KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY, KUMASI

FACULTY OF SCIENCE

DEPARTMENT OF MATHEMATICS

TIME SERIES ANALYSIS OF MALARIA CASES IN ASHATI REGION,A CASE STUDY OF OUT PATIENT DEPARTMENT OF KOMFO ANOKYE TEACHING HOSPITAL.

BY

KATE KORANKYE AFRIYIE

MAY 2012.

KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY, KUMASI

FACULTY OF SCIENCE

DEPARTMENT OF MATHEMATICS

TIME SERIES ANALYSIS OF MALARIA CASES IN ASHATI REGION,A CASE STUDY OF OUT PATIENT DEPARTMENT OF KOMFO ANOKYE TEACHING HOSPITAL

A thesis submitted to the Department of Mathematics of the College of Science,

Kwame Nkrumah University of Science and Technology, Kumasi in partial fulfillment of the requirement for the Degree of

MPHIL. MATHEMATICS

BY

KATE KORANKYE AFRIYIE

(PG3940709)

MAY 2012.

DECLARATION

I hereby declare that this thesis has not been submitted previously either wholly or partially for post Degree in the Kwame Nkrumah University of Science and Technology(KNUST) or anywhere else .

KATE AFRIYIE KORANKYE 23/05/12

(PG 3940709) Signature Date

DR. S.K. AMPONSAH ....

(SUPERVISOR) Signature Date

MR. K . F DARKWAH ....

(HEAD OF DEPARTMENT) Signature Date

ACKNOWLEDGEMENT

First and foremost , I wish to render my sincere gratitude to the Almighty Good through whose divine protection and guidance I have gone through this course successfully . I would like to express my sincere thanks to my supervisor, Dr. S.K Amponsah for his patience guidance , encouragement and time.

I am also indebted to many people who have contributed in diverse ways to the success of my research work . Among these are Mr. S. K. Appiah, Mr. Frank Kofi Owusu and Degraft all of Mathematics Department.

I express my gratitude to the Department Of Mathematics of KNUST for making available all relevant materials and assistance for the successful accomplishment of this thesis.

ABSTRACT

Malaria still remains a public health problem in some regions of Ghana. To strength the country's prevention and control measures, this study was carried out to develop forecasting and predication models of malaria cases in the Out Patient Department (OPD) of Komfo Anokye Teaching Hospital (KATH) using time series and ARIMA. This study was carried out retrospectively using the monthly reported malaria cases from Out Patient Department (OPD) of Komfo Anokye Teaching Hospital (KATH). Time series analysis was performed on monthly cases from 200-2010 in Out Patient Department (OPD) of Komfo Anokye Teaching Hospital (KATH). The time series models derived from parameters estimates and diagnostics of ARIMA was deployed to identify the best model. The ARIMA forecasting models could be used for planning and managing malaria prevention and control programme in Ashanti region by the use of computer software. It was found that parameters in $ARIMA(1,0,0)(2,0,2)_{[12]}$ model were all significant and the best. The forecasted of malaria cases from January to December. ARIMA models of time-series analysis were useful in forecasting the number of cases in the Out Patient Department (OPD) of Komfo Anokye Teaching Hospital (KATH). The objectives of this study is to use time series analysis specifically ARIMA model to model Malaria rate Ashanti Region from 2000 to 2010. The method used to analyze the data was the Box-Jenkins ARIMA. The best selected model is $ARIMA(1,0,0)(2,0,2)_{[12]}$ as compare to other models. The source was secondary data obtained from Komfo Anokye Teaching Hospital from 2000 to 2010. The statistical software used is matlab called R command . The thesis is written to analyze malaria rate in Ashanti Region in the case study at Out Patient Department (OPD) of Komfo Anokye Teaching Hospital (KATH). By the use of Box-jenkins ARIMA, $ARIMA(1,0,0)(2,0,2)_{[12]}$ was the best model. By using twenty four months to forecast, it was found out the Out Patient Department(OPD) malaria cases with an efficient administration of the appropriate drugs and good environmental sanitation could be minimized within the next twelve months.

DEDICATION

This book is dedicated to my children and husband whose encouragement and tolerance I have had my education.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

1.0 INTRODUCTION

Malaria is mosquito born diseases and is widespread in tropical and subtropical regions. Every year malaria kills about one and three million people world wide. In Ghana, every three hours somebody die of malaria (MOH). To strengthen the country's prevention and control measures, this study was carried out to develop forecasting and prediction models of malaria cases in the Out Patient Department (OPD) of Komfo Anokye Teaching Hospital (KATH) using time series and ARIMA.

1.1 BACKGROUND TO THE STUDY

Malaria has always been the subject of research for medical practitioners from time immemorial. Many ancient texts, especially medical literature, mention of various aspects of malaria and even of its possible link with mosquitoes and in early man, confronting the manifestations of malaria, attributed the fevers to supernatural influences: evil spirits, angered deities, or the black magic of sorcerers. The ancient Chinese believed the frightening symptoms and signs to be the work of three demons, one with a hammer, one with a pail of cold water, and a third with a stove. The ancient Romans worshiped a fever goddess, three demons rolled into one. The connection between malaria and swamps was known even in antiquity and the evil spirits or malaria gods were believed to live within the marshes.

One of the oldest scripts, written several thousand years ago in cuneiform script on clay tablets, attributes malaria to Nergal, the Babylonian god of destruction and pestilence, pictured as a double-winged, mosquito-like insect. In 800 B.C. the Indian sage Dhanvantari wrote that bites of mosquitoes could causes diseases, fever, shivering etc. The Charaka Samhita written in approximately 300 BC, classified the fevers into five different categories,

namely continuous fevers, remittent fevers, quotidian fevers, tertian fevers and quartan fevers. Susruta Samhita, written about 100 BC, associated fevers with the bites of the insects.

Hippocrates was probably the first malariologist. By 400BC, the author described the various malaria fevers of man. Hippocratic corpus distinguished the intermittent malarial fever from the continuous fever of other infectious diseases, and also noted the daily, every-other-day, and every-third-day temperature rise. The Hippocratic corpus was the first document to mention about splenic change in malaria and also attributed malaria to ingestion of stagnant water: "Those who drink [stagnant water] have always large, stiff spleens and hard, thin, hot stomachs, while their shoulders, collarbones, and faces are emaciated; the fact is that their flesh dissolves to feed the spleen..." Hippocrates also related the fever to the time of the year and to where the patients lived.

The recurrence of malaria is a phenomenon that was known to the ancients and first recorded by Roman Poet Horace (December 8, 65 BC - November 27, 8 BC) in his third satire. Malaria is a mosquito-borne infectious disease caused by a eukaryotic protist of the genus Plasmodium. It is widespread in tropical and subtropical regions, including parts of the Americas, Asia, and Africa. Each year, there are more than two hundred and fifty million cases of malaria, killing between one and three million people, the majority of whom are young children in sub-Saharan Africa Ninety percent of malaria-related deaths occur in sub-Saharan Africa. Malaria is commonly associated with poverty, and can indeed be a cause of poverty] and a major hindrance to economic development.

Five species of the plasmodium parasite can infect humans: the most serious forms of the disease are caused by Plasmodium falciparum. Malaria caused by *Plasmodium vivax*, *Plasmodium ovale* and *Plasmodium malariae* causes milder disease in humans that is not generally fatal. A fifth species, *Plasmodium knowlesi*, is a zoonosis that causes malaria in

macaques but can also infect human. Malaria is naturally transmitted by the bite of a female *Anopheles* mosquito. When a mosquito bites an infected person, a small amount of blood is taken, which contains malaria parasites. These develop within the mosquito, and about one week later, when the mosquito takes its next blood meal, the parasites are injected with the mosquito's saliva into the person being bitten. After a period of between two weeks and several months (occasionally years) spent in the liver, the malaria parasites start to multiply within red blood cells, causing symptoms that include fever, and headache. In severe cases the disease worsens leading to hallucinations, coma, and death. Malaria parasites contain apicoplasts, an organelle usually found in plants, complete with their own functioning genomes. These apicoplast are thought to have originated through the endosymbiosis of algae and play a crucial role in various aspects of parasite metabolism e.g. fatty acid bio-synthesis To date, 466 proteins have been found to be produced by apicoplasts and these are now being looked at as possible targets for novel anti-malarial drugs

The parasite's primary (definitive) hosts and transmission vectors are female mosquitoes of the *Anopheles* genus, while humans and other vertebrates are secondary hosts. Young mosquitoes first ingest the malaria parasite by feeding on an infected human carrier and the infected *Anopheles* mosquitoes carry *Plasmodium* sporozoites in their salivary glands. A mosquito becomes infected when it takes a blood meal from an infected human. Once ingested, the parasite gametocytes taken up in the blood will further differentiate into male or female gametes and then fuse in the mosquito gut. This produces an ookinete that penetrates the gut lining and produces an oocyst in the gut wall. When the oocyst ruptures, it releases sporozoites that migrate through the mosquito's body to the salivary glands, where they are then ready to infect a new human host. This type of transmission is occasionally referred to as anterior station transfer The sporozoites are injected into the skin, alongside saliva, when the mosquito takes a subsequent blood meal.

