention is the work of Ellner et al. (1998). They presented and evaluated an approach to analyzing population dynamics data. They used semi-mechanistic models. They used historical data on measles epidemics as a case study and showed how that approach could lead to better forecasts, better characterizations of the dynamics, and a better understanding of the factors causing complex population dynamics relative to either mechanistic models or purely descriptive statistical time series models. The semi-mechanistic models were found to have better forecasting accuracy than either of the model types used in previous analyses when tested on data not used to fit the models. The dynamics were characterized as being both non-linear and noisy. The dynamics oscillate between strong short-term stability and strong short-term chaos. There was statistically significant evidence for short term chaos in all data sets examined.

Stamp et al. (1990) presented a Mathematical model for the simulation of a localized measles epidemic. Their work was presented along with a computer simulation based on this model. The simulation results were compared with the results of a measles outbreak which occurred at Texas Tech University. The effectiveness of the vaccination program undertaken during the Texas Tech epidemic and the effect of altering the level of herd immunity were also considered.

A simple stochastic Mathematical model was developed and investigated for the dynamics of measles epidemic by Kassem and Ndam (2010). Their model, which was a multi-dimensional diffusion process, included susceptible individuals, latent (exposed), infected and removed individuals. Stochastic effects were assumed to arise in the process of infection of susceptible individuals. Using the best currently available parameter values, the intrinsic variability in response to a given initial infection was examined by solving the stochastic system numerically. The results of the simulation seemed to agree with the historical pattern of measles in Nigeria.

Bharti et al. (2008) stated that Mathematical models could help elucidate the spatio-temporal dynamics of epidemics as well as the impact of control measures. The model of gravity for directly transmitted diseases is currently one of the most parsimonious models for spatial epidemic spread. They used distance-weighted, population size-dependent coupling model to estimate host movement and disease incidence in meta-populations. Their model captured overall measles dynamics in terms of underlying human movement in pre-vaccination England and Wales. In spatial models, edges often present a special challenge. Therefore, to test the model's robustness, they analyzed gravity model incidence predictions for coastal cities in England and Wales. Their Results showed that, although predictions were accurate for inland towns, they significantly underestimated coastal persistence. They examined incidence, outbreak seasonality, and public transportation records, to show that the model's inaccuracies stem from an underestimation of total contacts per individual along the coast. They rescued this predicted 'edge effect' by

increasing coastal contacts to approximate the number of per capita inland contacts. Those results illustrated the impact of ‘edge effects’ on epidemic Meta-populations in general and illustrated directions for the refinement of spatiotemporal epidemic models.

Chen et al. (2007) proved that vaccination has a powerful defense against measles. They reappraised measles sero-epidemiological data in Taiwan from 1974 to 2004 having robust age-stratified serological information on exposure and immunity to quantitatively characterize measles vaccination programmes. They dynamically modeled measles sero-epidemiology to estimate age-dependent intensity of infection associated with the effects of different contact patterns on pre- and post-vaccination. The contact matrix was employed to describe the transmission between and within each age group. They used a deterministic Susceptible–Exposed–Infected–Recovery (SEIR) model to capture sub-population dynamics. Their study showed that mass regional or nationwide vaccination programmes could greatly reduce the potential for a major measles epidemic and have strong direct effects on the potential impact of childhood vaccination. They parameterized a predictive model that should reduce the socio-economic costs of measles surveillance in Taiwan and thereby encourage its continuance, especially for pre-school children.

Wallinga et al. (2005) estimated the measles reproduction ratio for eight Western European vaccination programmes. Because many plausible age-structured transmission patterns result in a similar description of the observations, it is not possible to estimate a unique value of the reproduction ratio. They developed a method to estimate bounds and

confidence intervals for plausible values of the reproduction ratios. They used maximum likelihood methods. Lower and upper bounds for plausible values of the basic reproduction ratio were estimated to be 7.17 (95% CI 7.14-7.20) and 45.41 (95% CI 9.77-49.57), corresponding to lower and upper bounds on critical vaccine coverage of 86.6% and 98.1%. Of the eight evaluated vaccination programmes, four have vaccine coverage below the lower bound and allow measles to persist, and four have vaccine coverage at the upper bound and may eventually eliminate measles.

Trottier and Philippe (2006) presented univariate time series analysis of pertussis, mumps, measles and rubella based on Box-Jenkins or Auto-Regressive Integrated Moving Average (ARIMA) modeling. The objective of their paper was to analyze the stochastic dynamics of childhood infectious disease using time series analysis. Their method, which enables the dependency structure embedded in time series data to be modeled, had potential research applications in studies of infectious disease dynamics. Canadian chronological series of pertussis, mumps, measles and rubella, before and after mass vaccination, were analyzed to characterize the statistical structure of those diseases. Despite the fact that those infectious diseases were biologically different, it was found that they were all represented by simple models with the same basic statistical structure. Aside from seasonal effects, the number of new cases was given by the incidence in the previous period and by periodically recurrent random factors. It was also shown that mass vaccination did not change that stochastic dependency. They concluded that the Box-Jenkins methodology identified the collective pattern of the dynamics, but not the specifics of the diseases at the biological individual level.

Cliff and Haggett (1993) reviewed the application of statistical models to the outbreaks of measles epidemic. They looked first at its epidemiological characteristics and assessed the extent to which those either aid or hinder modeling. They then turned to the models that had been developed to simulate geographical spread. A distinction was drawn between process-based and time series models. They provided applications from work, by using Icelandic data. Finally they considered the forecasting potential of the models described.

Greeffell (1992) examined the impact of seasonality and chaotic dynamics in simple models for the population dynamics of measles on the probability of fade-out of infection. Monte Carlo simulations of the seasonally forced SEIR model was used, with parameters appropriate to a city of 1 million people. The incidence of fade-out in a spatially homogeneous model was compared with simple spatial models involving various degrees of coupling between sub-areas. His results indicated a significant degree of fade-out of infection, which was not consistent with previously derived criteria for the persistence of measles. Lowering the degree of spatial coupling did not substantially reduce the extent of fade-out. A simple non-linear analysis of the simulated series was presented, and the epidemiological implications of those results were discussed.

Glass et al. (2003) presented the effect of Heterogeneity in Measles Vaccination on Population Immunity. They adopted a meta-population framework to model local aggregation of populations, and used that to investigate the effects of vaccination

heterogeneity. A recent survey of antibody levels in a community with low vaccination levels in The Netherlands enabled them to assess the relative importance of local and long-range infective contacts. They identified feasible levels of aggregation in the meta-population model. In the aggregated model, they found that heterogeneity in vaccination coverage could lead to a much increased rate of infection among unvaccinated individuals, with a simultaneous drop in the average age at infection.

KNUST

Tarwater and Martin (2001) evaluated the effect of population density on the epidemic outbreak of measles. They used average-number contacts with susceptible individuals per infectious individual as a measure of population density, an analytical model for the distribution of the non-stationary stochastic process of susceptible contact was presented. They used a 5-dimensional lattice simulation model of disease spread to evaluate the effects of four different population densities. They also used a zero-inflated Poisson probability model to quantify the non-stationarity of the contact rate in the stochastic epidemic process. Analysis of the simulation results identified a decrease in a susceptible contact rate from four to three, resulted in a dramatic effect on the distribution of contacts over time, the magnitude of the outbreak, and, ultimately, the spread of disease.

Keeling and Grenfell (2002) revealed that the use of constant infectious and incubation periods, rather than the more convenient exponential forms, had been presented as a simple means of obtaining realistic persistence levels. They considered the persistence of

measles: reconciling theory, simulation and observation. They used a deterministic approach to parameterize a variety of models to fit the observed biennial attractor that determined the level of seasonality by the choice of model. They used ‘best-fit’ parameters to compare fairly the persistence of the stochastic versions of those models. Finally, they considered the differences between the observed fade-out pattern and the more theoretically appealing ‘first passage time’.

KNUST

Lloyd (2000) illustrated how detailed dynamical properties of a model might depend in an important way on the assumptions made in the formulation of the model. According to his study most mathematical models used to understand the dynamical patterns seen in the incidence of childhood viral diseases, such as measles, employ a simple, but epidemiologically unrealistic, description of the infection and recovery process. The inclusion of more realistic descriptions of the recovery process was shown to cause a significant destabilization of the model. When there was seasonal variation in disease transmission that destabilization leads to the appearance of complex dynamical patterns with much lower levels of seasonality than previously predicted.

From Lloyd (2001), most mathematical models used to study the epidemiology of childhood viral diseases, such as measles. He described the period of infectiousness by an exponential distribution. He used Susceptible Infectious Recovered (*SIR*) model in his study. He reviewed that less dispersed distributions were seen to have two important

epidemiological consequences. First, less stable behaviour was seen within the model: incidence patterns became more complex. Second, disease persistence was diminished: in models with a finite population, the minimum population size needed to allow disease persistence increased. The assumption made concerning the infectious period distribution was of a kind routinely made in the formulation of mathematical models in population biology. He detected that a major effect on the central issues of population persistence and dynamics were observed. The results of his study have broad implications for mathematical modelers of a wide range of biological systems.

According to Zaman et al. (2007), almost all mathematical models of diseases start from the same basic premise. The population could be subdivided into a set of distinct classes dependent upon experience with respect to the relevant disease. They used Susceptible Infected Recovered (*SIR*). In their paper, they described an *SIR* epidemic model with three components; *S*, *I* and *R*. They described their study of stability analysis theory to find the equilibria for the model. In order to achieve control of the disease, they considered a control problem relative to the *SIR* model. A percentage of the susceptible populations was vaccinated in that model. They showed that an optimal control exists for the control problem and they used *Runge-Kutta* fourth order procedure to describe the numerical simulations. They finally described a real example showing the efficiency of that optimal control.

London and Yorke (1973) stated that recurrent outbreaks of measles, chickenpox and mumps in cities were studied using a mathematical model of ordinary differential delay equations. They estimated the mean contact rate from the monthly reported cases over a 30- to 35-year period. The mean monthly contact rate for each disease was 1.7 to 2 times higher in the winter months than in the summer months. They showed that the seasonal variation was attributed primarily to the gathering of children in school. Computer simulations that use the seasonally varying contact rates reproduce the observed pattern of un damped recurrent outbreaks: annual outbreaks of chickenpox and mumps and biennial outbreaks of measles. The two-year period of measles outbreaks was the signature of an endemic infectious disease that would exhaust itself and become non-endemic if there were a minor increase in infectivity or a decrease in the length of the incubation period. For populations in which most members were vaccinated, simulations showed that the persistence of the biennial pattern of measles outbreaks implies that the vaccine was not being used uniformly throughout the population.

Liverpool, U.K., 1863--1900, has been used as a model to explore the interaction between measles epidemics and the population dynamics in an overcrowded community with inadequate nutrition. They used a non-linear model that allowed the estimation of certain underlying demographic parameters. Their results were consistent with a system that was driven by an oscillation in the transmission parameter that was compounded of an oscillation in autumn and also by an oscillation in wheat prices (Duncan et al., 1999).

Infectious diseases provide a particularly clear illustration of the spatiotemporal underpinnings of consumer-resource dynamics. The paradigm was provided by extremely contagious, acute, immunizing childhood infections. Partially synchronized, unstable oscillations were punctuated by local extinctions. That, in turn, could result in spatial differentiation in the timing of epidemics and, depending on the nature of spatial contagion, might result in traveling waves. They used the basis of a gravity coupling model and a Time series Susceptible Infected- Recovered (TSIR) model for local dynamics. They proposed a meta-population model for regional measles dynamics. Their model could capture all the major spatiotemporal properties in pre-vaccination epidemics of measles in England and Wales (Yingcun et al., 2004).

Trottier and Philippe (2001) also presented a deterministic modeling as applied to the population dynamics of infectious diseases. They used SEIR deterministic model to provide useful insights into the mechanic of many common childhood diseases such as measles. They showed that deterministic models exhibit damped oscillations, showed random variations and predicted the spread of infectious diseases. Their paper provided an introduction to the theory and methods of deterministic modeling and would be followed by two other articles that would show how sensitivity analysis could be helpful for the forecast and control of common infectious diseases at the population scale.

The objective of Perez and Dragicevic (2009) study was to develop an agent-based modeling approach to integrate geographic information systems (GIS) to simulate the

spread of a communicable disease in an urban environment, as a result of individuals' interactions in a geospatial context. They used measles outbreak in an urban environment as a case study. Their results provided insights into the application of the model to calculate ratios of Susceptible - Infected in specific time frames and urban environments, due to its ability to depict the disease progression based on individuals' interactions. It was demonstrated that the dynamic spatial interactions within the population led to high numbers of exposed individuals who performed stationary activities in areas after they had finished commuting. The sick individuals were concentrated in geographical locations like schools and universities.

Souza (1982) proposed a new approach to forecasting based on the Bayesian principles of information theory and called the Poisson - gamma single - state model. In his paper, a two-state version of the Poisson - gamma model was formulated by considering the uncertainty not only in the parameters but also in the model itself. That model was particularly useful for modeling epidemic data such as measles by considering two different situations of the generating process at each time point.

Thacker and Millar (1991) stated that in 1983, 5 years after the inception of an aggressive national measles elimination strategy, the United States experienced its lowest level of reported numbers of cases of measles. That accomplishment was the result of an effective vaccination strategy coupled with surveillance and control efforts by local, state and national public health agencies. After 1983, however, the reported number of measles

cases slowly increased until 1989 and the number exceeded that of 1979. In 1990, they were experiencing epidemics throughout the United States and expected the reported number of cases of measles to exceed that of 1989. They felt it was timely to reflect on that experience in the light of previous measles control efforts. They looked back to the contributions of Professor George Macdonald, which were critical to the successful elimination of measles from The Gambia in 1969. As they enter the last decade of that century, the sensible merging of mathematics and epidemiology in useful models and the appropriate use of such models for planning, offered the best hope for achieving the elimination of measles either in that or the next century.

Bolker and Griefell (1995) presented a Space, Persistence and Dynamics of Measles Epidemic. Their paper explored the relations between persistence and dynamics in measles epidemics. Most current models, including the stochastic seasonally forced and age-structured models examined in their paper failed to capture simultaneously the observed dynamics and persistent characteristics of epidemics in large urban populations before vaccination. Summary measures of persistence and triennially allowed them to compare epidemics in England, New York and Copenhagen with results of non-spatial and spatial stochastic models. Spatial (meta-population) structure allowed persistence and triennial dynamics to coexist in that class of models. The spatial dynamics of measles, for which detailed spatiotemporal data were available, might serve as a useful test of ideas applicable to other epidemiological and ecological systems with an important spatial component.

Earn et al. (2000) showed that dramatic changes in patterns of epidemics had been observed throughout that century. They reviewed that for childhood infectious diseases such as measles, the major transitions were between regular cycles and irregular, possibly chaotic epidemics and from regionally synchronized oscillations to complex, spatially incoherent epidemics. A simple model could explain both kinds of transitions as the consequences of changes in birth and vaccination rates. Measles was a natural ecological system that exhibits different dynamic transitions at different times and places, yet all of those transitions could be predicted as bifurcations of a single nonlinear model.

Georgette and Allen (2009) researched sought to improve the cost effectiveness of Pre-Outbreak Immunization (POI). That end was achieved through the development of a novel quantification for cost effectiveness, the Morbidity Avoidance Ratio (MAR) that could be widely applied in impoverished nations most affected by vaccine preventable diseases. They used a simulation for disease spread programmed into MS Excel and calculated the MAR for idealized cases of measles, mumps, and rubella. They also determined based upon that analysis that the most cost effective POI rate was the herd immunity threshold. They found that as the POI rate increased, the cost effectiveness increased until the threshold was reached. Their research demonstrated a novel approach to analyzing POI and could help improve the cost effectiveness of outbreak control.