Only female mosquitoes feed on blood, thus males do not transmit the disease. The females of the *Anopheles* genus of mosquito prefer to feed at night. They usually start searching for a meal at dusk, and will continue throughout the night until taking a meal. Malaria parasites can also be transmitted by blood transfusions, although this is rare. The life cycle of malaria parasites in the human body. A mosquito infects a person by taking a blood meal. First, sporozoites enter the bloodstream, and migrate to the liver. They infect liver cells (hepatocytes), where they multiply into merozoites, rupture the liver cells, and escape back into the bloodstream. Then, the merozoites infect red blood cells, where they develop into ring forms, trophozoites and schizonts which in turn produce further merozoites. Sexual forms (gametocytes) are also produced, which, if taken up by a mosquito, will infect the insect and continue the life cycle. Malaria in humans develops via two phases: an exoerythrocytic and an erythrocytic phase. The exoerythrocytic phase involves infection of the hepatic system, or liver, whereas the erythrocytic phase involves infection of the erythrocytes, or red blood cells. When an infected mosquito pierces a person's skin to take a blood meal, sporozoites in the mosquito's saliva enter the bloodstream and migrate to the liver. Within 30 minutes of being introduced into the human host, the sporozoites infect hepatocytes, multiplying asexually and asymptomatically for a period of 6–15 days. Once in the liver, these organisms differentiate to yield thousands of merozoites, which, following rupture of their host cells, escape into the blood and infect red blood cells, thus beginning the erythrocytic stage of the life cycle The parasite escapes from the liver undetected by wrapping itself in the cell membrane of the infected host liver cell.

Within the red blood cells, the parasites multiply further, again asexually, periodically breaking out of their hosts to invade fresh red blood cells. Several such amplification cycles occur. Thus, classical descriptions of waves of fever arise from simultaneous waves of merozoites escaping and infecting red blood cells. Some *P. vivax* and *P. ovale* sporozoites do

not immediately develop into exoerythrocytic-phase merozoites, but instead produce hypnozoites that remain dormant for periods ranging from several months (6–12 months is typical) to as long as three years. After a period of dormancy, they reactivate and produce merozoites. Hypnozoites are responsible for long incubation and late relapses in these two species of malaria.

The parasite is relatively protected from attack by the body's immune system because for most of its human life cycle it resides within the liver and blood cells and is relatively invisible to immune surveillance. However, circulating infected blood cells are destroyed in the spleen. To avoid this fate, the *P. falciparum* parasite displays adhesive proteins on the surface of the infected blood cells, causing the blood cells to stick to the walls of small blood vessels, thereby sequestering the parasite from passage through the general circulation and the spleen. This "stickiness" is the main factor giving rise to hemorrhagic complications of malaria. High endothelial venules (the smallest branches of the circulatory system) can be blocked by the attachment of masses of these infected red blood cells. The blockage of these vessels causes symptoms such as in placental and cerebral malaria. In cerebral malaria the sequestrated red blood cells can breach the blood brain barrier possibly leading to coma. Although the red blood cell surface adhesive proteins (called PfEMP1, for *Plasmodium falciparum* erythrocyte membrane protein 1) are exposed to the immune system, they do not serve as good immune targets, because of their extreme diversity; there are at least 60 variations of the protein within a single parasite and effectively limitless versions within parasite population The parasite switches between a broad repertoire of *Plasmodium falciparum* erythrocyte membrane protein 1(PfEMP1) surface proteins, thus staying one step ahead of the pursuing immune system.

Some merozoites turn into male and female gametocytes. If a mosquito pierces the skin of an infected person, it potentially picks up gametocytes within the blood. Fertilization and sexual

recombination of the parasite occurs in the mosquito's gut, thereby defining the mosquito as the definitive host of the disease. New sporozoites develop and travel to the mosquito's salivary gland, completing the cycle. Pregnant women are especially attractive to the mosquitoes,and malaria in pregnant women is an important cause of stillbirths, infant mortality and low birth weigh particularly in *P. falciparum* infection, but also in other species infection, such as *P. vivax*

Throughout the world, about one hundred and thirty million babies are born worldwide every year and approximately four million of the above figure are stillborn. It is estimated that more than 98% of these occur in developing countries whilst stillbirth accounts for more than half of perinatal mortality in developing countries. In Sub-Saharan Africa, stillbirths account for more than 3% of deliveries each year. Countries in South-East Asia report the highest overall numbers of stillbirth, The average stillbirth rate in developing countries has been reported to be around 26 per 1000 live births, about five times higher than in developed countries (5 per 1000). One fourth to one third of all stillbirths is estimated to take place during delivery. Stillbirths occurring in the intrapartum period generally have a normal appearance and are often called "fresh" stillbirths. The skin not being intact implies death more than twenty four hours before delivery (antepartum), often called "macerated" stillbirths

Stillbirths have not been widely studied, have been under-reported, and rarely have been considered in attempts to improve birth outcomes in developing countries. There are many factors associated with stillbirth including inadequate access to obstetric care, inadequate care, malaria, hypertensive disease, poor nutritional status, history of stillbirth, congenital anomalies, sickle cell disease, and high burden of infectious comorbidities. Conceptually, infection may result in fetal death through several pathways, First, maternal infection may cause severe illness, leading to fetal death. Also, an infection in the uterus or anywhere else in the mother's body may precipitate preterm labor. Last, the placenta may be directly

infected, leading to reduced blood flow to the fetus, a likely cause of stillbirth associated with malaria infection.. When malaria parasites infect the placenta, placental insufficiency results because of lymphocyte and macrophage accumulation, and increased expression of pro-inflammatory cytokines; these impede maternal blood flow through the placenta. Intestinal helminths, including hookworms and *Trichuris trichura,* have been associated with anemia.

A study in Tanzania showed that 63 percent of stillbirths were attributable to maternal anemia. It has been suggested that low hemoglobin concentrations can cause a state of chronic hypoxia, which is presumably exacerbated in pregnancy when oxygen demands are particularly high because of the metabolism of the mother and the fetus, and that oxygen transfer to the fetus is probably reduced in anemic women. Folate deficiency causes megaloblastic anemia. Circulating folate concentrations decline in pregnant women, hence the need for supplementation. A strong association has been observed between maternal plasma, cord plasma, and placental folate concentrations, suggesting that transplacental folate delivery depends on maternal plasma folate concentrations.

According to the World Health Organization's Opportunities for Africa's Newborns 2006 report, the stillbirth rate for Ghana is 24 per 1000 deliveries. Even though stillbirths represent a large proportion of perinatal deaths, causes of stillbirths are poorly understood in Ghana. To our knowledge, the association between malaria and intestinal helminth coinfection in pregnancy and stillbirth has not been studied. Few studies have studied the association between malaria and helminths in pregnancy, with conflicting results. This study provides baseline data in this area. Given that 98 percent of stillbirths occur in developing countries, especially sub-Saharan Africa [which also has a high burden of malaria and intestinal helminth infections, it is important to investigate the role of these infections in contributing to stillbirth.

A wide variety of antimalarial drugs are available to treat malaria. In the last 5 years, treatment of *P. falciparum* infections in endemic countries has been transformed by the use of combinations of drugs containing an artemisinin derivative. Severe malaria is treated with intravenous or intramuscular quinine or, increasingly, the artemisinin derivative artesunate.Several drugs are also available to prevent malaria in travellers to malaria-endemic countries (prophylaxis). Resistance has developed to several antimalarial drugs, most notably chloroquine. Malaria transmission can be reduced by preventing mosquito bites by distribution of inexpensive mosquito nets and insect repellents, or by mosquito-control measures such as spraying insecticides inside houses and draining standing water where mosquitoes lay their eggs.