Sattenspiel and Dietz (1995) presented a model for the spread of infectious diseases among discrete geographic regions that incorporate a mobility process that describes how contact occurs between individuals from different regions. They described the general formulation of the mobility process and it was shown that the formulation encapsulates a range of mobility behavior from complete isolation of all regions to permanent migration between regions. They also showed how that mobility process fits into an SIR epidemic model. Their examples included a model for disease transmission in a population with two distinct mobility patterns operating and a model developed to describe a 1984 measles epidemic on the Caribbean island of Dominica.

Stone et al. (2000) based on a theory of population dynamics in perturbed environments. It was hypothesized that measles epidemics could be more efficiently controlled by pulse vaccination. They analyzed the rationale of the pulse vaccination strategy in the simple SIR epidemic model. It was possible to eradicate the measles infection from the entire model population. They derived the conditions for epidemic eradication under various constraints and showed their dependence on the parameters of the epidemic model.

Bauch (2008) presented the Role of Mathematical Models in Explaining Recurrent Outbreaks of Infectious Childhood Diseases. Childhood diseases such as measles were characterized by recurrent outbreaks. Mathematicians had long used models in an effort to better understand and predict those recurrent outbreak patterns. That paper summarized and commented upon those efforts, providing a historical outline of childhood disease models that had been developed since the start of the twentieth century. The paper also

discussed the influence of data analysis techniques, such as spectral analysis, on the understanding and modeling of childhood disease dynamics.

Grais et al. (2006) stated that the current World Health Organization recommendations for response during measles epidemics focus on case management rather than outbreak response vaccination (ORV) campaigns, which may occur too late to impact morbidity and mortality and have a high cost per case prevented. They explored the potential impact of an ORV campaign conducted during the 2003–2004 measles epidemic in Niamey, Niger. They measured the impact of this intervention and also the potential impact of alternative strategies. They used a unique geographical, epidemiologic and demographic dataset collected during the epidemic to develop an individual-based simulation model. They estimated that a median of 7.6% [4.9–8.9] of cases were potentially averted as a result of the outbreak response, which vaccinated approximately 57% (84 563 of an estimated 148 600) of children in the target age range (6–59 months), 23 weeks after the epidemic started. They found that intervening early (up to 60 days after the start of the epidemic) and expanding the age range to all children aged 6 months to 15 years may lead to a much larger (up to 90%) reduction in the number of cases in a West African urban setting like Niamey.

From the revealed analysis above no research work has been extended to model cases of prevalence of measles in the Ashanti Region of Ghana. I therefore, introduced a Statistical model which applies to measles cases in the Ashanti Region of Ghana.

CHAPTER 3

METHODOLOGY

3.0 Introduction

This chapter discusses the methods used to obtain the results presented in this thesis. All the formulae used in the study are discussed thoroughly in this chapter. The Box–Jenkins method for identifying a plausible ARIMA model is given in this chapter, along with techniques for parameter estimation and forecasting for these models. A partial theoretical justification of the use of ARMA models is also discussed in this chapter.

3.1 Definitions of Time Series

A time series is a collection of observations of well-defined data items obtained through repeated measurements over time. For example, measuring the value of retail sales each month of the year for a particular product in a particular geographical location results in a time series. This is because sales revenue is well defined, and consistently measured at equally-spaced intervals. Data collected irregularly or only once are not time series. Time series is therefore, defined mathematically as a time dependent sequence

$X_1, X_2, X_3, \dots, X_N$ or $\{X_t\}$, $t \in \{1, 2, 3, \dots, N\}$, where $1, 2, 3, \dots, N$ depicts time steps and assumed to be equally spaced (Cryer and Kung, 2008).

Time Series can be classified into: deterministic time series and stochastic time series.

- Time series which can be expressed as a known function, such as $X_t = f(t)$, is said to be deterministic time series.

The sequence of random variables $\{X_t : t = 0, \pm 1, \pm 2, \pm 3, \dots\}$ is called a stochastic process and serves as a model for an observed time series. Time series is said to be stochastic time series if it can be expressed as $X_t = X(t)$, where X is a random variable.

Here the mean function is defined by $\mu_t = E(X_t)$ for $t = 0, \pm 1, \pm 2, \pm 3, \dots$

(Cryer and Kung, 2008).

3.2 Objective of Time Series Analysis

There are several possible objectives in analyzing a time series. These objectives may be classified as description, explanation, prediction and control. Below, we briefly describe each of them sequentially.

3.2-1 Description

For time series, the most obvious graphical form is a time plot in which the data are plotted over time. A time plot immediately reveals any trends over time, any regular seasonal behavior, and other systematic features of the data. These need to be identified so they can be incorporated into the Statistical model. Apart from trend and seasonal variation, the outlier to look for in the graph of the time series is the possible presence of turning point, where for example, a downward trend suddenly changes to an upward trend. An important step in selecting an appropriate forecasting method is to consider the types of data patterns, so that the methods most appropriate to those patterns can be utilized. Four types of time series data patterns can be identified in the literature: horizontal, seasonal, cyclical and trend (Gottman, 1981).

3.2-2 Explanation

When observations are taken on two or more variables, it may be possible to use the variation in one time series variable to explain the variation in the other time series variable. This may lead to a deeper understanding of the mechanism which generated a given time series (Gottman, 1981).

3.2-3 Prediction

When an observed time series is given, one may want to predict the future values of the series. This is an important task in sales forecasting and in the analysis of economic and industrial time series. Prediction is closely related to control problem in many situations. For example, if one can predict that measles epidemic in the Ashanti Region of Ghana is going to increase, then appropriate corrective measures can be taken ahead of time to confront the situation (Gottman, 1981)

3.2-4 Control

When a time series is generated which measures the quality of a manufacturing process, the aim of the analysis may be to control the process. Control procedures are of several different kinds. In statistical quality control, for instance, the observations are plotted on control charts and the controller takes action as a result of studying the charts. Box and Jenkins have described a more sophisticated control strategy which is based on fitting a stochastic model to the series, from which future values of the series are predicted. The values of process variables predicted by the model are taken as target values and the variables conform to the target values (Gottman, 1981).

3.3 Components of a Time Series

The four components of time series are secular trend, seasonal variation, cyclical variation and irregular variation. We describe these briefly below.

3.3-1 Secular Trend

A time series data may show upward trend or downward trend for a period of years and this may be due to factors such as increase in population, change in technological progress, large scale shift in consumer demands, etc. For example, population increases over a period of time, price increases over a period of years, production of goods on the capital market of the country increases over a period of years. These are the examples of upward trend. The sales of a commodity may decrease over a period of time because better products coming have been released into the market. This is an example of declining trend or downward trend. The increase or decrease in the movements of a time series is called Secular trend (Blog, 2008).

3.3-2 Seasonal Variation

Seasonal variations are short-term fluctuation in a time series which occur periodically in a year. This continues to repeat year after year. The major factors that are responsible for the repetitive pattern of seasonal variations are weather conditions and customs of people. More woolen clothes are sold in winter than in the season of summer. Regardless of the trend we can observe that in each year more ice creams are sold in summer and very little in winter season. The sales in the departmental stores are more during festive seasons than in the normal days. In general seasonality is defined as a pattern that repeats itself over fixed intervals of time (Blog, 2008).

3.3-3 Cyclical Variations

Cyclical variations are recurrent upward or downward movements in a time series but the period of cycle is greater than a year. Also, these variations are not regular as seasonal variation. There are different types of cycles of varying in length and size. The ups and downs in business activities are the effects of cyclical variation. A business cycle showing these oscillatory movements has to pass through the four phases: prosperity, recession, depression and recovery. In a business, these four phases are completed by passing one to another in this order (Blog, 2008).

3.3-4 Irregular Variation

Irregular variations are fluctuations in time series that are short in duration, erratic in nature and follow no regularity in the occurrence pattern. These variations are also referred to as residual variations since, by definition, they represent what is left out in a time series after trend, cyclical and seasonal variations. Irregular fluctuations results due to the occurrence of unforeseen events such as floods, earthquakes, wars, famines, etc. (Blog, 2008).

3.4-1 Stationary Time Series

Stationarity means that there is no growth or decline in the data. The data fluctuate around a constant mean, independent of time, and the variance of the fluctuation remains essentially constant over time. A time series is said to be strictly stationary if the joint distribution of $X_{t_1}, X_{t_2}, \dots, X_{t_n}$ is the same as the joint distribution of $X_{t_1+T}, X_{t_2+T}, \dots, X_{t_n+T}$ for all $t_{1+T}, t_{2+T}, \dots, t_{n+T}$. We can usually assess stationarity using a time plot (Cryer, 2008).

3.4-2 Achieving Stationarity (Differencing)

It is important to remove the non-stationarity in data to be analyzed so that other correlation structure can be seen before proceeding with time series model building. If there is a trend in the mean then differencing the time series data will remove the trend and stationarity will be achieved. For non-seasonal data, first differencing is usually sufficient to attain stationarity.

$$\text{The first difference is denoted as } \nabla X_t = X_t - X_{t-1} \quad (3.1)$$

$$\text{For second differencing we have } \nabla^2 X_t = \nabla(\nabla X_t) = X_t - 2X_{t-1} + X_{t-2} \quad (3.2)$$

From the above definitions, we can see that the first difference eliminates a linear trend, the second difference eliminates a quadratic trend and so on (Shumway and Stoffer, 2006).

We define the *backshift operator* as $BX_t = X_{t-1}$ and extend it to powers

$$B^2 X_t = B(BX_t) = B(X_{t-1}) = X_{t-2},$$

$$B^3 X_t = X_{t-3}$$

.

$$B^k X_t = X_{t-k}$$

We may then rewrite $\nabla X_t = X_t - X_{t-1}$ as $\nabla X_t = (1-B)X_t$ and the notion can be extended further as follows.

The second difference becomes:

$$\begin{aligned} \nabla^2 X_t &= (1-B)^2 X_t = (1-2B+B^2)X_t \\ &= X_t - 2BX_t + B^2 X_t \end{aligned} \quad (3.3)$$

Differences of order d are defined as $\nabla^d = (1 - B)^d$ where the operator $(1 - B)^d$ may be expanded algebraically to evaluate for higher integer values of d . When $d = 1$, we drop it from the notation (Shumway and Stoffer, 2006).

3.5 Autocorrelation and Partial Autocorrelation Functions

3.5-1 Autocorrelation Function (ACF)

The autocorrelation function measures the degree of correlation between neighboring observations in a time series.

The autocorrelation function (ACF), $\rho_{t,s}$, is given by the formula

$$\rho_{t,s} = \text{Corr}(X_t, X_s) = \frac{\text{Cov}(X_t, X_s)}{\sqrt{\text{Var}(X_t)\text{Var}(X_s)}} \quad \forall t, s \in \{0, \pm 1, \pm 2, \pm 3, \dots\} \quad (3.4)$$

where $\text{Cov}(X_t, X_s) = E[(X_t - \mu_t)(X_s - \mu_s)] = E(X_t, X_s) - \mu_t \mu_s$

The autocorrelation coefficient estimated from sample observations at lag k is

$$\gamma_k = \frac{\sum_{t=k+1}^n (X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{t=1}^n (X_t - \bar{X})^2} \quad (\text{Spyros et al., 1998}). \quad (3.5)$$

3.5-2 The Sample Distribution of Autocorrelation Function Coefficients

The autocorrelation coefficients of a random data are approximately normal with

mean $\mu_{\rho_k} = 0$ and $\sigma_{\rho_k} = \frac{1}{\sqrt{n}}$, where n is the size of the sample (Cryer and Kung, 2008).

3.5-3 Partial Autocorrelation Function (PACF)

Partial autocorrelation function (PACF) measures the degree of association between X_t and X_{t+k} when the effects of other time lags on X are held constant. The partial autocorrelation function (PACF) at lag k is then defined to be the correlation between the prediction errors: that is,

$$\phi_{kk} = \text{Corr}(X_t - \beta_1 X_{t-1} - \beta_2 X_{t-2} - \dots - \beta_{k-1} X_{t-k+1}, X_{t-k} - \beta_1 X_{t-k+1} - \beta_2 X_{t-k+2} - \dots - \beta_{k-1} X_{t-1})$$

$$= \frac{|P_k^*|}{|P|}, \text{ where } P_k \text{ is the } k \times k \text{ autocorrelation matrix, and } P_k^* \text{ is } P_k \text{ with the last}$$

column replaced by $[\rho_1, \rho_2, \dots, \rho_k]^T$ and

$$P_k = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \dots & 1 \end{bmatrix} \quad (3.8)$$

For convenience, we take $\phi_{11} = 1$ and obtain $\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$ (3.9)

3.5-5 White Noise

A very important example of a stationary process is the so-called white noise process, which is defined as a sequence of independent, identically-distributed random variables $\{X_t\}$. Its importance stems not from the fact that it is an increasing model itself, but from the fact that many useful processes can be constructed from white noise.

A stationary time series analysis for which X_t and X_{t+k} are uncorrelated is called “White Noise”, where $k = 1, 2, 3, \dots$. Such a process will have autocorrelation function r_k

$$r_k = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases}, \quad (3.10)$$

A white noise process is sometimes called a purely random process (Nagpaul, 2005).

3.6 Autoregressive Integrated Moving Average Models (ARIMA)

The autoregressive integrated moving average (ARIMA) models have assumed great importance in modeling real-world processes. The introduction of correlation as a phenomenon that may be generated through lagged linear relations leads to the autoregressive (AR) and autoregressive moving average (ARMA) models.

3.6-1 An Autoregressive Model of Order p , AR (p)

An autoregressive model of order p , denoted by AR (p), is of the form

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \dots + \phi_p X_{t-p} + e_t \quad (3.11)$$

where X_t is stationary random process/variable, $\phi_1, \phi_2, \phi_3, \dots, \phi_p$ are parameters or constants ($\phi_p \neq 0$).

Unless otherwise stated, we assume that e_t is a Gaussian white noise series with mean zero and variance σ_e^2 . The mean of X_t in equation (3.11) is zero.

If the mean, μ , of X_t is not zero, then we replace X_t by $X_t - \mu$ in equation (3.11).

That is $X_t - \mu = \phi_1 (X_{t-1} - \mu) + \phi_2 (X_{t-2} - \mu) + \phi_3 (X_{t-3} - \mu) + \dots + \phi_p (X_{t-p} - \mu) + e_t$

$$\begin{aligned}
X_t &= \alpha + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \dots + \phi_p X_{t-p} + e_t \\
&= \sum_{k=1}^p \phi_k X_{t-k} + \alpha + e_t
\end{aligned} \tag{3.12}$$

where $\alpha = \mu(1 - \phi_1 - \phi_2 - \phi_3 - \dots - \phi_p)$ (Shumway and Stoffer, 2006).

The order of an AR(p) process is determined by the partial autocorrelation function (PACF). An AR (p) process has its PACF cutting off after lag p and the ACF decays.