Although many are under development, the challenge of producing a widely available vaccine that provides a high level of protection for a sustained period is still to be met. The problem of coming out with a reliable vaccine to combat this menance has nessciated a study into malaraia and the need to forecast malaria rate in Ashanti Region using ARIMA model.

1.2 STATEMENT OF THE PROBLEM

Malaria is frequently referred to as a disease of the poor or a disease of poverty. A better understanding of the linkages between malaria and poverty is needed to guide the design of coherent and effective policies and tools to tackle malaria and poverty together. The malaria burden is a challenge to human development. It is both a cause and consequence of under-development. In Ghana, malaria is the number one cause of morbidity accounting for 40-60% of outpatient. It is therefore a major public health problem in terms of both morbidity and mortality with around 2.5 millions clinical cases and more than 15.000 deaths each year in the country. It is also the leading cause of mortality in children under five years, a significant cause of adult morbidity, and the leading cause of workdays lost due to illness. Despite its devastating effects, the importance of a malaria-free environment in promoting economic

development and poverty reduction has not been fully appreciated in Ghana. Perhaps the reason may be that the impact of the burden of malaria has not been demonstrated in quantitative terms to convince politicians, policy makers, programme managers and development partners to devote the needed attention to this dreadful disease. It is the single main cause of mortality among pregnant women and children below the age of five, accounting for more than 50 % of all cases. It continues to ravage millions of rural Ghanaians, despite concerted efforts to reduce malaria mortality. This is often attributed to a number of factors such as the limited access to basic health care due to poverty, limited specialized health facilities, the cost-sharing system, and underfunding of the health sector by the government.

Malaria contributes substantially to the poor health situation in Africa. It is on record that, Sub-Saharan Africa accounts for 90% of the world's 300 – 500 million cases and 1.5 – 2.7 million deaths annually. About 90% of all these deaths in Africa occur in young children. This has serious demographic consequences for the continent. Between 20 and 40 percent of outpatient visits and between 10 and 15 percent of hospital admissions in Africa are attributed to malaria (WHO, 1999). This burdens the health system. In general, it is estimated that malaria accounts for an average of 3% of the total global disease burden as a single disease in 1990. In Sub-Saharan Africa (SSA), 10.8% of all Disability – Adjusted live years (DALYs) were lost to malaria in 1990. Again, among the ten leading causes of DALYS in the world in 1998, malaria ranked eighth with a share of 2.8% of the global disease burden.

In South Saharan Africa, however, Malaria is ranked second after HIV/AIDS accounting for 10.6% of the disease burden. According to the World Bank, Malaria accounted for an estimated 35 million DALYs lost in Africa in 1990 due to ill health and premature deaths (World Bank, 1993). This loss was again estimated at 39 million DALYs in 1998 and 36 million DALYs in 1999 (WHO, 1998, 1999, 2000). Furthermore, while malaria contributed

2.05% to the total global deaths in 2000, it was responsible for 9.0% of all deaths in Africa (WHO, 2002). The World Health Organization also estimated that the total cost of malaria to Africa was US$ 1.8 billion in 1995 and US$ 2 billion in 1997 (WHO, 1997).

Malaria is a massive problem, which plagues all segments of the society. The effect of malaria on people of all ages is quite immense. It is however very serious among pregnant women and children because they have less immunity. When malaria infection is not properly treated in pregnant women, it can cause anaemia and also lead to miscarriages, stillbirths, underweight babies and maternal deaths. Also, frequent cerebral malaria can lead to disabling neurological sequelae. Further, malaria in school children is a major cause of absenteeism in endemic countries. It is estimated that about 2% of children who recover from cerebral malaria suffer brain damage including epilepsy (WHO/UNICEF, 2003). Hence, among young children, frequent episodes of severe malaria may negatively impact on their learning abilities and educational attainment. This is a threat to human capital accumulation, which constitutes a key factor in economic development. The debilitating effects of malaria on adult victims are very much disturbing. In addition to time and money spent on preventing and treating malaria, it causes considerable pain and weakness among its victims. This can reduce peoples working abilities. The adverse impact of the disease on household production and gross domestic product can be substantial.

Fontaine et al., (1961) indicated a well-documented major epidemic which resulted in an estimated 3 million cases of which 150 000 died. He explained further that such a large-scale epidemic has been known to return at some irregular intervals of years; for example, during the 1980s and 1990s, severe epidemics were recorded in 1981, 1988, 1991, 1992 and 1998. It had not been possible to forecast any of these events, especially in highland areas which are normally considered non-malarious or had very low transmission. As an example, in 1988, a major epidemic affected most highland areas in the country following normal or below

normal transmission the previous year (manuscript in preparation). A similar epidemic in 1998 resulted in high levels of mortality in highland areas where the disease had been absent for years. In areas with unstable transmission, setting up systems for epidemic early warning has become essential. The quantification and use in early warning of the effect of epidemic-precipitating factors such as weather patterns has been difficult in epidemic-prone areas where slight changes might cause devastating epidemics. Currently there are efforts to develop early warning systems that use weather monitoring and climate forecasts and other factors (Thomson and Connor 2001). In some countries, epidemics have been associated with the occurrence of the weather phenomenon known as the El Niño (Bouma & van der Kaay 1996; Bouma and Dye 1997). While predicting El Niño years is not a simple task, even such predictions would be too global in nature to be useful as an early warning in specific areas. Moreover, there are variations in the magnitude and timing of the effects of El Niño on malaria incidence according to geographical conditions.

Despite considerable efforts throughout the century to eradicate or control malaria it is still the most prevalent and most devastating disease in the tropics. The disease has a crippling effect on the economic growth and perpetuates vicious cycles of poverty. It costs Africa US$10 – 12 billion every year in lost domestic product even though it could be controlled for a fraction of that sum (UNICEF 2004). In Africa, malaria causes approximately 20% of cerebral conditions leading to coma and death. One important strategy to prevent people from the risk of malaria infection is the use of insecticide treated mosquito net (ITNs). Recent studies have shown that the use of bed nets, especially the ITNs can reduce both transmission and mortality by at least 25% when used properly (Sagoe-Moses 2005).

The 2003 Ghana Demographic Health Survey (GDHS) revealed that among the 6,251 households surveyed 17.6% had a bed net and only 3.2% had ITNs. Malaria presents a serious health problem in Ghana; it is hyperendemic with a crude parasite rate ranging from

10 – 70% and plasmodium falciparum the major malaria parasite, dominating3. Although numerous efforts have been made to fight malaria in the country as in many endemic countries, achievements have been minimal. It is the number one cause of morbidity accounting for 40% of outpatient attendance with annual reported cases of 2.2 million between 1995 -2001, and over 10% ending up on admission (Kusi 2003). From the UN classification of childhood diseases it ranks third in Africa (Ministry of Health 2002).

It is known that initial management and treatment of the disease often takes place outside formal service without consulting trained professionals largely because of transportation, finance, and time, social and physical distance, among others. Treatment received from these sources is often inappropriate leading to clinical complications such as severe anaemia, cerebral malaria, which may lead to coma and/or other parasitological complication such as the development of drug resistant parasites, whose emergence in the country in 1986 has been a major obstacle to the control strategy of prompt recognition and adequate treatment. It is also the leading cause of workday loss due to illness in the country. For instance, it accounts for 3.6 ill days in a month, 1.3 workdays absent and 6.4% of potential income loss to Ghana for 1998/99. The disease is again responsible for 10.2% of all healthy life lost from other diseases making it the chief cause of lost days of healthy life in Ghana it concluded. As part of measures to eradicate the disease, WHO initiated a ``Roll Back Malaria`` project, of which Ghana is benefiting, to expand availability and coverage of insecticide-treated mosquito nets, which includes forecasting and procurement of these nets by Non Governmental Organizations (NGOs).