KNUST

3.6-2 Autoregressive Process of Order 2 Denoted by AR(2)

Assume the series is stationary then autoregressive process of second order is given by

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + e_t \tag{3.13}$$

3.6-3 Moving Average Model of Order q , MA(q)

As an alternative to the autoregressive representation in which the X_t on the left-hand side of the equation are assumed to be combined linearly, the moving average model of order q , denoted by MA(q), assumes the white noise ω_t on the right-hand side of the defining equation are combined linearly to form the observed data. The moving average model of order q , denoted by MA (q) model, is defined as

$$X_t = \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \theta_3 \omega_{t-3} + \dots + \theta_q \omega_{t-q} + \omega_t \tag{3.14}$$

Where there are q lags in the moving average and $\theta_1, \theta_2, \theta_3, \dots, \theta_q$ ($\theta_q \neq 0$) are parameters. The noise ω_t is assumed to be Gaussian white noise with mean zero and variance σ_ω^2 (Shumway and Stoffer, 2006).

3.6-4 Autoregressive Moving Average Model (ARMA)

We now proceed with the general development of autoregressive, moving average, and mixed autoregressive moving average (ARMA), models for stationary time series.

A time series $\{X_t; t = 0, \pm 1, \pm 2, \pm 3, \dots\}$ is ARMA(p, q) if it is stationary and

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_{t-p} + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \dots + \theta_q \omega_{t-q} + \omega_t \quad (3.15)$$

with $\phi_p \neq 0$, $\theta_q \neq 0$ and $\rho_\omega^2 > 0$. The parameters p and q are called the autoregressive and the moving average orders, respectively. If X_t has a non-zero mean μ , we set $\alpha = \mu(1 - \phi_1 - \phi_2 - \dots - \phi_p)$ and write the model as

$$X_t = \alpha + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_{t-p} + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \dots + \theta_q \omega_{t-q} + \omega_t \quad (3.16)$$

$\{\omega_t; t = 0, \pm 1, \pm 2, \dots\}$ is a Gaussian white noise sequence (Shumway and Stoffer, 2006).

3.6-5 ARMA(2, 1) Model

An example of an ARMA(p, q) model is the ARMA(2, 1) model given by

$$X_t = \alpha + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \theta_1 \omega_{t-1} + \omega_t \quad (3.17)$$

The ARMA(2, 1) model is stationary if $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$ and $|\phi_2| < 1$, and it is invertible if $-1 < \theta_1 < 1$. In an ARMA(2, 1) model ACF cuts off after lag 1 and the PACF cuts off after lag 2.

3.7 The AR and MA Polynomials

The AR polynomial is defined as $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p$, $\phi_p \neq 0$ and the MA polynomial is defined as $\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q$, where $\theta_q \neq 0$ and z is a complex number. To address the first problem, we will henceforth refer to an ARMA(p, q) model to mean that it is in its simplest form. That is, in addition to the original definition given in equation (3.18).

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_{t-p} + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \dots + \theta_q \omega_{t-q} + \omega_t, \quad (3.18)$$

We will also require that $\phi(z)$ and $\theta(z)$ have no common factors. So, the process,

$X_t = 0.7X_{t-1} - 0.7\omega_{t-1} + \omega_t$ is not referred to as an ARMA(1, 1) process because, in its reduced form, X_t is white noise. To address the problem of future-dependent models, we formally introduce the concept of causality and Invertibility (Shumway and Stoffer, 2006).

3.8 Causality and Invertibility of an ARMA(p, q) Model

An ARMA(p, q) model is causal and stationary if and only if $\phi(z) \neq 0$ for $|z| \leq 1$.

In other words, an ARMA process is **causal** and stationary only when the roots of $\phi(z)$ lie outside the unit circle; that is, $\phi(z) = 0$ only when $|z| > 1$.

An ARMA (p, q) model is invertible if and only if $\theta(z) \neq 0$ for $|z| \leq 1$. In other words, an ARMA process is invertible only when the roots of $\theta(z)$ lie outside the unit circle; that is, $\theta(z) = 0$ only when $|z| > 1$ (Shumway and Stoffer, 2006).

Table 3.1: Causality and Invertibility Conditions of Specific Time Series Model

ARMA Model	Causality Condition	Invertibility Condition	ACF Coefficient	PACF Coefficient
(1, d, 0)	$-1 < \phi_1 < 1$	None	Tails off	Cuts off After lag 1
(2, d, 0)	$\phi_1 + \phi_2 < 1$ $\phi_2 - \phi_1 < 1$ $ \phi_2 < 1$	None	Tails off	Cuts off After lag 2
(0, d, 1)	None	$-1 < \theta_1 < 1$	Cuts off After lag 1	Tails off
(0, d, 2)	None	$\theta_1 + \theta_2 > -1$ $\theta_2 - \theta_1 > -1$ $ \theta_2 < 1$	Cuts off After lag 2	Tails off
(2, d, 1)	$\phi_1 + \phi_2 < 1$ $\phi_2 - \phi_1 < 1$ $ \phi_2 < 1$	$-1 < \theta_1 < 1$	Cuts off After lag 1	Cuts off After lag 2

Table 3.1 depicts the Causality and Invertibility conditions of specific time series models and the behavior of their theoretical ACF and PACF functions.

3.9 Estimating the Parameters of an ARMA Model

The process for estimating the parameters of the ARMA model is like the one for the MA model, it is an iterative method. Like the MA the residual sum of squares is calculated at every point on a suitable grid of the parameter values, and the values which give the minimum sum of squares are the estimates. For an ARMA(1, 1) the model is given by

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \theta_1\omega_{t-1} + \omega_t \quad (3.19)$$

Given N observations $X_1, X_2, X_3, \dots, X_N$, we guess values for μ, ϕ_1, θ_1 , set $\omega_0 = 0$

and $X_0 = \mu$ and then calculate the residuals recursively by

$$\omega_1 = X_1 - \mu$$

$$\omega_2 = X_2 - \mu - \phi_1(X_1 - \mu) - \theta_1\omega_1$$

:

$$\omega_N = X_N - \mu - \phi_1(X_{N-1} - \mu) - \theta_1\omega_{N-1}$$

It follows that the residual sum of squares $\sum_{t=1}^N \omega_t^2$ is calculated. The other values of μ, ϕ_1

and θ_1 are tried until the minimum residual sum of squares is found (Cryer and Kung, 2008).

3.10 The Autoregressive Integrated Moving Average Model (ARIMA)

A time series $\{X_t\}$ is said to follow an **integrated autoregressive moving average** model, if the d^{th} difference $W_t = \nabla^d X_t = (1-B)^d X_t$ is a stationary ARMA process. If $\{W_t\}$ follows an ARMA(p, q) model, we say that $\{X_t\}$ is an ARIMA(p, d, q) process. Fortunately, for practical purposes, we can usually take $d = 1$ or at most 2. Consider then an ARIMA($p, 1, q$) process. With $W_t = X_t - X_{t-1}$, we have

$$\begin{aligned} W_t &= \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} + e_t \\ \Rightarrow W_t &= \sum_{i=1}^p \phi_i W_{t-i} - \sum_{j=1}^q \theta_j e_{t-j} + \mu + e_t \end{aligned} \quad (3.20)$$

3.10-1 ARIMA(1, 1, 1) Process

An example of ARIMA(p, d, q) is the ARIMA(1, 1, 1) which has one autoregressive (AR) parameter, one level of differencing and one moving average (MA) parameter is given by

$$W_t = \phi_1 X_{t-1} + \theta_1 \omega_{t-1} + \mu + \omega_t$$

$$\Rightarrow (1-B)X_t = \phi_1(1-B)X_{t-1} + \theta_1 \omega_{t-1} + \mu + \omega_t$$

which can be simplified as $X_t - X_{t-1} = \phi_1 X_{t-1} - \phi_1 X_{t-2} + \theta_1 \omega_{t-1} + \mu + \omega_t$

$$\Rightarrow X_t - X_{t-1} = \phi_1(X_{t-1} - X_{t-2}) + \theta_1 \omega_{t-1} + \mu + \omega_t \quad (3.21)$$

3.11 Seasonal Autoregressive Integrated Moving Average Model

For multiplicative Seasonal Autoregressive Integrated Moving Average (SARIMA) model, we have the general notation $(p, d, q), (P, D, Q)^S$ where (p, d, q) is the non-seasonal part and $(P, D, Q)^S$ is the seasonal part with p, d, q having their usual meaning and P is the order of the seasonal AR process. D is the differencing of the seasonal process, Q is the order of seasonal MA process of the time series, and S is the order of seasonality. For the purpose of identifying a seasonal ARMA process, we divide the process into two parts. To identify the seasonal pattern, we ignore the non-seasonal process and determine whether the seasonality is determined by an AR or an MA process by focusing on the coefficients of the seasonal terms. Suppose that the non-seasonal part is an ARIMA(1, 0, 1) and the time series shows a yearly seasonal pattern, then the complete model becomes

$$(1 - \phi_1 \beta)(1 - \phi_{12} \beta^{12})X_t = (1 - \theta_1 \beta)\omega_t, \text{ if seasonality is on the AR portion, and}$$

$$(1 - \phi_1 \beta)X_t = (1 - \theta_1 \beta)(1 - \theta_{12} \beta^{12})\omega_t, \text{ if seasonality is on the MA portion, where}$$

$$(1 - \phi_{12} \beta^{12})X_t = X_t - \theta_{12} \beta^{12} X_t$$

It can also be written as $(1 - \theta_{12}\beta^{12})\omega_t = \omega_t - \theta_{12}\beta^{12}\omega_t = \omega_t - \theta_4\omega_{t-4}$

An MA (2) seasonal process with one level of differencing is expressed by

$$(1 - \beta)(1 - \beta_{12})X_t = (1 - \theta_1\beta - \theta_2\beta^2)\omega_t$$

One level of differencing seasonal pattern in AR = MA (2) non-seasonal. OR

$$(1 - \beta)X_t = (1 - \theta_1\beta - \theta_2\beta^2)\omega_t \quad (1 - \theta_{12}\beta^{12})\omega_t \text{ (Shumway and Stoffer, 2006).}$$

Table 3.2: Causality and Invertibility conditions of Purely Seasonal Time Series Model

ARMA Model	Causality Condition	Invertibility Condition	ACF Coefficient	PACF Coefficient
$(1, D, 0)^s$	$-1 < \phi_1 < 1$	None	Tails off	Cuts off after one seasonal lag
$(2, D, 0)^s$	$\phi_1 + \phi_2 < 1$ $\phi_2 - \phi_1 < 1$ $ \phi_2 < 1$	None	Tails off	Cuts off after two seasonal lags
$(0, D, 1)^s$	None	$-1 < \theta_1 < 1$	Cuts off after one seasonal lag	Tails off
$(0, D, 2)^s$	None	$\theta_1 + \theta_2 > -1$ $\theta_2 - \theta_1 > -1$ $ \theta_2 < 1$	Cuts off after two seasonal lags	Tails off
$(2, D, 1)^s$	$\phi_1 + \phi_2 < 1$ $\phi_2 - \phi_1 < 1$ $ \phi_2 < 1$	$-1 < \theta_1 < 1$	Cuts off after one seasonal lag	Cuts off after two seasonal lags

Table 3.2 depicts the Causality and Invertibility conditions of specific purely seasonal time series models and the behavior of their theoretical ACF and PACF functions.

3.11-1 Purely Seasonal Model

A purely seasonal time series is one that has only seasonal AR or MA parameters.

Seasonal autoregressive models are built with parameter called seasonal autoregressive

(SAR) parameters. The SAR parameters represent autoregressive relationships that exist between time series data separated by multiples of the number of periods per season. For example, a model with one SAR parameter is written as

$$X_t = \phi_s X_{t-s} + \omega_t \quad (3.22)$$

That is $ARIMA(P, D, Q)^S = ARIMA(1, 0, 0)^S$

where S is the number of periods per season. The parameter is called the SAR parameter with order s . A general seasonal autoregressive model with P SAR parameters is written

as follows:
$$X_t = \sum_{i=1}^p \phi_{is} X_{t-is} + \omega_t \quad (3.23)$$

where X_{t-s} is order s , X_{t-2s} is of order $2s$, .., and X_{t-ps} is of order ps (Spyros et al., 1998)

3.11-2 Seasonal Moving Average Models

The Seasonal Moving Average (SMA) Model with Q parameters is given by

$$X_t = \sum_{i=1}^q \phi_{is} e_{t-is} + e_t \quad (3.24)$$

Considering $ARIMA(0, 0, 1)^4$, the model is a quarterly seasonal moving average of order one, that is it has one seasonal moving average parameter. A model with one seasonal

moving average parameter is written as
$$X_t = \phi_s e_{t-s} + e_t \quad (3.25)$$

3.12-3 Mixed SAR and SMA Models

A mixed Seasonal Autoregressive (SAR) and Seasonal Moving Average (SMA) model is

given by;
$$X_t = \sum_{i=1}^p \phi_{is} X_{t-is} + \sum_{i=1}^q \theta_{is} \omega_{t-is} + \omega_t \quad (3.26)$$

The order of the seasonal ARMA model is expressed in terms of both PS and QS .

3.13 The Box–Jenkins Methodology for ARIMA Models

The basis of the Box-Jenkins approach to modeling time series is summarized in Figure 3.1 and consists of three phases: identification, estimation and testing, and application (Spyros et al., 1998).

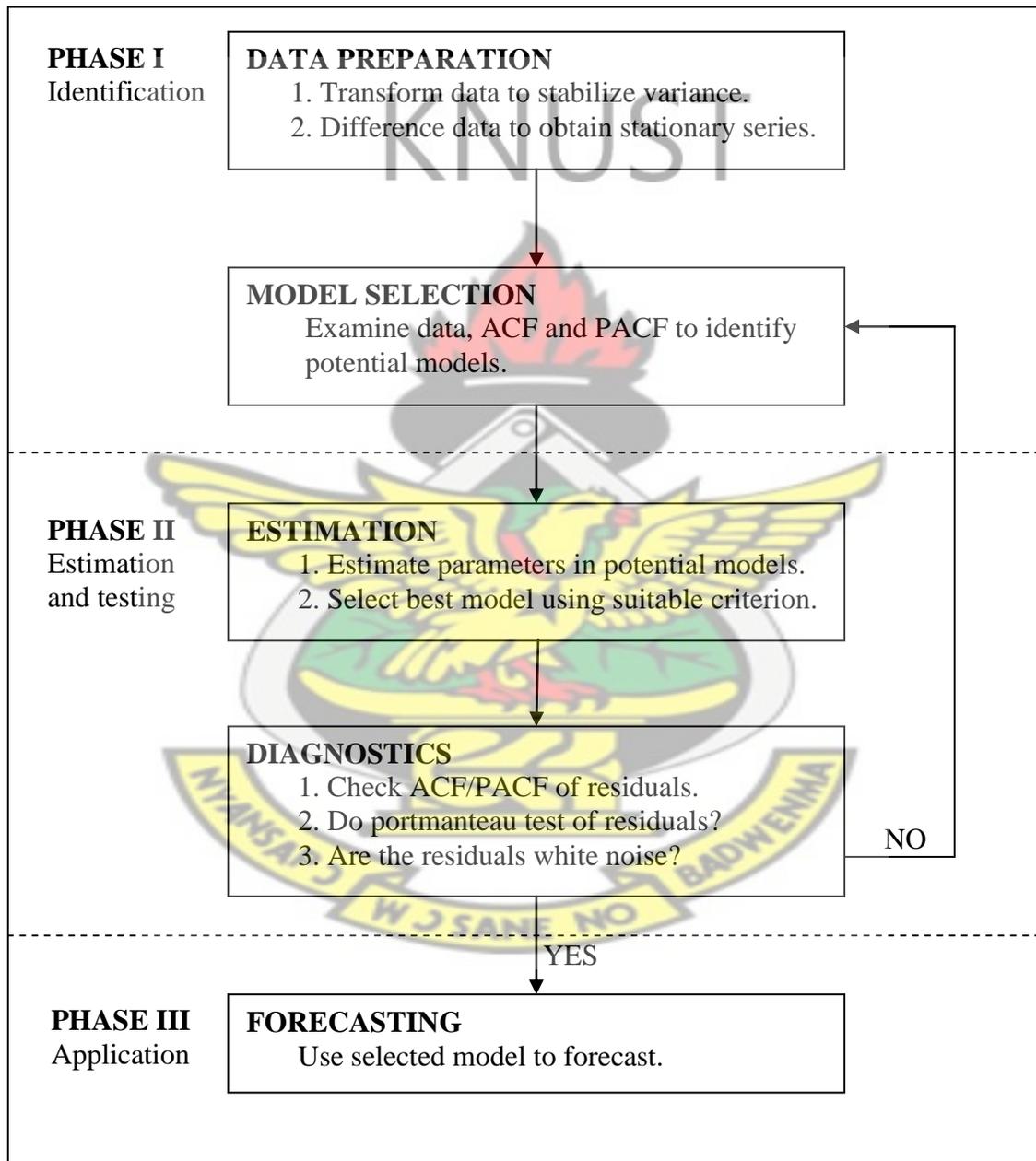


Figure 3.1: Schematic Representation of the Box-Jenkins Methodology for Time Series Modeling (Spyros et al., 1998).