As a result of this effort, the Ministry of Health has drafted a policy on insecticide treated bed nets, and is adopting these nets as an additional control measure to back the RBM project (Chinbuah 1999). Almost everybody in Ghana is vulnerable to the disease and about 50% of Ghanaians spend money on mosquito control products such as coils, sprays, fly proof nets,

mosquito repellent etc, while almost every household spend money on the curative treatment of malaria (Ghana Social Marketing Foundation 1999). Since the cost of treatment of the disease is directly proportional to the size of the potential benefits to be derived for the country, for a successful malaria control programme, this study will try to identify areas of high prevalence of malaria in the region. Without a continuous assessment or provision of baseline data, the Region will not be able to effectively plan, implement and evaluate programmes.

The direct economic costs of malaria that result from treatment and from time away from work or school are enormous, but the overall economic impact of malaria is likely to be much more substantial than suggested by estimates of direct costs alone Malaria therefore is not only a public health problem but also a developmental problem. At the national level, apart from the negative effect of lost productivity on the major sectors of the economy, malaria has negative effects on the growth of tourism, investments and trade especially in endemic regions. Malaria still remains a public health problem in Ghana despite marked reduction of cases in last few years.

It can so far be generalized that very young children and pregnant women are the population groups at highest risk with respect to malaria morbidity and mortality. Most children experience their first malaria infections during the first year or two of life, when they have not yet acquired adequate clinical immunity, which makes these early years particularly dangerous. Adult women have a high level of immunity, but this is impaired especially in the first pregnancy, with the result that risk of infection increases. Pregnant women and infants may experience a variety of adverse consequences from malaria infection including maternal anaemia, placental accumulation of malaria parasites, low birth weight from prematurity and intrauterine growth retardation, foetal parasite exposure and congenital infection, leading to increased risk of maternal and infant mortality. To strengthen the country's prevention and

control measures, this study was carried out to develop forecasting and prediction models of malaria incidence in Ghana using time series model from 2000 to 2010.

1.3 OBJECTIVES

The objectives of the thesis are:

(i) to build a times series data for Malaria in Ashanti Region and

(ii) to model Malaria rate from in Ashanti Region from 2000 to 2010

1.4 METHODOLOGY

The Box- Jenkins ARIMA approach is the methodology that will be employed for the study. The selected model will be compared to other models and the best model will be selected made on its forecasting performance. The data source is a secondary data obtained from the Komfo Anokye Teaching Hospital from 2000 to 2010. The statistical software that will be used is Matlab code called R command.

1.5 JUSTIFICATION

From a macroeconomic perspective, malaria mortality and morbidity have been observed to slow economic growth by reducing capacity and efficiency of the labour force. Basic economic theory postulates that the quantity of a given output that is produced is a function of several factors including the capital stock, labour force and the quality of labour available. Based on this, it could be argued that the effects of malaria on labour diminishes total output and for that matter national income. Gallup and Sachs (2001) in a cross-country econometric estimation of the effects of malaria on national income concluded that countries with substantial level of malaria grew 1.3% less per person per year for the period 1965 - 1990. The study also confirmed that a 10% reduction in malaria was associated with 0.3% higher growth in the economy.

It is therefore justifiable based on the devastating effects of malaria across the length and breadth of the country, hence modeling malaria using the Box- Jenkins ARIMA approach will provide stakeholders with reliable forecast which can be used to predict malaria rate in the country. It is again to help Stakeholders gain useful information through the models which are capable of predicting malaria

1.6 LIMITATION OF THE STUDY

The following are some of the limitations encountered in the course of the study; Lack of econometric software to analyzed data, Unwillingness on the part of institutions concerned to release data and lack of accurate data and monitoring of patients affected with malaria.

1.7 ORGANIZATION OF THESIS

In chapter one, we considered the background of the study, the problem stated and the objective of the study. The methodology, justification and limitation of the study were also put forward.

Chapter two presents the review of relevant literature. Chapter three is denoted for the methodology of the study. In chapter four we shall put forward the data collection and analysis of the study. Chapter five, the final chapter of the study presents the conclusion and recommendations of the study.

1.8 SUMMARY

In this chapter, we considered the background of the study, the problem statement, objectives of the study, the methodology, justification of the study and limitations of the study.

In the next chapter, we shall put forward pertinent literature on time series analysis.

CHAPTER TWO

LITERATURE REVIEW

2.0 INTRODUCTION

This chapter reviews the relevant literature and applications on time series and forecasting Time series forecasting has the largest literature and number of applications of any approach to forecasting. Production planning, budgeting, inventory management, sales, marketing, and distribution all depend on accurate, short-term, time series forecasts. Many policy-level decisions in the energy, tourism, agricultural, and other areas depend on multivariate time series forecasts. While several avenues have been explored to improve time series forecasting, one of the most promising – pooling data from analogous time series – has had but scant attention (Fildes and Beard 1992). Conventional time series methods – exponential smoothing, Kalman filters, Box-Jenkins ARIMA methods, Census X11, Parzen's ARARMA method (Parzen 1982), multiple regression methods, etc.– all forecast single series in isolation. This chapter is subdivided into the followings areas: Malaria and its effects and Malaria models

2.1 MALARIA AND ITS EFFECTS

This section deals with the review of literature associated with the effects of malaria on an individual and the nation as a whole. Malaria is a parasitic disease that affect humans. It effects lead to acute symptoms of infection such as fever, headache, and nausea. The main chronic symptom is anemia. Malaria results in death in some cases, when the disease is caused by the vivax parasite and to a lesser extent a low fatality rates when the disease is caused by the falciparum parasite. The parasite has a complicated life-cycle that is partly spent in a mosquito vector and partly in the human host. The disease is transmitted when a

mosquito takes a blood meal from an infected person and, some time later, bites another person. Because of the crucial role played by mosquitoes in the transmission cycle, warm and wetter climates are more likely to sustain endemic malaria.

Numerous studies have taken a look at the effects of malaria and one of such include Conly (1975) who in his analysis dealt with the spillovers within a particular household (e.g. parents' caring for sick children) in Paraguay and concluded that time allotted for caring for the sick has significant impact on production. Although environmental management projects to control malaria should be aimed at reducing anopheline density.

Lacey and Lacey, (1990) discovered in rural areas that, fish may be appropriate components of malaria control if breeding sites are well known and limited in number. They concluded that use of fish may be less feasible where natural breeding sites are extremely numerous. Brinkman and Brinkman (1991) estimated that approximately 40% of fever can be attributed to malaria in Africa, though this figure will vary across transmission zones.

Research by Fletcher et al., (1992) suggested that the control of nuisance-biting species is also important for fostering community support and satisfaction. Meltzer (1992), for example, showed that health at different points in the life cycle will impact investment and fertility decisions differentially. Furthermore, Malthusian effects (Acemoglu and Johnson's main interpretation for their results) will be attenuated for a disease with comparatively high childhood morbidity, such as is the case for the malaria in the Americas.

More recently, time-series analyses have shown that the risk of a malaria epidemic increased approximately five-fold during the year following an El Niño event in the Indian region (Bouma and van der Kaay, 1994). Experience in community-based programs to control Chagas' disease through housing improvements in Brazil, Venezuela, and Bolivia suggests that developing a local organizational structure may have many benefits (Bryan et al., 1994).

The researchers found that the active participation of community members made the interventions more cost-effective than the traditional vertically organized program and also fostered community pride and solidarity. Najera (1994) argues that the disappearance of malaria in parts of Europe was associated with economic development related to agricultural expansion rather than vector control or chemoprophylaxis.

McCombie,(1996), In his contribution to the effects of malaria argued that biasness arises from using clinic or hospital based data because only around 20-40% of malaria cases and deaths are estimated to receive treatment in formal health facilities. Most malaria cases are diagnosed and treated in the home. As a result, data from health facilities reflects any inequity in access to those facilities, though the extent of the inequity cannot be determined without comparisons to population based data Akhtar and McMichael (1996), discovered a strong correlation between annual rainfall and the number of rainy days and the incidence of malaria in most districts of Rajasthan and in some districts. House spraying programs are most likely to be effective when the principal vectors are endophilic and endophagic and where strong financial support can ensure timely chemical application by well-trained operators, using appropriate equipment and insecticides (Chavasse and Yap, 1997).