Box and Jenkins effectively put together in a comprehensive manner the relevant information required to understand and use univariate time series ARIMA models. The theoretical underpinnings described by Box and Jenkins and later by Box, Jenkins, and Reinsel (1994) are quite sophisticated, but it is possible for the non-specialist to get a clear understanding of the essence of ARIMA methodology. Application of a general class of forecasting methods involves two basic tasks:

- (a) analysis of the data series and
- (b) selection of the forecasting model that best fits the data series.

Thus, in using a smoothing method, analysis of the data series for seasonality, aids in selection of a specific smoothing method that can handle the seasonality.

(Spyros et al., 1998)

3.13 Other Specification Methods

A number of other approaches to model specification have been proposed since Box and Jenkins' seminal work. Some of the most studied are Akaike's (1974) Information Criterion (AIC), AICc and Bayesian Information Criterion (BIC). Others include the Augmented Dickey Fuller (ADF) unit root test and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Tests. We describe these briefly below sequentially.

Augmented Dickey-Fuller (ADF) Test

The hypotheses $H_0: X_t$ is non – Stationary and $H_1: X_t$ is Stationary can be tested in

$$\text{the regression equation } \Delta X_t = \beta_0 + \alpha t + \beta_1 X_{t-1} + \sum_{i=1}^p \gamma_i \Delta X_{t-i} + \varepsilon_t \quad (3.27)$$

Accept H_0 if $P - \text{value} > 0.05$, else accept H_1 .

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test

An alternative approach to the ADF test is the KPSS test. A hypotheses of $H_0: X_t$ is level or trend stationary is tested against $H_1: X_t$ is non- stationary in the regression equation in (3.28).

$$X_t = \alpha_t + \beta t + \mu_t \quad (3.28)$$

where a random walk, $\alpha_t = \alpha_{t-1} + \varepsilon_t$ is allowed.

Accept H_0 if $P\text{-value} > 0.05$, else accept H_1 .

3.13-1 The Akaike's Information Criteria (AIC)

The Akaike's Information Criteria (AIC) provides a measure of the goodness-of-fit of a model which takes into account the number of terms in the model. It is commonly used with ARIMA models to determine the appropriate model order. The AIC is equal to twice the number of parameters in the model minus twice the logarithm of the likelihood function. Mathematically, AIC is calculated as follows:

$$AIC(p, q) = 2k - 2\log(\text{Maximum Likelihood}) \quad (3.29)$$

where $k = p + q + 1$ if the model contains an intercept or constant term and $k = p + q$ otherwise. Given two or more competing models, the one with the smaller AIC value will be deemed more appropriate (Spyros et al., 1998).

Thus the AIC is a biased estimator, and the bias can be appreciable for large parameter per data ratios. Hurvich and Tsai showed that the bias can be approximately eliminated

by adding another non-stochastic penalty term to the AIC, resulting in the corrected AIC, denoted by AICc and defined by the formula

$$AICc = AIC + \frac{2(k+1)(k+2)}{n-k-2} \quad (3.30)$$

Here n is the (effective) sample size and again k is the total number of parameters as above excluding the noise variance. Simulation results by Hurvich and Tsai suggest that for cases with k/n greater than 10%, the AICc outperforms many other models selection criteria, including both the AIC and BIC (Spyros et al., 1998).

3.13-2 The Schwartz's Bayesian Information Criteria (BIC)

Another approach to determining the ARMA orders is to select a model that minimizes the Schwarz Bayesian Information Criterion (BIC). BIC like the AIC, the BIC is an order selection criteria for ARIMA models. It was invented by Schwarz and sometimes leads to less complex models than AIC. It is defined mathematically as

$$BIC(p, q) = k \times \log(n) - 2 \times \log(\text{Maximum Likelihood}) \quad (3.31)$$

If the true process follows an ARMA (p, q) model, then it is known that the orders specified by minimizing the BIC are consistent; that is, they approach the true orders as the sample size increases. However, if the true process is not a finite-order ARMA process, then minimizing AIC among an increasingly large class of ARMA models enjoys the appealing property that it will lead to an optimal ARMA model that is closest

to the true process among the class of models under study. The BIC imposes a greater penalty for the number of estimated model parameters than does 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 (Spyros et al., 1998).

3.13-3 Estimation of the Parameters of the Tentative Models

Once a model is identified, the next stage of the Box-Jenkins approach is to estimate the parameters. In this study, all the coefficients of the estimated parameters were done using the R statistical software package.

3.13-4 Testing the Model for Adequacy (Portmanteau Test)

After identifying an appropriate model for a time series data, it is very important to check that the model is adequate. The error terms e_t are examined and for the model to be adequate, the errors should be random. Ljung and Box provided a modified portmanteau test statistic for checking the randomness of the error terms. Their statistic is given by equation (3.32),

$$Q^* = n(n+2) \times \sum_{k=1}^h \left(\frac{r_k^2}{n-k} \right) \quad (3.32)$$

which is approximately distributed as a χ^2 with $h-p-q$ degrees of freedom, where n is the length of the time series, h is the first h autocorrelations being checked, p is the order of the AR process and q is the order of the MA process and r is the estimated autocorrelation coefficient of the k^{th} residual term. If the calculated value of Q^* is

greater than χ^2 for $h-p-q$ degrees of freedom, then the model is considered inadequate and the model is adequate if Q^* calculated is less than χ^2 for $h-p-q$ degrees of freedom. If the model is tested inadequate, then the forecaster should select an alternative model and test for the adequacy of the model (Spyros et al., 1998).

3.14 Measuring Forecasting Accuracy (Error Metrics)

Mean Absolute Error (MAE)

The Mean Absolute Error (MAE) is defined mathematically by equation (3.33).

$$MAE = \frac{1}{n} \sum_{t=1}^n |e_t| \quad (3.33)$$

where X_t is the actual observation for time period t , F_t is the forecast value for the same period and $e_t = X_t - F_t$ is the error term and n is the number of forecasting values (Spyros et al., 1998).

Mean Square Error (MSE)

Symbolically, the Mean Square Error (MSE) is defined by equation (3.34)

$$MSE = \frac{1}{n} \sum_{t=1}^n e_t^2 \quad (3.34)$$

where X_t is the actual observation for time period t , F_t is the forecast value for the same period and $e_t = X_t - F_t$ is the error term and n is the number of forecasting values (Spyros et al., 1998).

3.15 The Diebold-Mariano Statistic for Comparing Predictive Accuracy

Let $\{X_t\}$ denote the series to be forecast and let $X^1_{t+h/t}$ and $X^2_{t+h/t}$ denote two competing forecasts of $X_{t+h/t}$ based on I_t .

For example, $X^1_{t+h/t}$ could be computed from an AR(p) model and $X^2_{t+h/t}$ could be computed from an ARMA(p, q) model. The forecast errors from the two models are

$$\begin{aligned}\mathcal{E}^1_{t+h/t} &= X_{t+h} - X^1_{t+h/t} \\ \mathcal{E}^2_{t+h/t} &= X_{t+h} - X^2_{t+h/t}\end{aligned}$$

The h -step forecasts are assumed to be computed for $t = t_0, \dots, T$ for a total of T_0 forecasts giving $\{\mathcal{E}^1_{t+h/t}\}_{t_0}^T, \{\mathcal{E}^2_{t+h/t}\}_{t_0}^T$.

Because the h -step forecasts use overlapping data the forecast errors in $\{\mathcal{E}^1_{t+h/t}\}_{t_0}^T$ and $\{\mathcal{E}^2_{t+h/t}\}_{t_0}^T$ will be serially correlated (Zivot, 2004).

The accuracy of each forecast is measured by a particular loss function

$$L(X_{t+h}, X^i_{t+h/t}) = L(\mathcal{E}^i_{t+h/t}), \quad i = 1, 2. \text{ Some popular loss functions are}$$

- Squared error loss: $L(\mathcal{E}^i_{t+h/t}) = L(\mathcal{E}^i_{t+h/t})^2$
- Absolute error loss: $L(\mathcal{E}^i_{t+h/t}) = |\mathcal{E}^i_{t+h/t}|$

To determine if one model predicts better than another we may test null hypotheses

$$H_0 : E[L(\mathcal{E}^1_{t+h/t})] = E[L(\mathcal{E}^2_{t+h/t})]$$

against the alternative

$$H_1 : E[L(\mathcal{E}^1_{t+h/t})] \neq E[L(\mathcal{E}^2_{t+h/t})] \quad (\text{Zivot, 2004})$$

Accept H_0 if $P\text{-value} > 0.05$, else accept H_1 .

CHAPTER 4

ANALYSIS AND RESULTS

4.0 Introduction

This chapter discusses the data analysis and modeling the data collected on measles cases in the Ashanti Region's time series. Here we applied the **R** Statistical Package in modeling the time series. That is, all the plots and numerical output displayed in this thesis have been produced with the **R** software. Most of the numerical outputs have been edited for additional clarity or for simplicity. Actual measles data drawn from various hospitals in the Ashanti Region of Ghana are used throughout in this chapter to illustrate the methodology explained in chapter 3.

4.1 Time Plot of Prevalence of Measles in the Ashanti Region of Ghana

Figure 4.1 displays the time series plot of the measles data from January 2001 to November 2011. It can be observed that measles increased sharply by a large amount from January to April 2001, but decreased from May to December 2002. A gradual increase in measles was observed from January to March 2003 followed by an irregular pattern in measles (that is upward and downward trends of the measles cases). In general, the trend in measles prevalence in the Ashanti Region of Ghana seems to be decreasing, but not always the case. The annual measles time plot in Figure 4.1 does not exhibit seasonal variation, and it is non-stationary due to the trend component. Looking at the end of Figure 4.1, decrease in measles was observed from September to November 2011.

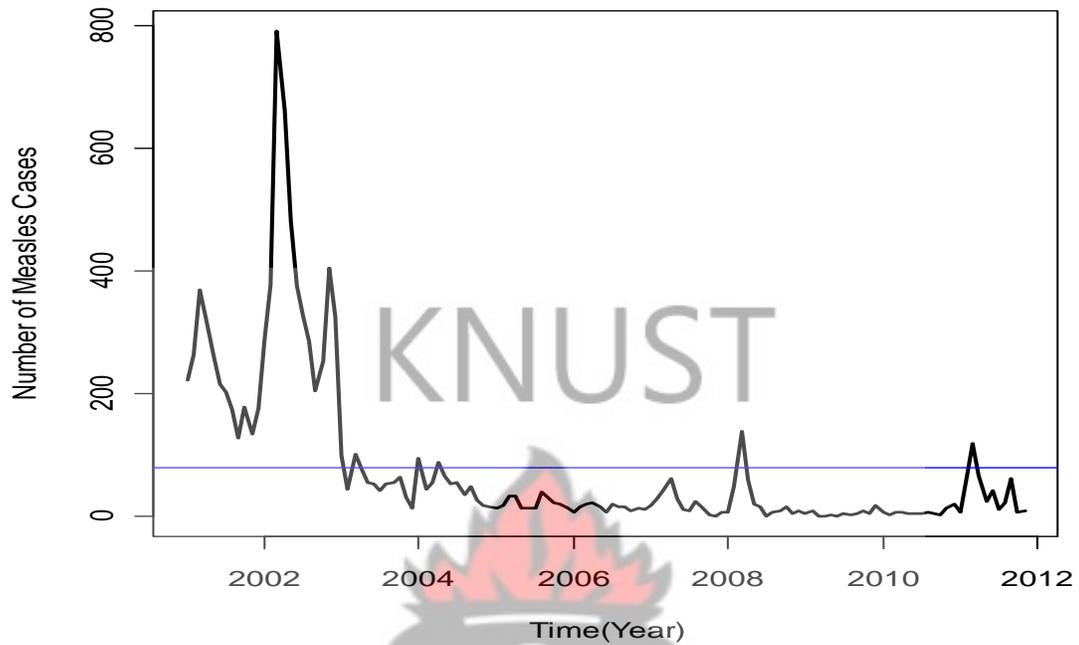


Figure 4.1: Time plot of measles prevalence in the Ashanti-Region of Ghana from January 2001 to November 2011.

Most of the data points are a little bit far apart from the mean. This indicates that there is a clear case of non-stationarity in the mean. It follows that the measles series is non-stationary in the mean.

4.2 Stationarity checks using the ACF, PACF, KPSS and Dickey-Fuller

Figure 4.2 depicts the autocorrelation function (ACF) of the measles data which describes the correlation between values of the measles at different points in time, as a function of the two times or of the time difference.

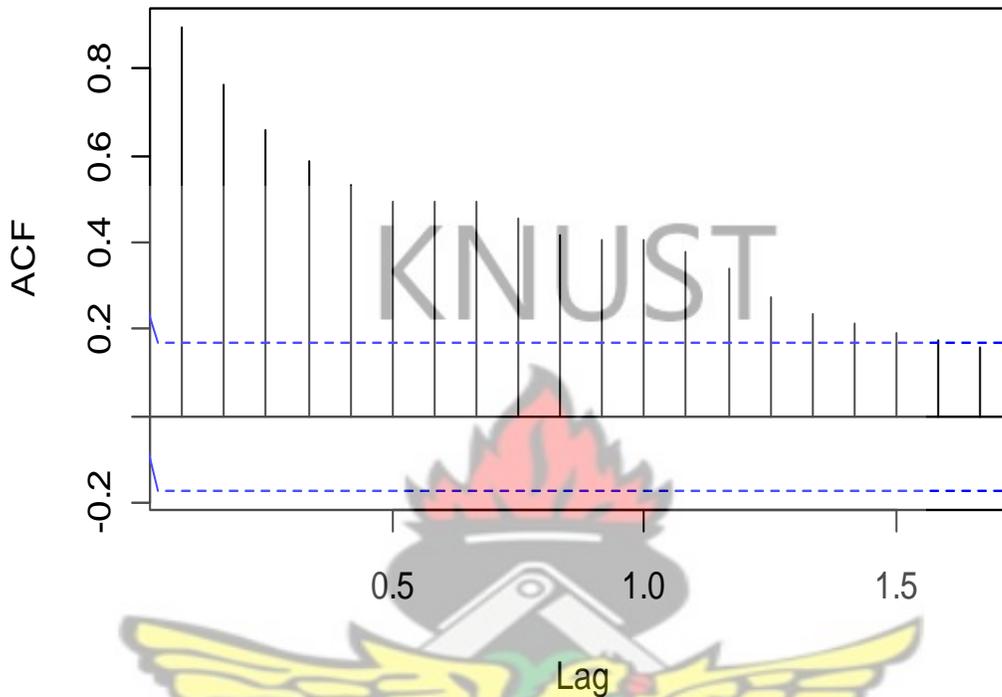


Figure 4.2: *The Autocorrelation Function (ACF) of Prevalence of Measles*

The lags from 1 to 18 autocorrelation exceed two standard errors above zero (they are significantly far from zero). The autocorrelation function is decreasing gradually with time and that shows that there is a non-stationarity in the measles data. The ACF plot confirms non-stationarity in the measles series.