Rozendaal (1997), reacted to Chavasse contributiob by saying that, even when insecticides are applied as directed, other factors, such as the type of house structure and building material, can reduce the impact of Indoor Residual Spraying(IRS). He further commented that insecticides tend to last longest on wooden or painted walls and to breaks down quickly on mud walls with a high clay content. Gujarat Akhtar and McMichael. Ribbands, 1998 indicated that removing vegetation from around houses may control mosquitoes by removing resting sites. He later realized that such practice does not work or had no effect on anopheline. Lindsay and Martens (1998) considered the progressive rise in the incidence of

malaria over the last decades in African highlands. The authors use mathematical models to identify epidemic-prone regions in highlands Africa, and to quantify the difference expected to occur as a consequence of projected global climate change. To make estimates about the areas that are vulnerable to epidemic outbreaks of malaria, they use data and models from Geographic Information Systems (GIS) (computerized mapping systems) and Remotely Sensed (RS) imagery data from earth-orbiting satellites. Correlations among variables are found. However, the authors observe that since correlation doesn't imply causality the results are not conclusive and require further investigation.

The effects of soil moisture to determine the causal links between weather and malaria transmission has been studied by Patz et al, (1998). He discovered that the most common mosquito species is Anopheles gambiae, which accounts for about 45% and 56% of the variability of human biting rate and entomological inoculation rate, respectively. In rural Gambia, Lindsay and Snow (1988) found that children sleeping in houses with closed eaves and metal roofs experienced fewer malaria infections than children sleeping in houses with open eaves. The authors concluded however that not all forms of housing improvement will reduce malaria transmission. In his attempts to test for causality in the spread of malaria, Chowdhury (1988) finds no consistent results across his thirty-five country sample. Fourteen of his cases support the hypothesis that infant mortality causally impacts fertility, while only two cases support the opposite hypothesis. The remaining cases indicate feedback between the two variables, or the absence of a relationship between fertility and mortality. It can be acknowledged that his results provide stronger support for the hypothesis that mortality affects fertility, but they are notably (and admittedly) inconclusive.

Gratz and Pal (1988), focused on using Chemical larviciding and biological control, particularly using larvivorous fish, concluded that the methods were important to malaria

control programs in the early part of the 20th century, particularly in urban and Periurban areas. Lindsay, 1999 hypothesized that increasing temperatures could be part of the reason why malaria can now survive at higher altitudes. Many other confounding factors, however, could be causing the increase in malaria in these areas. Sanchez and Nunez (2000) considered the role of geographic and health factors in explaining the cross-municipio income differences within Colombia at a point in time, by using local resources in childhood as an instrument for adult height. Nunez again teamed up with Ribero in (2000) analyzed the effect of health endowments on income in Colombia. Gwatkin et al. (2000) used asset data collected in DHS surveys to examine socio-economic differences in health, nutrition and population indicators.

Rashed et al., (2000), conducted a study in Benin by using the results of a malaria KAP survey to analyse the relationship between a number of demographic and social variables (including income and expenditure) and incidence of febrile episodes in children. They concluded that only Insecticide Treated Nets(ITN) use and the age of the child were found to be significantly related to fever incidence. Sychareun et al., (2000) conducted a test in Lao PDR using questionnaires to obtain income information from pregnant women in a remote district hospital. They observed a significant difference in malaria prevalence between socio-economic groups, where 87.5% of women with positive slide parasitaemia were classified as having low income whilst only 12.5% of the women were classified as having high income. They concluded that with only 16 women testing positive for malaria and the majority of women sampled - regardless of malaria diagnosis - falling into the low-income category, the results were not statistically significant. This was supported by Biritwum et al., (2000) conducted a study in Ghana by comparing malaria incidence in two communities, one with relatively low average income of 73,824 cedis per year and one with relatively higher average income of 138,167 cedis per year. He realized that not all income-based studies have been

consistent in their results, despite lower formal education levels among caregivers and fewer children in nursery. He however discovered that malaria incidence in the poorer community was not significantly different from that in the richer community, though incidence of ill-health in general was higher among the poor than the rich.

Doumani et al., (2001) argued that the control of malaria transmission was a key factor in the development of Zambian copper mining, Venture, 2001contributed to malaria rate and poverty by arguing that because the burden of malaria is concentrated in poor countries there is inequity in allocation of global research funds especially by the pharmaceutical industry, since domestic purchasing power for new malaria products is very limited especially for anti malarial drugs. Abdulla et al. 2001, studied malaria level in Tanzania and used a data from a demographic surveillance site (DSS) in rural Tanzania. He also used asset index to measure SES, which supports Filmer's finding, revealing that fever seems to affect the poor and less poor approximately equally.

In 2002, Salomon and Murray analyzed the patterns of diseases and mortality rates in the framework of the literature on epidemiologic transition. Tol (2002) in his contribution considered the General Circulation Model's results to estimate (and evaluate in monetary terms) the impacts of climate change for a wide range of market and non-market sectors (agriculture, forestry, water, energy, costal zones and ecosystems, as well as mortality due to vector-borne diseases, heat stress and cold stress). He discovered that small increases in temperatures would bring some benefits (mainly for the developed world). He concluded that the global impact of climate change depends crucially on the weights used to aggregate the regional values. Shanks et al. (2002) investigate whether the reemergence of malaria in Western Kenya could be attributed to changes in meteorological conditions. The existence of trends in a continuous 30-year monthly malaria incidence dataset (1966–1995) is tested for.

Sachs and Malaney (2002) demonstrated a correlation between the presence of malaria in a country and that country's per capita GDP. They argued that there is an inverse relationship between the two and that malaria causes underdevelopment

Malaria incidence increased significantly during the 1966–1995 period. In contrast, no aspect of climate is found to have changed significantly-neither the temperature extremes (maximum and minimum) nor the periods when meteorological data were transformed into months when malaria transmission is possible. Therefore, the authors conclude that climate changes have not caused the highland malaria resurgence in Western Kenya. They suggest that two other factors may have influenced the increase in malaria hospitalizations: an increase in malaria severity indicated by an increased case-fatality rate (from 1.3% in the 1960s to 6% in the 1990s) that is most likely linked to chloroquine resistance. Secondly, travel to and from the Lake Victoria region by a minority of the tea estate workers also exerts an upward influence on malaria transmission in Kericho, Kenya, since such travel increases the numbers of workers asymptomatically carrying gametocytes, which infect. Guerin et al., (2002) concluded that without the necessary treatment, however, malaria causes recurring and debilitating infections, which increase vulnerability to other diseases, affect educational performance of children, and reduce labor market productivity of adults (Guerin et al., 2002).

However, recent evidence from another Tanzanian DSS site by Savigny et al., 2002 indicates that the poorest infants and children under five -years of age had higher risks of death than those in the least-poor socio-economic quintiles. They emphaised that One of reasons for this conflicting results may be that there is insufficient variation in socio-economic status in some areas to allow significant differences to be detected. Malaney, (2003) in his contribution explained that, the fever is simply the most acute realization of morbidity and can cause anemia which can persists for a considerable time after infection

Since malaria infection can cause anemia that persists for a considerable time after infection, Thomas et al., (2003) showed that anemia has depressing effects on contemporaneous adult productivity and physical capital and land improvement are suppressed by the threat of malaria. Lopez and Porras (2003) and Meisel (2004) used data on stature to analyze changes in health across birth cohorts in Mexico and Colombia, respectively. Malaney, Spielman and Sachs (2004) found a correlation between economic growth rates and the presence of malaria is not sufficient to establish the direction of casualty since the presence of malaria is endogenous.

Hanson et al., (2004) contributed to the effects of malaria by saying that one of its barriers associated to the preventive services is the lack of information on the part of consumers Hanson et al., (2004). Kovats et al., (2005) considered the WHO 2004 estimates and remarked that to generate consistent estimates the models need to incorporate: geographical variation in the vulnerability to climate; future changes in the disease rates due to factors other than climate (e.g. decreases rates of infectious diseases due to technological advances); assumptions on a country's ability to control a disease such as malaria, dengue fever or diarrheal disease; uncertainties around the exposure-response relationship. Moreover, they claim that, even controlling for the above mentioned (potentially positive or negative) issues, no model can take into account the possibility of irreversibility or plausible low probability events with potentially high impact on human health. As a main consequence, threshold health effects to regulate "tolerable" amount of climate change cannot be identified.