Figure 4.3 exhibits the partial autocorrelation function (PACF) of the measles data. The first lag is almost unity which confirms that the measles time series is non-stationary.

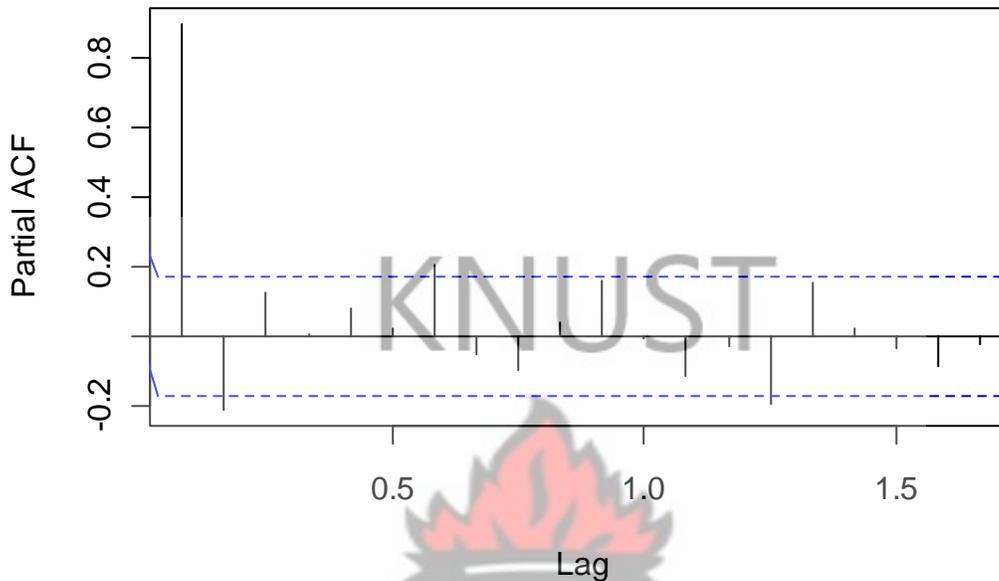


Figure 4.3: The Partial Auto-Correlation Function (PACF) of prevalence of measles

Table 4.1: Augmented Dickey-Fuller (ADF) and KPSS Tests

Augmented Dickey – Fuller Test		
Dickey-Fuller	Lag Order	P-Value
-2.7315	5	0.2723
KPSS Test		
KPSS Level	Lag Parameter	P-Value
2.0023	2	0.01

Table 4.1 shows both Augmented Dickey-Fuller and KPSS test results. There was no stationarity in the original measles data; since p-value for ADF test was greater than 0.05 and that of KPSS was less than 0.05 using a 5% significant level. The two tests confirmed that there was non-stationarity in the original measles data which needs to be differenced to achieve mean stationarity.

4.3 First-Order Difference of Prevalence of Measles in the Ashanti Region from January 2001 to November 2011.

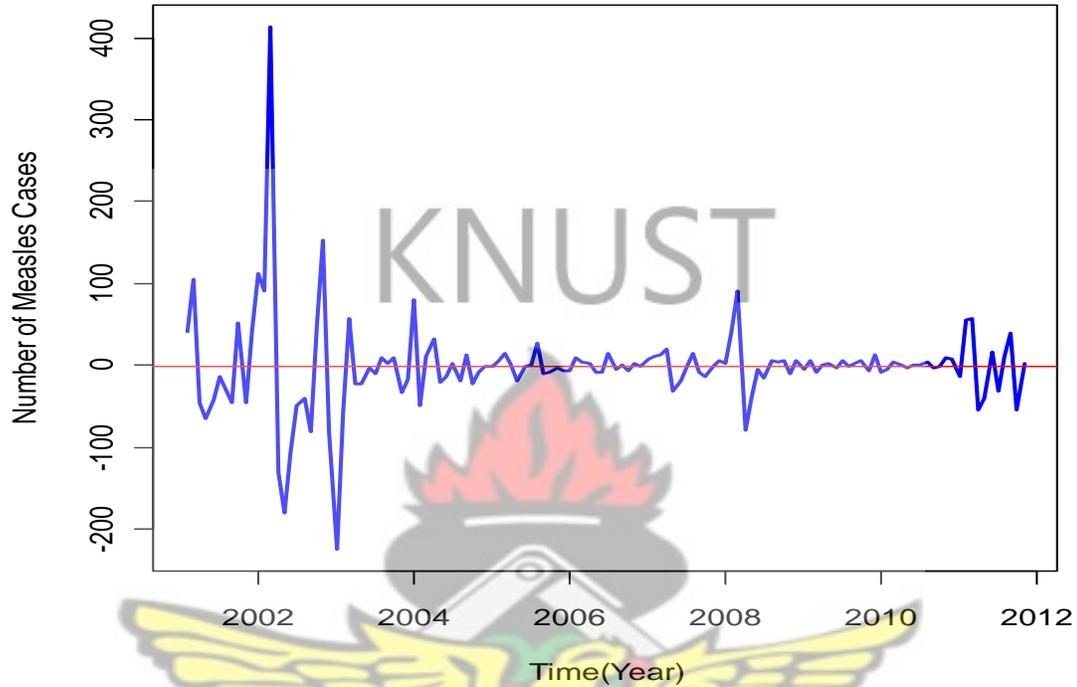


Figure 4.4: *First-Order Differencing of Prevalence of Measles in the Ashanti Region*

Figure 4.4 shows the first-order difference of the measles data. A transformation of the measles data using first-order difference was performed to remove the non-stationarity in the original measles data. The data fluctuate around a constant mean, independent of time, and the variance of the fluctuation remains essentially constant over time. There was not any seasonal behavior in the time plot, and the measles data now looks to be approximately stable for further investigations.

4.4 Objective Test for Stationarity for the First-Order Differenced Series

Table 4.2: Augmented Dickey-Fuller (ADF) and KPSS Tests

Augmented Dickey – Fuller Test		
Dickey-Fuller	Lag Order	P-Value
-7.003	5	0.01

KPSS Test		
KPSS Level	Lag Parameter	P-Value
0.0273	2	0.1

From Table 4.2 we can observe that there is a stationarity in the measles data, since p-value for ADF test was less than 0.05 and that of KPSS was greater than 0.05. The measles data now looks to be approximately stationary in the mean for further investigations.

4.5 Selecting Competing Models Using ACF and PACF of the First-Order Differencing of Measles Prevalence

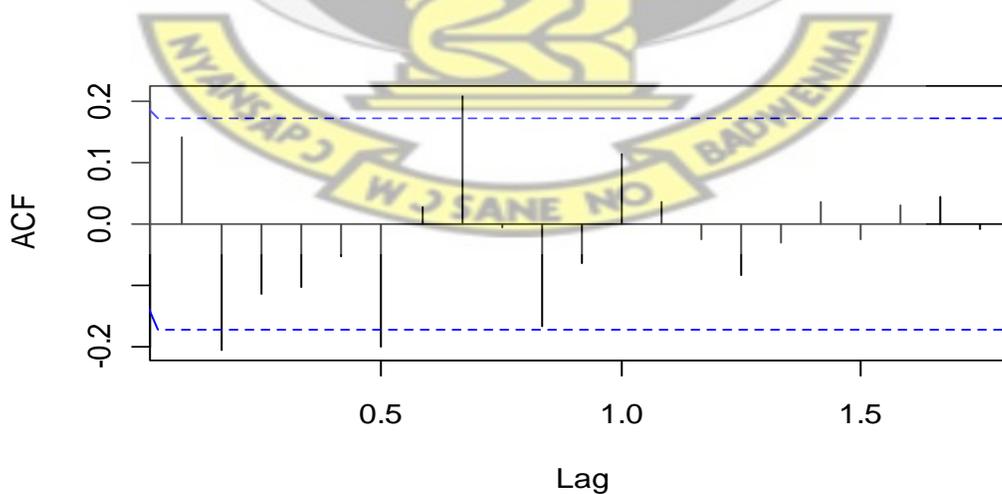


Figure 4.5: ACF of the first-order differencing of prevalence of measles

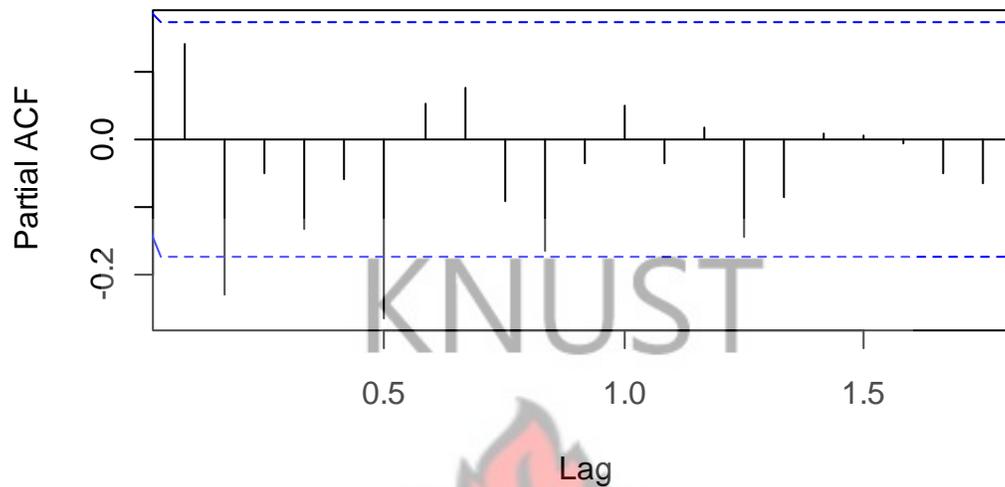


Figure 4.6: PACF of first-order differencing of prevalence of measles

Figure 4.5 shows the sample Auto-Correlation Function (ACF) of the first order differencing of the prevalence of measles. Except for marginal significance at lags 2, 6 and 8, the model seems to have captured the essence of the dependence in the series. Inspecting the sample ACF, we see that PACF is tailing off and the ACF is cutting off at lags 2, 6 and 8. This would suggest that the prevalence of measles follows an MA (2) model, MA (6) model or MA (8) model.

Figure 4.6 shows the sample Partial Auto-Correlation Function (PACF) of the first order differencing of the prevalence of measles in the Ashanti Region of Ghana at different lags. Inspecting the sample PACF, we see that the ACF is tailing off and the PACF is cutting off at lags 2 and 6. Except for marginal significance at lags 2 and 6, the model seems to have captured the essence of the dependence in the series. This suggests an

AR(2) or AR(6) model for the measles prevalence. As a preliminary analysis, we will fit both models. It follows that, in both the ACF and the PACF of the first order differencing of the measles data, the following models were suggested:

(i) ARIMA (0, 1, 2)

(ii) ARIMA (2, 1, 0)

(iii) ARIMA (1, 1, 0)

(iv) ARIMA (2, 1, 1)

4.6 Estimation of Tentative Models

4.6.1 Parameter estimate and diagnostics of ARIMA (0, 1, 2) model

Table 4.3: Parameter estimate for ARIMA (0, 1, 2) with non-zero mean

Coefficient	Estimate	Standard Error	t – Value
ma1	0.1311	0.0947	1.3844
ma2	-0.2968	0.1122	2.6453
AIC	AICc	BIC	Constant
1419.93	1420.25	1431.40	-1.8115

The coefficients of the estimated MA (2) parameters are within the invertibility condition bounds. The estimated ARMA(0, 2) model can be written as shown in equation (4.1):

$$X_t = 0.1311\omega_{t-1} - 0.2968\omega_{t-2} - 1.8115 + \omega_t \quad (4.1)$$

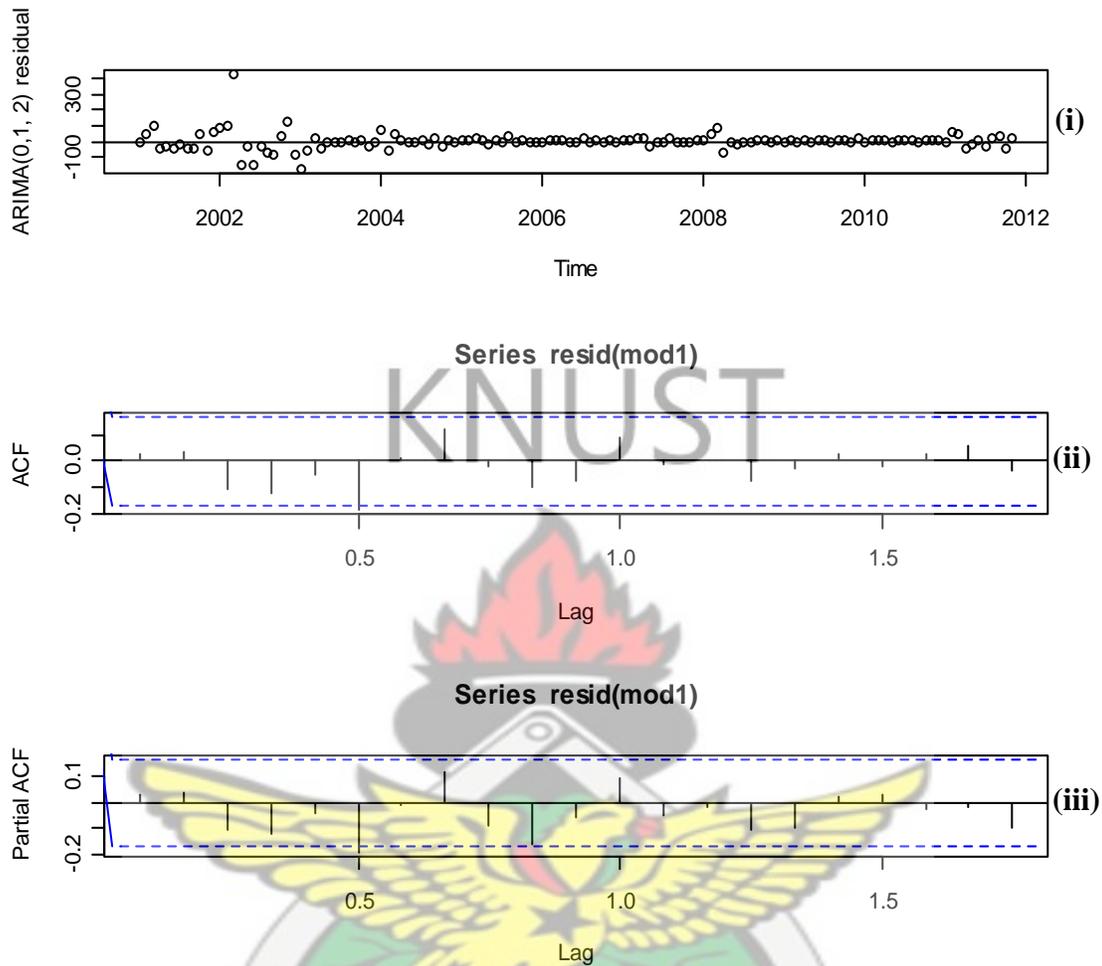


Figure 4.7: Diagnostic plot (ts-diagram) of the fitted ARIMA (0, 1, 2)

Figure 4.7 shows the diagnostics of the residuals of ARIMA (0, 1, 2). Figure 4.7 (i) is the time plot of the residuals against time. There is no obvious pattern in the plot except for a possible outlier, and looks like an independently and identically distributed (i. i. d) sequence of zero mean with a constant variance. Figure 4.7 (ii) is the plot of the ACF of the residuals. Here, there was not enough evidence of significant spikes which clearly shows that the residuals are white noise.

Figure 4.7 (iii) is the plot of the PACF of the residuals. Here, there was not enough evidence of significant spikes which clearly shows that the residuals are white noise.

Table 4.4: Box-Ljung test and Forecasts from ARIMA (0, 1, 2) with non- zero mean

Box - Ljung Test		
X Squared	Degrees of Freedom	P-Value
17.0893	20	0.6472

Results from Table 4.4 showed that the model's residuals were non - significant with Ljung Box test statistic of 17.0893 and a p-value of 0.6472. Hence the model was adequate for forecasting.

4.6.2 Parameter estimate and diagnostics of ARIMA(2, 1, 0) model

Table 4.5: Parameter estimate for ARIMA(2, 1, 0) with non-zero mean

Coefficient	Estimate	Standard Error	t – Value
ar1	0.1703	0.0852	1.9988
ar2	-0.2311	0.0861	2.6841
AIC	AICc	BIC	Constant
1421.28	1421.60	1432.75	-1.7820

The coefficients of the estimated AR(2) parameters are within the causality condition bounds. The estimated model for ARMA(2) was given by equation (4.2):

$$X_t = 0.1703X_{t-1} - 0.2311X_{t-2} - 1.7820 + e_t \quad (4.2)$$

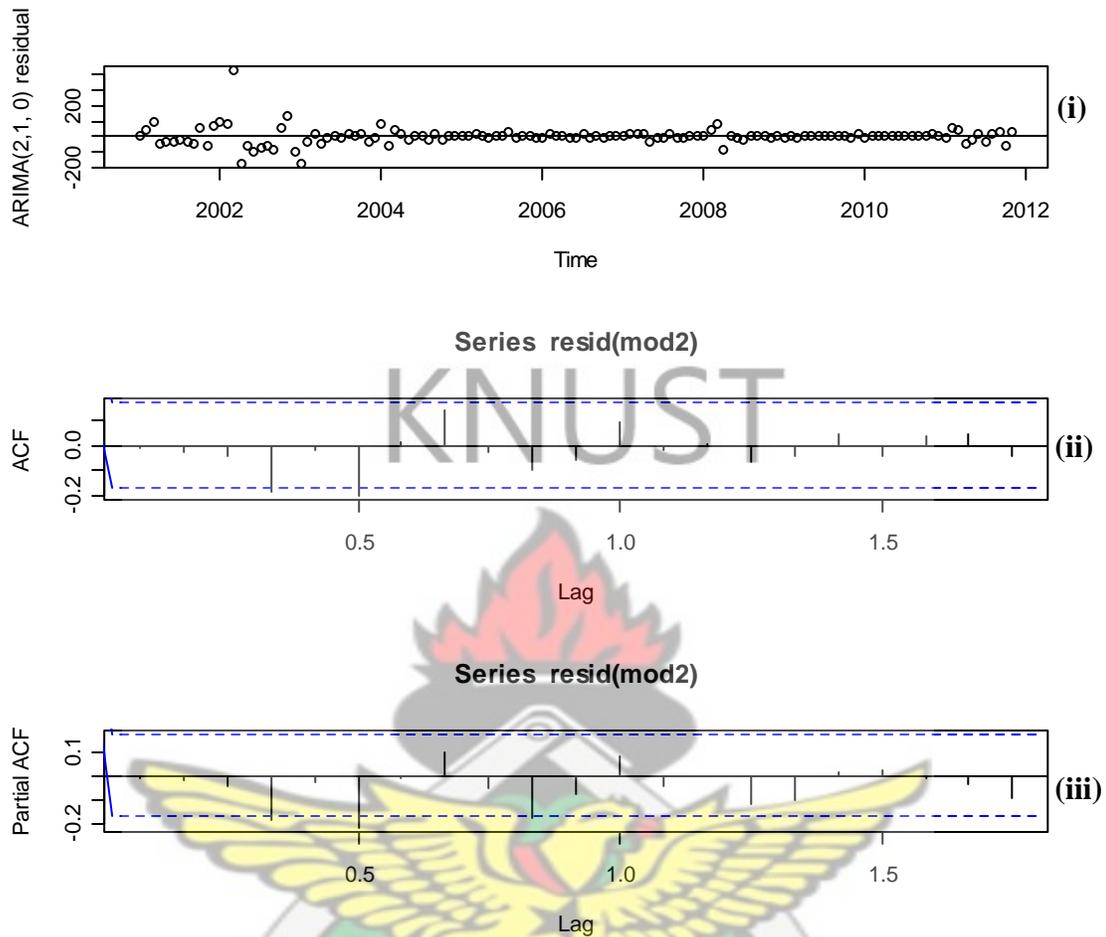


Figure 4.9: Diagnostic plot (ts-diagram) of the fitted ARIMA(2, 1, 0)

Figure 4.9 shows the diagnostics of the residuals of ARIMA(2, 1, 0). Figure 4.9 (i) is the time plot of the residuals against time. There is no obvious pattern in the plot except for a possible outlier, and looks like an independently and identically distributed (i. i. d) sequence of zero mean with a constant variance. Figure 4.9 (ii) is the plot of the ACF of the residuals. Here, there was not enough evidence of significant spikes which clearly shows that the residuals are white noise. Figure 4.9 (iii) is the plot of the PACF of the

residuals. Here, there was not enough evidence of significant spikes which clearly shows that the residuals are white noise.

Table 4.6: Box-Ljung test and Forecasts from ARIMA(2, 1, 0) with non- zero mean

Box - Ljung Test		
X Squared	Degrees of Freedom	P-Value
18.4227	20	0.5596

Results from Table 4.6 shows that the model’s residuals were non-significant with Ljung Box test statistic of 18.4227 and a p-value of 0.5596. Hence the model was adequate for forecasting.

4.6.3 Parameter estimates and diagnostics of ARIMA(1, 1, 0) model

The coefficient of the estimated AR(1) parameter is within the causality condition bounds. The estimated ARMA(1, 0) model can be written as shown in equation (4.3).

$$X_t = 0.1399X_{t-1} - 2.0022 + e_t \quad (4.3)$$

Table 4.7: Parameter estimate for ARIMA(1, 1, 0) with non-zero mean

Coefficient	Estimate	Standard Error	t – Value
ar1	0.1399	0.0867	1.6136
AIC	AICc	BIC	Constant
1426.28	1426.47	1434.88	-2.0022

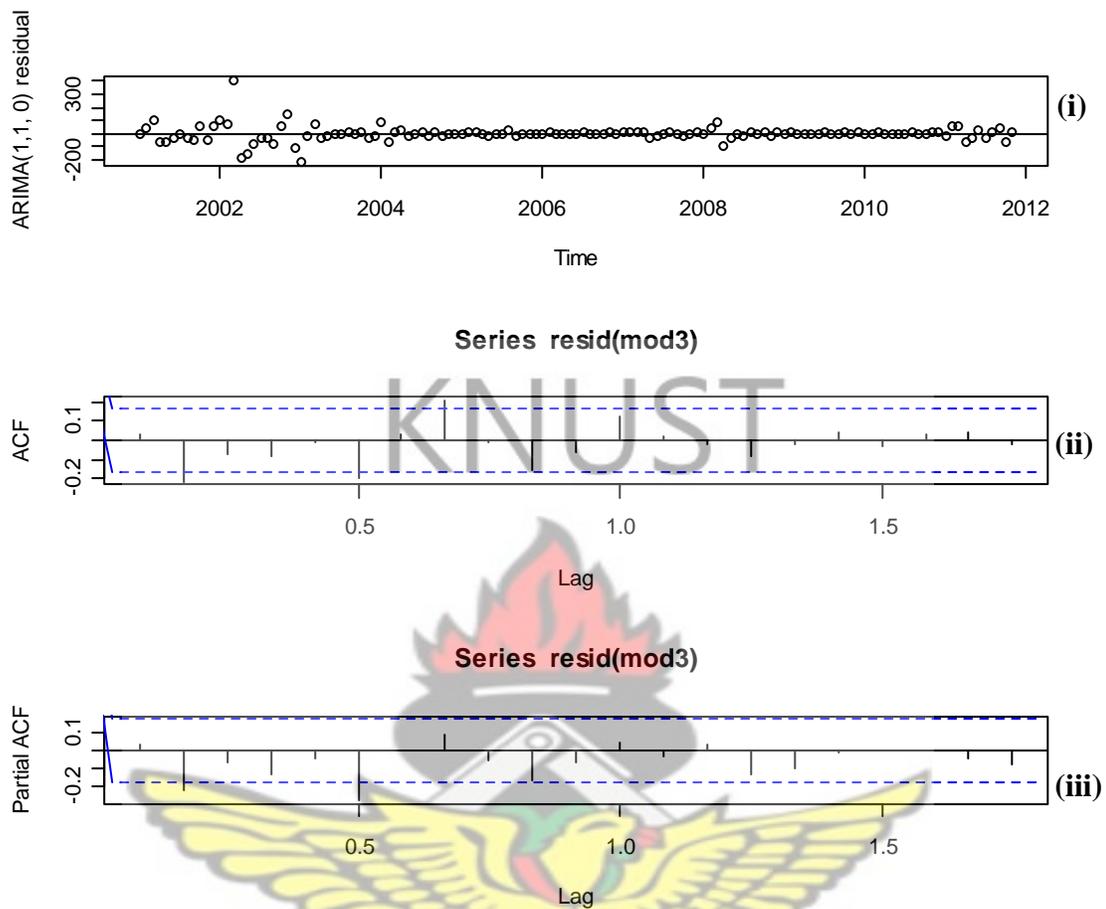


Figure 4.11: Diagnostic plot (*ts-diagram*) of the fitted ARIMA(1, 1, 0)

Figure 4.11 depicts the diagnostics of the residuals of ARIMA(1, 1, 0). Figure 4.11 (i) is the time plot of the residuals with time. There is no obvious pattern in the plot except for a possible outlier, and looks like an independently and identically distributed (i. i. d) sequence of zero mean with a constant variance. Figure 4.11 (ii) is the plot of the ACF of the residuals. Here, there was not enough evidence of significant spikes which clearly shows that the residuals are white noise. Figure 4.11 (iii) is the plot of the PACF of the residuals. Here, there was not enough evidence of significant spikes which clearly shows that the residuals are white noise.

Table 4.8: Box-Ljung test and Forecasts from ARIMA (1, 1, 0) with non- zero mean

Box - Ljung Test		
X Squared	Degrees of Freedom	P-Value
29.0519	20	0.0864

Results from Table 4.8 shows that the model's residuals were non - significant with Ljung Box test statistic of 29.0519 and a p-value of 0.0864. Hence the model was adequate for forecasting.

4.6.4 Parameter estimate and diagnostics of ARIMA(2, 1, 1) model

Table 4.9: Parameter estimate for ARIMA(2, 1, 1) with non-zero mean

Coefficient	Estimate	Standard Error	t – Value
ar1	0.9800	0.1051	9.3245
ar2	-0.2693	0.0890	3.0258
ma1	-0.9107	0.0781	11.6607
AIC	AICc	BIC	Constant
1417.05	1417.53	1431.39	-1.9980

The coefficients of the estimated ARMA (2, 1) parameters are within the causality and invertibility condition bounds. The estimated ARMA(2, 1) model can be written as shown in equation (4.4).

$$X_t = 0.98X_{t-1} - 0.2693X_{t-2} - 0.9107\omega_{t-1} - 1.9980 + \omega_t \quad (4.4)$$

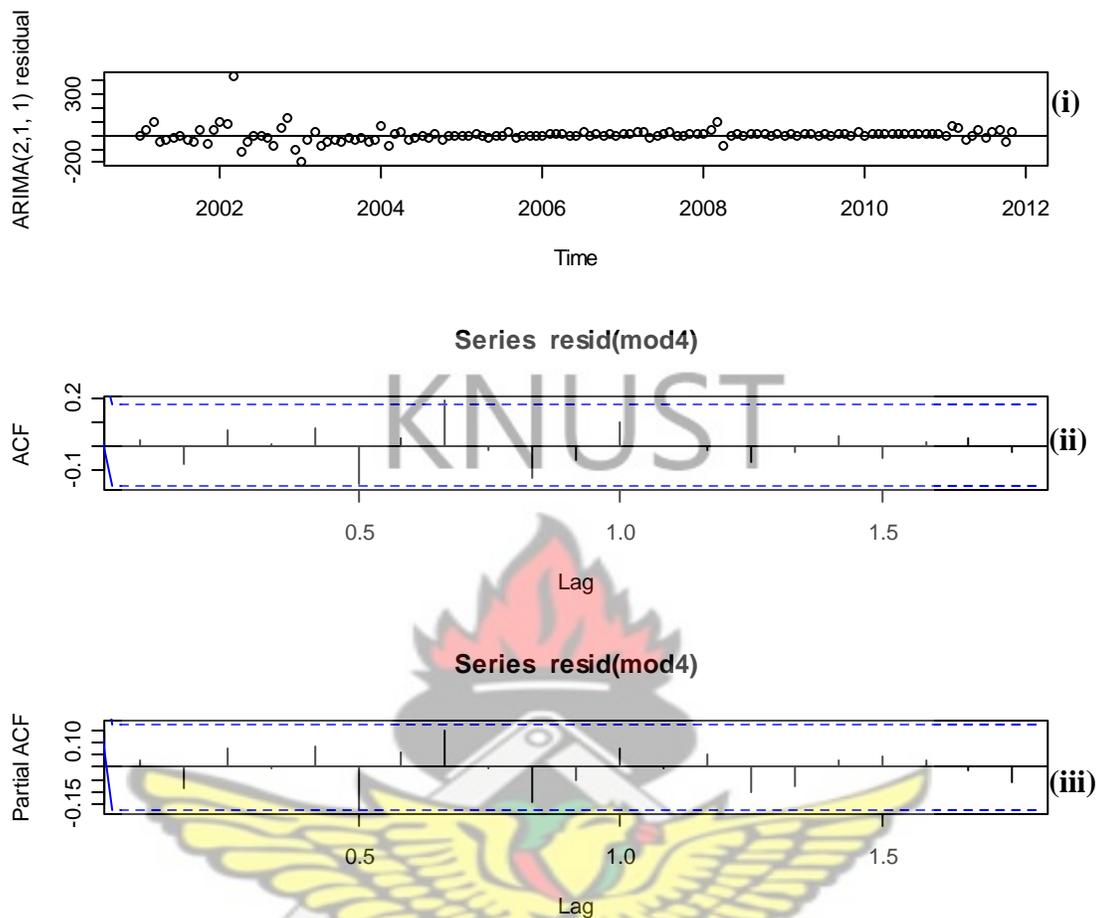


Figure 4.13: Diagnostic plot (ts-diagram) of the fitted ARIMA(2, 1, 1)

Figure 4.13 exhibits the diagnostics of the residuals of ARIMA (2, 1, 1). Figure 4.13 (i) is the time plot of the residuals against time. There is no obvious pattern in the plot except for a possible outlier, and looks like an independently and identically distributed (i. i. d) sequence of zero mean with a constant variance. Figure 4.13 (ii) is the plot of the ACF of the residuals. Here, there was not enough evidence of significant spikes which clearly shows that the residuals are white noise. Figure 4.13 (iii) is the plot of the PACF of the residuals. Here, there was not enough evidence of significant spikes which clearly shows that the residuals are white noise.