Lucas (2005) indicated that, women born after malaria eradication in Sri Lanka completed more years of schooling, suggesting that returns to education rose faster than child wages in that episode. Llanos-Zavalaga et. al 2005 also conducted a study on the use of bed nets in Piura and realized that one of the factors associated with the non-use of bed nets was the lack

of comprehensive knowledge of the disease. Acemoglu & Johnson (2005), find no discernible gain in education attainment with the improvements in the disease environment that resulted from international health interventions.

Acemoglu and Johnson (2006) show that the global decline in mortality from several diseases (including malaria) in the 1950s increased life expectancy in developing countries. The authors argued, however, that this increase in life expectancy was not matched with an increase in income, relative to high-income countries. Nevertheless, their study does not analyze the impact of malaria separately from other mortality risks. Furthermore, it relies on an instrument that is mechanically weighted towards diseases with greater mortality (in large measure a mong infants, as it turns out).

Coorey et al., (2006) used transdisciplinary research to develop more effective policies to control malaria, protect forests, and alleviate poverty. They described the malaria problem, including its etiologic roots, and its social toll, by examining some shortcomings of contemporary societal responses. They discussed the need for understanding the role of deforestation which has a link with malaria. Their conclusion was a proposal for strategically linking research and policy, at the "malaria-deforestation-poverty nexus in a comprehensive decision-analysis framework that channels research to the most pressing policy needs, informs policy with the most conclusive research, and ensures stakeholders are effectively informed about their options.

Pascual et al., (2006) analyzed the dynamics of the geographical spread of malaria and focused on the most important malaria species for humans, *Plasmodium falciparum* and *Plasmodium vivax*, whose range is limited at high altitudes by low temperatures. They investigate whether global warming could drive the geographical spread of the disease and produce an increase in incidence at higher-altitude sites. They use data for four high-altitude

sites in East Africa from 1950 to 2006. A nonparametric analysis that decomposes the variability in the data into different components is performed and reveals that the dominant signal in three of the sites and the subdominant signal in the fourth one correspond to a warming trend. To assess the biological significance of this trend, the authors drive a dynamical model for the population dynamics of the mosquito vector with the temperature time series and the corresponding detrended versions. The authors concluded that the observed temperature changes would be significantly amplified by the mosquito population dynamics with a difference in the biological response at least, one order of magnitude larger than that in the environmental variable.

Another interesting study in this context is the one of Bhattacharya et al., (2006) who projected malaria transmission in new geographical regions in India and concluded that the spread of malaria is expected to move from central regions towards South Western and Northern Regions by 2050.

Recently, Hong (2007) in then contribution to the effects of malaria used a database of Union Army veterans and correlated the early-life exposure to malaria with numerous later-life health outcomes in the 19th century America. Still on later life outcome Barreca (2007) considered the effects on later-life outcomes of specially "in utero" exposure to short-term fluctuations in malaria caused by within-year variation in rainfall and temperature in the United States. He concluded that inter annual variation in malaria is likely to have economic impacts that differ from those long-term changes induced by eradication.

Bleakley (2007) finds that hookworm eradication in the U.S. South was followed by an increase in school attendance, literacy, and income.

Dalton et.al., (2007) worked on estimation (particularly on the effect of the child mortality variable) by deploying exogenous variation in the ecology of malaria transmission and in

agricultural productivity through the staggered introduction of Green Revolution, high-yield seed varieties. The authors observed that child mortality (proxied by infant mortality) is by far the most important factor among those explaining aggregate such as total fertility rates, followed by farm productivity. The authors concluded that female literacy and aggregate income do not seem to matter as much, comparatively. Cutler et al. (2007) also analyzed the malaria-eradication campaign in India and concluded that malaria exposure in early life reduced educational attainment.

Gaudert et al.,(2009) conducted a study on the risk of *Plasmodium falciparum* infection, which is variable over space and time. The authors used remote-sensed data coupled with field study data in order to drive a malaria transmission mode. They observed that environmental factors affect the biological cycle of both vector and parasite. Their conclusion was that non-linear mode, combining with environmental variables, predisposition factors and transmission pattern can be used for community.

Asenso et al., (2010) surveyed the linkages between malaria and agriculture, by focusing on the economic impacts of the disease. They examined the potential impacts of malaria in the agricultural sector, including direct and indirect impacts and observed that in addition to health care costs, malaria causes loss to agricultural labor and slows adoption of improved practices in agriculture. They further identified some gaps in malaria combating and concluded that a better understanding of malaria's impact on agricultural productivity, coupled with efforts to strengthen capacity to deal with malaria, would enhance policies and programs aimed at combating malaria and curbing its impacts on agricultural productivity.

2.2 MATHEMATICAL APPLICATION

Statistical models are important tools for analyzing the complex relationship between malaria rate and its effects in an economy like Ghana, since the model allows researchers to link

crucial climate variables (such as temperature, rainy season and other environmental factors.), at global or regional levels to the occurrence of the disease under scrutiny. Several researchers have conducted studies into this area using models such as ARIMA, SARIMA,GARCH,ARFMA, VAR and a host of other related models.

In relation to the climatic factors which leads to malaria outbreak, Keatinge et al. (2000) estimated the heat-related mortality due to climate change in Europe, using time-series data and taking into account the threshold temperature where mortality is lowest. The authors suggested from their findings that, European population have adapted to average summer temperatures, and might adapt to future higher temperatures with only a minor increase in heat-related deaths. The authors suggested further that the process of acclimatization should be taken into account when assessing the impact of heat waves and increased temperatures.

Curriero et al., (2002) used time series analyses to estimate the temperature-mortality association for eleven eastern US cities from 1973 to 1994 by using log-linear models. They observed that current and recent days temperature are associated with mortality, secondly, threshold temperature appears to exist below, hence mortality tends to decrease as temperature increases from the coldest days. Again it is observed that that mortality risk increases as temperature increases and finally, strong association exists between mortality , extreme temperatures and latitude.

Rodó et al., (2002) presented a time-series analysis of the relationship between El Niño/Southern Oscillation (ENSO) and the prevalence of cholera in Bangladesh using mortality data recorded on a monthly period from 1893 to 1940. They used Singular spectrum analysis (SSA) to capture discontinuous dynamics and trends. The technique allowed the decomposing of the irregular dynamics the time series and also isolate the inter-annual variability of the data. Their findings suggested that ENSO is responsible for more than 70%

of the dynamics of the disease. Hijioka et al., (2002) contributed by using multiple regression analysis and included the effect of water supply and sanitation coverage, annual average temperature and per capita GDP. They took into account different IPCC climate scenarios and water-borne diseases with temperature in 14 countries, Their results showed large regional differences in the impacts. Shakoor et al. (2006) in their contribution to mortality associated with malaria used time-series models to analyze mortality due to thermal stresses during heat waves, compared to total mortality occurring throughout the whole summer. They observed this trend in order to understand the fraction of the total impact attributable to temperature extremes.

Bosello et al., (2006) made use of the General Equilibrium Model (GTAP) in an unconventional approach in to analyse how health impacts would affect the general economy. Their aim was to estimate the indirect costs on the economic system derived from the health effects as a result of an increase of one degree Celsius in global mean temperature. They concluded that, the impact on labour productivity and health care expenditures for both the public system and private households, have impact on GDP.

The spatial variation of malaria was analyzed by Kazembe et al., (2006), who examine malaria-related hospital admissions and in-hospital mortalities among children in Africa. The authors used pediatric ward register data from Zomba district, Malawi, between 2002 and 2003, as a case study and developed two spatial models. The first is a Poisson model applied to analyze hospitalization and minimum mortality rates, with age and sex as covariates. The second is a logistic model applied to individual level data to analyze case-fatality rate, adjusting for individual covariates. They concluded that the rate of hospital admission and in-hospital mortality decreases with age.

In the field of climate based Early Warning Systems (EWS), which are used to predict the occurrence of epidemics of infectious diseases, Chaves and Pascual (2007) reviewed and compared linear and non-linear models for forecasting seasonal time series of diseases. They used American cutaneous leishmaniasis, as an example, and concluded that the models are evaluated based on the predictive R2 for forecasting the data "out-of-fit". A different strategy for predicting the pattern of diseases is given by Medina et al., (2007), who investigate the dynamics of diarrhea, acute respiratory infection (ARI), and malaria in Niono,(Mali). The authors observed that in using time-series model, figures that are generated are mostly non-stationarity, they exhibit large inter-annual plus seasonal fluctuations; and requires disease-specific tailoring of forecasting methods. The authors observed that in order to accommodate these characteristics, the non-parametric technique (the multiplicative Holt Winters method (MHW) should be used. This method was accepted because is based on past information and pseudo-parameters initialization. Secondly it points out forecasts which are recursively revised through residuals bootstrap to produce median forecasts and at 95% confidence interval bounds. Finally the method decompose the time series (TS) into level, trend (rate of change), seasonal, and approximately serially uncorrelated residual TS components

Linwei et al., (2008) conducted a study on ecological time series using data from 1971 to 1998. They used Auto-regressive integrated moving average (ARIMA) models to evaluate the relationship between weather factors and malaria incidence. They therefore explored the impact of climate variability on the transmission of malaria in the tropical rain forest area of Mengla County, south-west China. They observed that fog day frequency was for the first time found to be a predictor of malaria incidence in a rain forest area. The one-year delayed effect of fog on malaria transmission may involve providing water input and maintaining aquatic breeding sites for mosquitoes in vulnerable times when there is little rainfall in the 6-

month dry seasons. They concluded that the findings should be considered in the prediction of future patterns of malaria for similar tropical rain forest areas worldwide

Farida et.al., (2010) conducted a study using Provincial malaria epidemiological data from 2004 to 2007 which has been collected by the health posts in 23 provinces. This was used in conjunction with space-borne observations from NASA satellites Specifically, the environmental variables, including precipitation, temperature and vegetation index measured by the Tropical Rainfall. Measuring Mission and the Moderate Resolution Imaging Spectoradiometer, were used. They then employed regression techniques were to model malaria cases as a function of environmental predictors. The authors the entire time series except the last 6 months for training, whilst the remaining 6-month data was also used for prediction and validation. The authors concluded that Vegetation index, in general, is the strongest predictor, reflecting the fact that irrigation is the main factor that promotes malaria transmission in Afghanistan. Further surface temperature is the second strongest predictor whilst precipitation is not shown as a significant predictor, as it may not directly lead to higher larval population. Autoregressiveness of the malaria epidemiological data is apparent from the analysis hence the malaria time series are modelled well, with provincial average R^2 of 0.845. Although the R^2 for prediction has larger variation, the total 6-month cases prediction is only 8.9% higher than the actual cases.

2.3 SUMMARY

In this chapter, we considered the relevant literature review of the study, the effects and its mathematical application of the study.

In the next chapter, we shall put forward the methodology of the study.

CHAPTER THREE

METHODOLOGY

3.0 INTRODUCTION

This section undertakes a review of general literature models or analytical techniques of this study. It is meant to explain some of the terms and methodology used to achieve the results of the study.

3.1 TIME SERIES ANALYSIS

Time series by definition refers to a set of observations taken at specific times usually at equal intervals. A *time series* is a chronological sequence of observations on a particular variable. Usually the observations are taken at regular intervals (days, months, years), but the sampling could be irregular. Common examples of time series are the Dow Jones Industrial Average, Gross Domestic Product, unemployment rate, and airline passenger loads. A time series analysis consists of two steps: (1) building a model that represents a time series, and (2) using the model to predict (forecast) future values. If a time series has a regular pattern, then a value of the series should be a function of previous values. If Y is the target value that we are trying to model and predict, and Y_t is the value of Y at time t, then the goal is to create a model of the form:

$Y_t = f(Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots, Y_{t-n}) + e_t$ .Where Y_{t-1} is the value of Y for the previous observation, Y_{t-2} is the value two observations ago, etc., and e_t represents noise that does not follow a predictable pattern (this is called a *random shock*). Values of variables occurring prior to the current observation are called *lag values*. If a time series follows a repeating pattern, then the value of Y_t is usually highly correlated with $Y_{t-cycle}$ where *cycle* is the number of observations

in the regular cycle. For example, monthly observations with an annual cycle often can be modeled by $Y_t = f(Y_{t-12})$ The goal of building a time series model is the same as the goal for other types of predictive models which is to create a model such that the error between the predicted value of the target variable and the actual value is as small as possible.

The primary difference between time series models and other types of models is that lag values of the target variable are used as predictor variables, whereas traditional models use other variables as predictors, and the concept of a lag value doesn't apply because the observations don't represent a chronological sequence. An example of a time series is the total number of accidents recorded for 12 months and the total number of fire outbreaks during the dry season. A time series can be represented mathematically as $X_1, X_2, X_3, \ldots, X_n$ where X is the data point and 1, 2, 3; …, n are the time steps. There are two types of time series, they are; Deterministic time series which deals with series which can be predicted. This is of the function $Y = f(t)$ Next is stochastic time series which is defined as, a random variable plus an error term.

3.1.2 CONTINUOUS TIME SERIES

A time series is said to be continuous when observation are continuously made in time. When even the measured variable can take only discrete sets of value, the term continuous is still used for series of this type.

3.1.3 DISCRETE TIME SERIES.

A time series is said to be discrete when observations are taken only at specific time which is usually spaced, the term discrete is used for events of this type even when the measured variable is continuous variables. Sample series when continuous values are read off or digitized to give a discrete series.

3.1.4 OUTLIERS

A perfectly valid but extreme observation which may for example, indicate that the data are not normally distributed is called an outliers.

3.2 DEFINTIONS OF TERMS

Time series analysis has several uses, the following are a few of the objectives among others classified as follows : Description, Explanation, Forecasting, Control.

3.2.1 DESCRIPTION

Usually, the first step in analyzing a time series is to plot the data and to obtain simple descriptive measures of the main properties of the series such as: trend and seasonal variations.

Not only will a graph show trend and seasonal variation, but it also enables one to look out for "wild observation and outliers". Another feature to look out for in a time series graph is the possible presence of turning points, where for example upward trend suddenly changes to a downward trend.

3.2.2 EXPLANATION

When observations are taken on two or more variables, it may be possible to use the variables in one time series to explain the variations in another series. This leads to a deeper understanding mechanism which generated a given time series.

Multiple regression, models and linear systems may be helpful here. A linear system converts an input series to an output series by a linear operation as shown below in figure 3.1 as shown below. In the figure below a data is inputted into the system through a linear system to give an output.

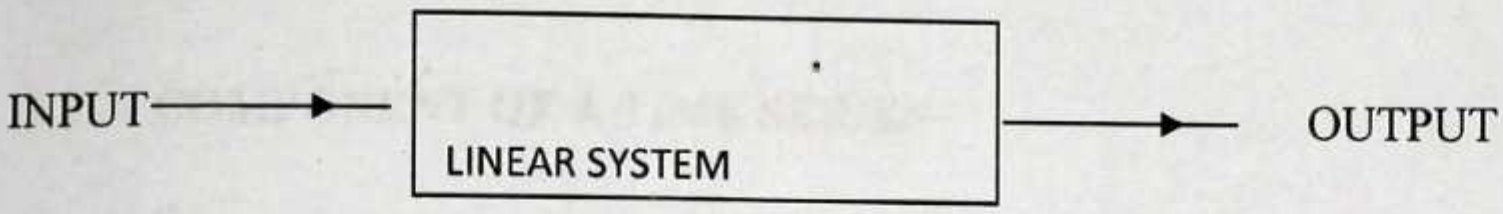


Figure 3.1 A Linear system of an input and output Time series

3.2.3 FORECASTING AND PREDICTING

Given an observed time series, one can predict the future values of the series. This is important in sales forecasting and in the analysis of economic and industrial time series. Prediction and Control are closely related. For example if one can predict that, the spread of diseases to move up, then appropriate corrective action can therefore be taken in that regard.

3.2.4 CONTROL

Control procedures are of several kinds. In statistical quality control, the observations are plotted on control charts and the controller takes action as a result of studying the charts. A more supplicated control strategy has been described by Box Jenkins (1970). This emphasizes a step by step approach of going through the control procedures. It is however emphasized that a time series which is generated to measure "quality" of manufacturing is aimed at controlling the process.

3.3 COMPONENT OF A TIME SERIES

The decomposition of the variation in a series into Trend, Seasonal effect, Cyclic, change and Irregular fluctuation is the main concern in time series analysis.