Table 4.10: Box-Ljung test and Forecasts from ARIMA (2, 1, 1) with non- zero mean

Box - Ljung Test		
X Squared	Degrees of Freedom	P-Value
18.0093	20	0.5868

Results from Table 4.10 showed that the model's residuals were non - significant with Ljung Box test statistic of 18.0093 and a p-value of 0.5868. Thus the model was adequate for forecasting.

4.7 Forecasting From ARIMA (2, 1, 1)

Table 4.11: Point Forecasts from ARIMA (2, 1, 1) with non-zero mean

Point Forecast From ARIMA (2, 1, 1)					
Dec-2011	Jan-2012	Feb-2012	Mar-2012	Apr-2012	May-2012
15.9037	20.6717	23.5551	24.9658	25.053	25.6206

Results from Table 4.11 shows that measles prevalence in the Ashanti Region will increase from December 2011 to May 2012.

From Figure 4.14, the yellow line depicts the 95% confidence interval, the red line is the 85% confidence interval and the blue line is the forecasting points. The model was used to forecast six months ahead and showed that the measles prevalence in the Ashanti Region will be increased as shown in the blue line.

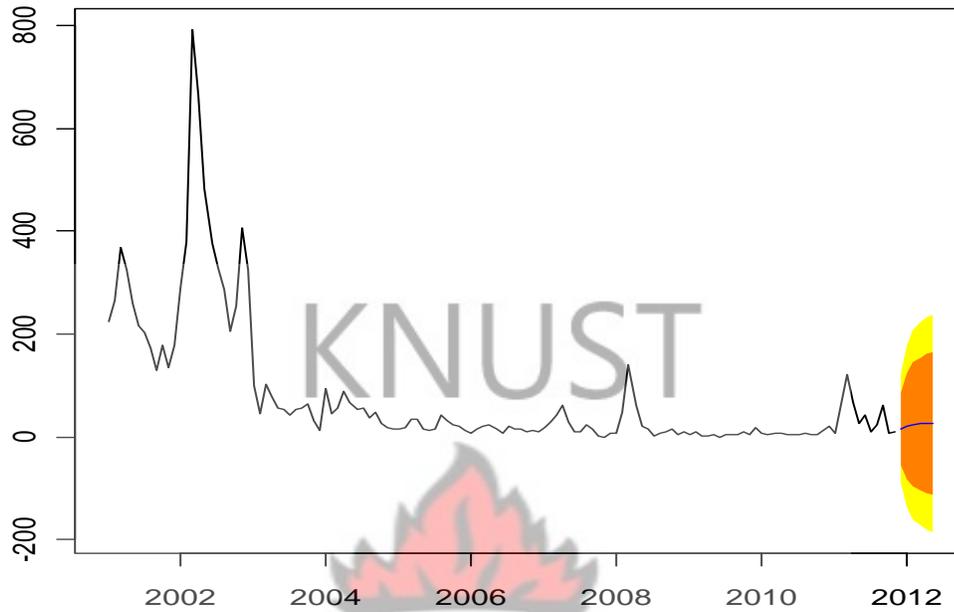


Figure 4.14: Forecasts from ARIMA (2, 1, 1) with non-zero mean

4.8 The Error Metrics

Table 4.12: The Mean Absolute Error (MAE) and the Mean Squared Error (MSE)

Mean Absolute Error (MAE)			
ARIMA(1, 1, 0)	ARIMA(2, 1, 0)	ARIMA(2, 1, 0)	ARIMA(0, 1, 2)
29.39361	28.44202	28.1141	27.98298
Mean Squared Error (MSE)			
ARIMA(1, 1, 0)	ARIMA(2, 1, 0)	ARIMA(0, 1, 2)	ARIMA(2, 1, 1)
3277.671	3103.249	3069.110	2947.151

Forecasting accuracy based on the Mean Absolute Error (MAPE) of the forecasted values was checked for each fitted model as shown in Table 4.12. It highly favored the forecasted value of ARIMA(0, 1, 2). This means that, the ARIMA(0, 1, 2) forecast error of 27.98298 out-performed all the forecast errors so far as the MAE is concerned. Similarly, the forecasting accuracy based on the Mean Square Error (MSE) of the forecasted values also favored ARIMA(2, 1, 1), the best selected model. This means that, the ARIMA(2, 1, 1) forecast error of 2947.151 out-performed all the forecast errors so far as the MSE is concerned. Hence ARIMA(2, 1, 1) was confirmed to be the best model.

4.9 Diebold – Mariano (DM) Test

Table 4.13: Diebold – Mariano Test for Comparing MAE from each Model

DATA	DM	FORECAST HORIZON	P - VALUE
e1 verses e2	-1.4418	2	0.9253 (0.05)
e1 verses e3	-1.1685	2	0.8787 (0.05)
e1 verses e4	-0.0784	2	0.5312 (0.05)
e2 verses e3	-0.8730	2	0.8087 (0.05)
e2 verses e4	0.1810	2	0.4282 (0.05)
e3 verses e4	0.6404	2	0.2610 (0.05)

From Table 4.13, e1 is the in-sample error from ARIMA(0, 1, 2) model, e2 is the in-sample error from ARIMA(2, 1, 0) model, e3 is the in-sample error from ARIMA(1, 1, 0) model and e4 is the in-sample error from ARIMA(2, 1, 1) model. The test rejected the null hypotheses of the pairs of errors compared for the models selected. The results show that all the models predict equally.

4.10 Time plot of actual measles data and the fitted Models

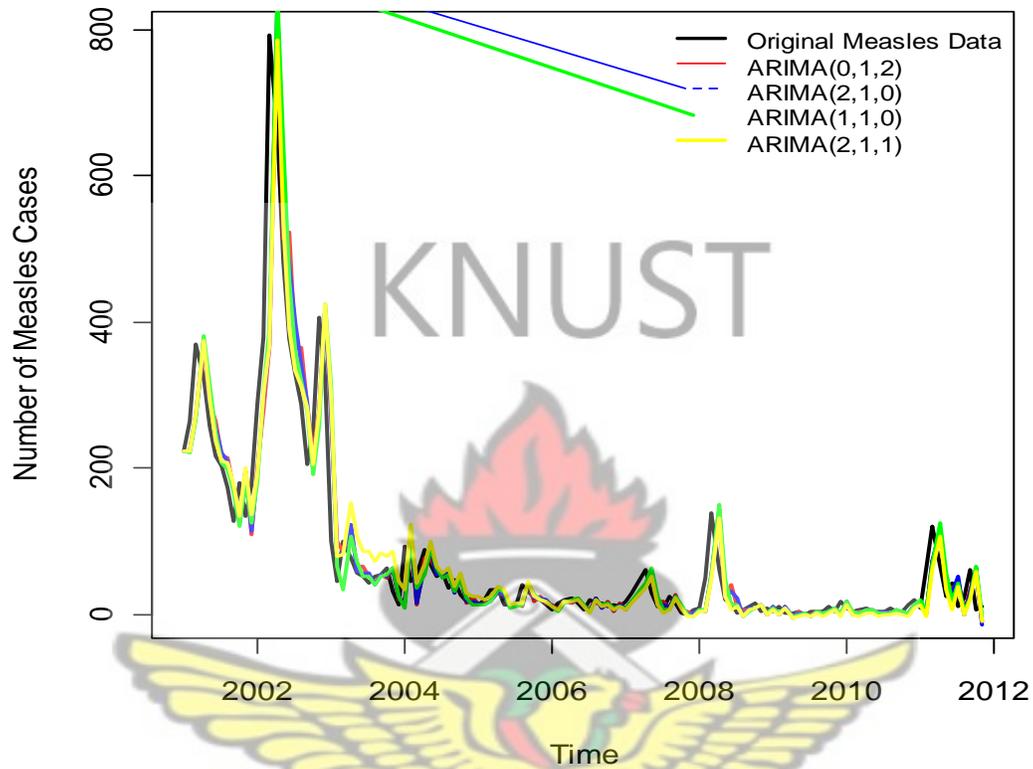


Figure 4.15: *Time plot of actual measles data and the fitted Models*

Figure 4.15 exhibits the time plot of actual measles data and the fitted ARIMA models.

All the models seem to fit the data quite well.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

This chapter contains the conclusion, summary and findings and the recommendation of the thesis.

5.1 Conclusion

KNUST

Results from this thesis shows that increasing use of measles vaccine is having a significant impact on the rate of measles transmission and its related complications in the Ashanti region of Ghana. Increasing the measles vaccination coverage rate in the region will further decrease the prevalence of measles in the region and decreasing the vaccination coverage will increase the rate of transmission of measles in the region which will affect the development of human resources in the country.

This model did not consider mass vaccination as one of the methods to prevent the prevalence of measles in the region. It concentrated on herd immunity due to the huge sum of money that needs to be spent in carrying out mass vaccination. The results of this thesis can be used as a tool to facilitate the introduction of measles vaccine and improve measles vaccination in the country as a whole.

Table 5.1: Summary of Diagnostics Tests

MODEL	ARIMA(0, 1, 2)	ARIMA(2, 1, 0)	ARIMA(1, 1, 0)	ARIMA(2, 1, 1)
AIC	1419.93	1421.28	1426.28	1417.05
AICc	1420.25	1421.6	1426.47	1417.53

From Table 5.1 ARIMA(2, 1, 1) had the lowest AIC value of 1417.05. It was the best model so far as the AIC was concerned. Hence the best selected model was

$$X_t = 0.98X_{t-1} - 0.2693X_{t-2} - 0.9107\omega_{t-1} - 1.9980 + \omega_t$$

The Mean Absolute Error (MAE) for ARIMA(0, 1, 2) of 27.98298 was the lowest. Its forecast error of 27.98298 out-performed all the forecast errors. However, the forecasting accuracy based on the Mean Squared Error (MSE) for ARIMA(2, 1, 1), the best selected model, was the lowest. Its forecast error of 2947.151 out-performed all the forecast errors.

In conclusion, the research study reported in this monograph has found that measles data in the Ashanti Region of Ghana could best be modelled with ARIMA(2, 1, 1). The study again found out that measles prevalence in the Ashanti Region is expected to increase if no preventative measures are taken.

5.2 Recommendations

The following recommendation is being made to the stakeholders of Ashanti Region:

- Ghana Health Service should continue the mass measles vaccination in the region to possibly eradicate the disease.

REFERENCES

- Akaike, H. (1974). "A new look at the statistical model identification." *IEEE Transactions on Automatic Control*, 19, 716 – 723.
- Allen L.J., Lewis T., Martin C. F., Jones M.A., Lo C.K., Stamp M., Mundel G. and Way A.B. (1993), "Analysis of a Measles Epidemic", *Statistic in Medicine*, Volume 12, Issue 3 – 4, Pages 229 – 239, DOI: 10.1002 / Sim, 4780120307, Texas Tech University.
- Allen L. J. S., Jones M. A. and Martin C. F. (1991). "A discrete-time model with vaccination for a measles epidemic.", Department of Mathematics, Texas Tech University, Lubbock, Texas 79409, USA.
- Bauch, C. T. (2008), "Role of Mathematical Models in Explaining Recurrent Outbreaks of Infectious Childhood Diseases" ,*Lecture Notes in Mathematics*, 2008, Volume 1945 / 2008, 297-319, DOI: 10.1007/978-3-540-78911-6_11
- Blog, M.Y. (2008), "Components of Time Series", www.jas.blogspot.com/./components-of-time-series.htm
- Box, B. and Greefell B. (1995). "Space, Persistence and Dynamics of Measles Epidemics." Zoology Department, Cambridge University, Downing Street, Cambridge CB2 3EJ, UK.
- Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994), *Time Series Analysis: Forecasting and Control*, 3rd edition, Prentice Hall, Englewood Cliffs, New Jersey.
- Centre for Disease Control and Prevention (2011), www.cdc.gov/measles/, 24/7.
- Chen S. C, Chang C. F., Jou L. J. and Liao C. M. (2007), "Modeling vaccination programmes against measles in Taiwan", *Epidemiology and Infection* (2007), 135 : pp 775-786, DOI: 10.1017/S0950268806007369.
- Cliff A. D. and Haggett P. (1993), "Statistical Modeling of Measles and Influenza outbreaks", Department of Geography, University of Cambridge, U. K.
- Cryer D. Jonathan and Kung – Sik Chan (2008), "Time Series Analysis With Applications in R", Second Edition, U.S.A., Springer.
- Duncan S. R., Scott S., and Duncan C. J. (1999), "A Demographic model of Measles Epidemics", *European Journal of Population*, Volume 15, Number 2, 185 – 198, Doi:10,1023 / A: 1006224902376, U. K.

- Earn D. J. D, Rohani P., Bolker B. M and Grenfell B. T. (2000), "A Simple Model for Complex Dynamical Transitions in Epidemics", *Journal of Original Scientific Research*, Vol. 287, no. 5453, pp. 667-670, DOI: 10.1126/science.287.5453.667
- Ellner S.P, Bailey B.A, Bobashev G.V, Gallant A.R, Grenfell R.T and Nyehka D.W (1998), "Noise and Nonlinearity in Measles Epidemics: Combining Mechanistic and Statistical approaches to Population Modeling", Department of Zoology, Cambridge University, U.K.
- Georgette, N and Allen D (2009): Determination of the Most Cost Effective Pre-Outbreak Immunization Rate through a Novel Approach. *The Internet Journal of Epidemiology* ISSN: 1540-2614, . 2009, Volume 6, Number 2
- Glass, K., Kappey, J. and Grenfell B. T. (2003). "The effect of heterogeneity in measles vaccination on population immunity." National Centre for Epidemiology and Population Health, Australian National University, Canberra 0200, Australia.
- Gottman, M. J. (1981), *Time Series Analysis, A comprehensive Introduction for Social Scientists*, Cambridge University Press.
- Grais, R. F., Conlan A. J. K., Ferrari M. J., Djibo A., Menach A. L., Bjørnstad O. N. and B.T Grenfell B. T. (2006). "Time is of the essence: exploring a measles outbreak response vaccination in Niamey, Niger." (rebecca.grais@epicentre.msf.org).
- Grenfell, B. T. (1992). " *Journal of the Royal Statistical Society.*" Series B (Methodological), University of Cambridge, UK.
- Kassem, T. G. and Ndam J.N. (2010), "A Stochastic Modeling of Recurrent Measles Epidemics", *Science World Journal*, Volume 3 Number 4, 2008, www.sciecnceworldjournal.org, ISSN 1597-6343
- Lloyd, A. L. (2000), "Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods." Program in Theoretical Biology, Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540, USA.
- Lloyd, A. L. (2001). "Realistic Distributions of Infectious Periods in Epidemic Models: Changing Patterns of Persistence and Dynamics." Program in Theoretical Biology, Institute for Advanced Study, Einstein Drive, Princeton, New Jersey, 08540, USA.

- London W. P. and Yorke J. A. (1973) “Recurrent Outbreaks of Measles, Chickenpox and Mumps - Seasonal Variation in Contact Rates”, University of Maryland College Park. Md. 20740.
- Keeling, M. J. and Grenfell, B. T. (4 Feb, 2002) “Understanding the persistence of measles: reconciling theory, simulation and observation” *Department of Zoology, University of Cambridge, Downing Street, Cambridge CB23EJ, UK*
- Nagpaul, P. S. (2005), “Time Series Analysis in WinIDAMS”, New Delhi, India.
- Nettleman, M. (2008), emedicinehealth, www.emedicinehealth.com , children’s health az list.
- NHS Choices (2010), www.nhs.uk/conditions/measles/pages/symptoms.aspx
- Nita Bharti, Yingcun Xia, Ottar N Bjornstad and Bryan T. Grenfell (2008), “Measles on the Edge: Coastal Heterogeneities and Infection Dynamics”, *PLoS ONE* 3(4); e1941,doi:10.1371/journal.pone.0001941, University of Oxford, UK.
- Perez L. and Dragicevic S. (2009), “An agent-based approach for modeling dynamics of contagious disease spread”, *International Journal of Health Geographic’s*, 8:50 doi: 10.1186 / 1476-072X – 8 – 50
- Rhodes C. J. and Anderson R. M. (1996). “A Scaling Analysis of Measles Epidemics in a Small Population.” *Phil. Trans. R. Soc. Lond. B* 29 December 1996 vol. 351 no. 1348, 1679-1688.
- Roberts M. G. and Tobias M. I. (2000) “Predicting and Preventing Measles Epidemics in New Zealand: Application of a Mathematical Model”, *Epidemiol infct* (2000), 124, 279 – 287, New Zealand, Cambridge University press, U. K.
- Sattenspiel L. and Dietz K. (1995), “A structured epidemic model incorporating geographic mobility among regions”. Department of Anthropology, University of Missouri, Columbia, USA.
- Schoenstadt, A. (2006), eMedTV, measles.amedtv.com/measles/cause-of-measles.html
- Shumway H. R. and Stoffer S. D. (2006), “Time Series Analysis and Its Applications With R Examples”, Second Edition, U.S.A., Springer.
- Spyros M., Wheelwright S. C. and Hyndman R. J. (1998), “Forecasting: Methods and Applications”, Third Edition, U.S.A., John Wiley & Sons, Inc.
- Stamp, M. (1990). “A mathematical analysis and simulation of a localized measles epidemic.” Texas Tech University, Lubbock, Texas 79409-1042, USA.

- Stone L., Shulgin B. and Agur Z. (2000) “Theoretical Examination of the Pulse Vaccination Policy in the SIR Epidemic Model”, *Mathematical and Computer Modeling* 31 (2000) 207-215, www.elsevier.nl/locate/mcm
- Souza, R. C. (1982), "Forecasting the progress of epidemics by means of a Bayesian-entropy framework" *Environment and Planning A* 14(1) 49 – 60.
- Tarwater P. M. and Martin C. F. (2001). “Effects of population density on the spread of disease.” DOI: 10.1002/cplx.10003
- Thacker S. B. and Millar J. D. (1991), “Mathematical Modeling and Attempts to Eliminate Measles: A Tribute to the Late Professor George Macdonald”, *American Journal of Epidemiology*, volume 133, issue 6, Pp. 517 – 525, Atlanta, GA 30333
- Trottier H. and Philippe P. (2001), “Deterministic Modeling of Infectious Diseases: Theory And Methods”, *The Internet Journal of Infectious Diseases*, ISSN: 1528-8366, Volume 1 Number 2.
- Trottier H. and Philippe P. (2006), “Univariate time series analysis of pertussis, mumps measles and rubella: Theory and Methods”, *The Internet Journal of Infectious Diseases*, ISSN: 2528-7836, Volume 3 Number 4.
- Wallinga J., Janneke C., Heijne M. and Kretzschmar (2005), “A Measles Epidemic Threshold in highly Vaccinated Population”, *Centre for Infectious Diseases Epidemiology, National Institute for Public health and Environment, Bilthoven, The Netherlands.*
- Wikipedia, the free encyclopedia (2008), en.wikipedia.org/wiki/Measles
- Wikipedia, the free encyclopedia (2011), en.wikipedia.org/wiki/Ashanti_Region
- World Health Organization (2011), WHO Media Centre, Fact Sheet, Number 286, www.who.int/mediacentre/factsheets/fs/en/,
- Yingcun X, Bjornstad O. N. and Grenfell B. T. (2004), “Measles Metapopulation Dynamics: A gravity model for Epidemiological Coupling and dynamics” *The American Naturalist*, Volume 164, Number 2, University of Cambridge, Cambridge, CB2 3EJ, U. K.
- Zaman G., Kang Y. H. and Jung I. H. (2007), “Stability analysis and optimal vaccination of an *SIR* epidemic model”, Department of Mathematics, University of Ulsan, Ulsan 680-749, South Korea.
- Zivot, E. (2004), ‘Notes on Forecasting’ faculty.washington.edu/ezivot/econ584/notes/forecast.pdf

APPENDIX

Table 1: Measles data in the Ashanti Region from January, 2001 to December, 2011.

Period	Cases	Period	Cases	Period	Cases	Period	Cases
1	223	34	64	67	20	100	1
2	264	35	31	68	15	101	3
3	369	36	13	69	16	102	0
4	324	37	94	70	10	103	5
5	259	38	44	71	13	104	3
6	216	39	55	72	11	105	5
7	203	40	88	73	18	106	10
8	174	41	67	74	29	107	4
9	128	42	53	75	42	108	17
10	179	43	56	76	62	109	8
11	134	44	36	77	30	110	3
12	177	45	48	78	11	111	7
13	289	46	26	79	9	112	8
14	379	47	17	80	24	113	4
15	793	48	16	81	15	114	4
16	663	49	14	82	2	115	4
17	483	50	19	83	0	116	8
18	377	51	33	84	6	117	5
19	328	52	33	85	8	118	3
20	287	53	14	86	49	119	13
21	205	54	13	87	139	120	21
22	253	55	14	88	60	121	8
23	406	56	41	89	21	122	63
24	325	57	31	90	16	123	120
25	100	58	23	91	1	124	66
26	44	59	20	92	6	125	25
27	101	60	13	93	10	126	42
28	78	61	7	94	15	127	11
29	56	62	16	95	4	128	22
30	53	63	20	96	9	129	62
31	44	64	23	97	4	130	7
32	53	65	15	98	10	131	10
33	55	66	6	99	1	132	16

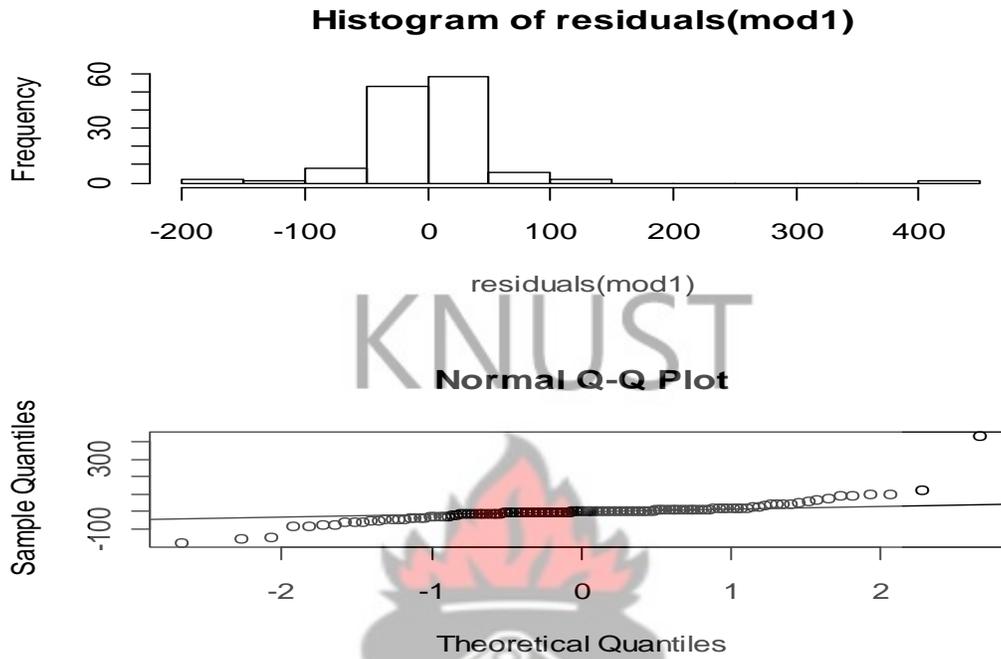


Figure 1: Histogram plot and the Normal Q-Q plot of the fitted ARIMA(0, 1, 2)

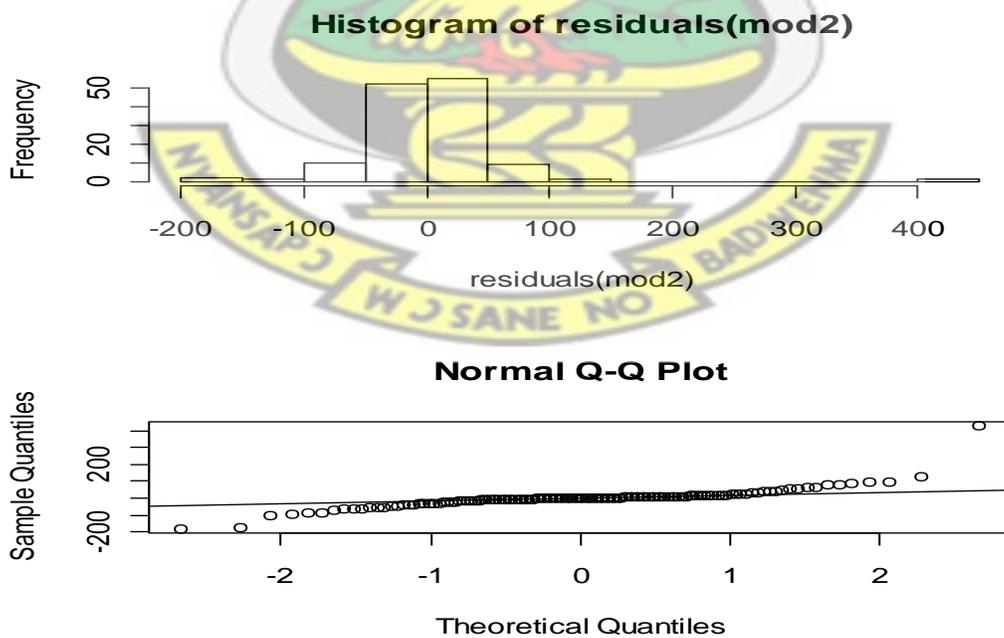


Figure 2: Histogram plot and the Normal Q-Q plot of the fitted ARIMA(2, 1, 0)

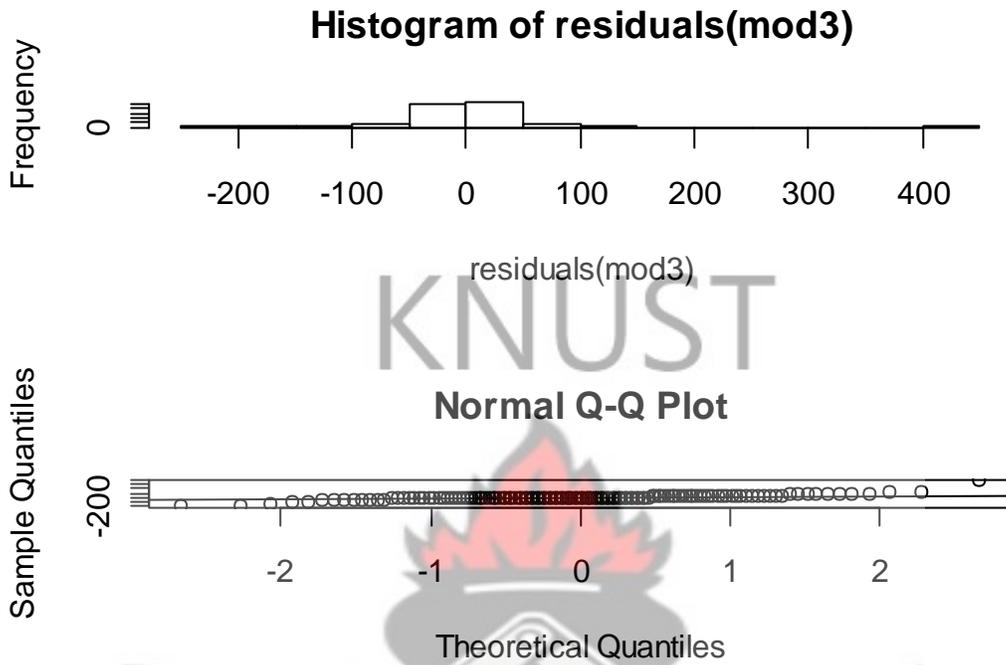


Figure 3: Histogram plot and the Normal Q-Q plot of the fitted ARIMA(1, 1, 0)

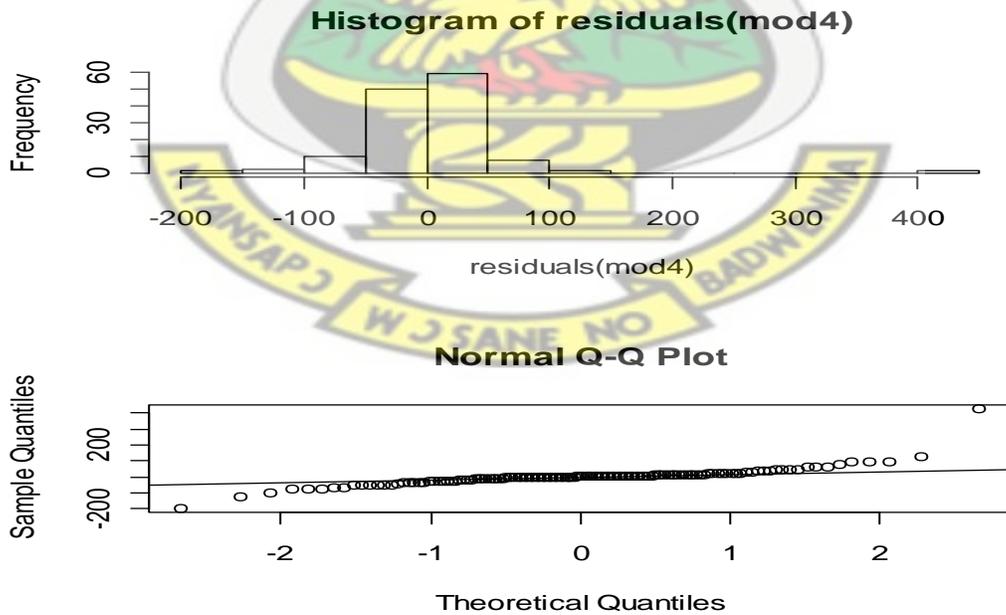


Figure 4: Histogram plot and the Normal Q-Q plot of the fitted ARIMA(2, 1, 1)