3.3.1 TREND

A trend may be loosely as "long term in the mean level". In speaking of trend, we must take into account the number of observations available and make a subjective assessment of what is "long term". For example climate variation sometimes exhibit cyclic variation over a very long period time such as 50 years. If one just had 20yrs data, this long term oscillation would appear to be a trend, but if several 100years data were available the long term oscillation would be visible. Nevertheless in the short term it may still be more meaningful to think of such a long term oscillation as a trend. An example of an increasing trend is shown in figure 3.2.

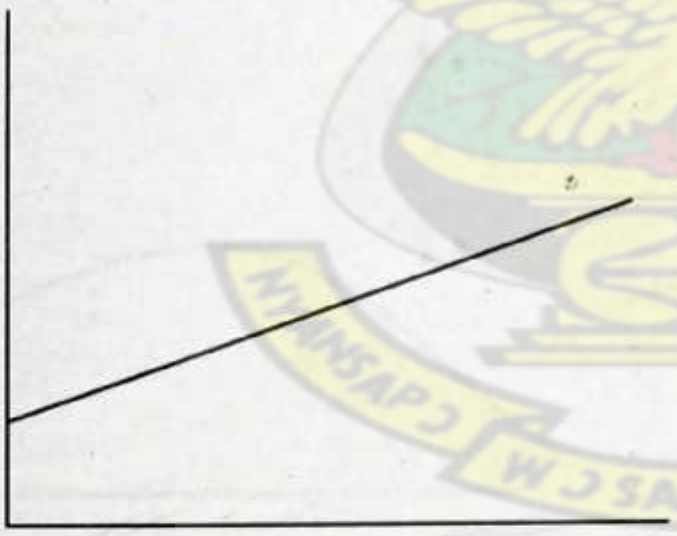


Figure 3.2 An increasing trend of a time series.

3.3.2 Seasonality

Many times series such as sales figures and temperature readings exhibit variations which is annual in period. For example, sales of umbrellas are typically high during the rainy seasons and low during the dry season. This yearly variations is easy to understand and can be

measured explicitly. When this is removed from the data, the data becomes deseasonalized. An example of a deseasonalized data is shown below in figure 3.3

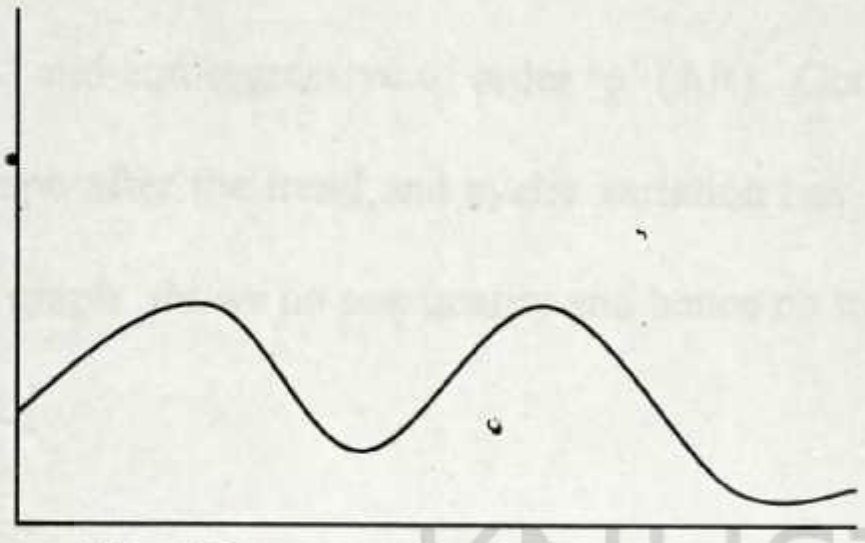


Figure 3.3 Deseasonalized Time Series Data

3.3.3 CYCLIC EFFECTS

The cyclical effects, symbolled C, generally describes fluctuations of the time series about the secular trends that are attributable and economic condition of that time. Example of cyclical factors is the general business cycle which is shown below in figure 3.4

Figure 3.4 Cyclic Effects of Times Series

3.3.4 IRREGULAR FLUCTUATIONS

After trend and cyclic variation have been removed from a set of data, we are left with a series of residents, which may or may not be "random". There are various techniques for

analyzing series of this type, to find out if some of the apparently irregular variations may be explained in terms of probability models. The concerned models which are associated with irregular fluctuations after the removal of the trend and cyclic variations are moving average of order "q" (MA) and autoregressive of order 'p' (AR). Consider for example the nature of the time series graph after the trend and cyclic variation has been removed as shown below in figure 3.5. The graph shows no seasonality and hence no trend.

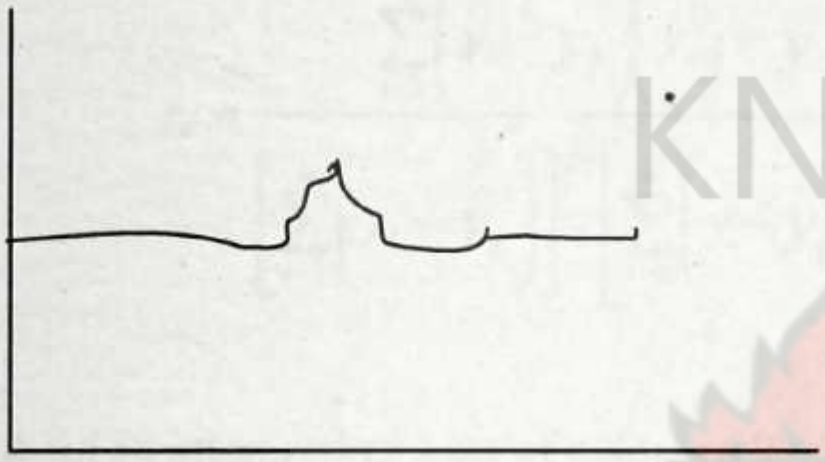


Figure 3.5 A detrended and Deseasonalized graph

3.4 THE AUTOCORRELATION FUNCTION

It is generally impossible to obtain a complete description of a time series process, the autocorrelation function is extremely useful in helping us to obtain a partial descriptions if the process for developing a forecasting model. Moreover, the autocorrelation coefficient measures the degree of correlation between neighboring data observations in a time series. We begin by defining the autocorrelations with log k as

$$P_k = \frac{E(Y_t - \mu_y)(Y_{t-l} - \mu_y)}{\left[E(Y_t - \mu_y)^2 E(Y_{t+k} - \mu_y)^2\right]^{1/2}} \qquad (3.4.1)$$

$$P_k = \frac{Cov(Y_t - Y_{t+k})}{\sigma Y_t \sigma Y_{t+k}} \qquad (3.4.2)$$

If the numerators are equal, $P_o = 1$ this will occur when k = 0 From the above relation, if K = 0 it implies $P_o = 1$ at that time the numerator and the denominator will be equal

In short autocorrelation function is usually found by normalizing the autoconvariances; thus dividing each y_k by the variance $y_o = \sigma_y^2$ of the process.

Thus, simple put $P_k = y_k$, K = 0, +1, +2, …

The autocorrelation coefficient is estimated from sample observation as follows;

$$r_k = \frac{\sum_{t-2}^{n}\left(y_t - \hat{y}_t\right)\left(y_{t-1} - \hat{y}_{t-1}\right)}{\left[\sum_{t=2}^{n}\left(y_1 - \hat{y}_t\right)\right]^{1/2}\left[\sum_{t-2}^{n}\left(y - y_{t-1}\right)^2\right]^{1/2}} \qquad (3.4.3)$$

A useful aid in interpreting a set of autocorrelation coefficient is a graph, called correlogram, which is plotted against the lag K. Below is and example to illustrate. A time series diagram showing short-term correlation together with its correlation

3.4.1 SAMPLING DISTRIBUTION OF AUTOCORRELATION COEFFICIENT

Consider a time series constituting a set of completely random and independent observations, and then the autocorrelation for all orders must be few. However, if we sample say 40 observation out of random, the estimated autocorrelation coefficient for any order could be different from zero. Out of the whole population, we can select an infinite number or samples; say 40 each time selected at random

It could be observed that each random sample of 40 observations is likely to have different sample autocorrelation coefficient of order 1, … 16, this resulting value would be near zero. Thus if K is denoted as the autocorrelation for the population, the autocorrelation for different even sample of observation would form a distribution of values around K.

The distribution would be called the "sampling distribution of values around K.

This distribution would be called the "sampling distribution of autocorrelations". Statistical theories teach us that the sampling distribution of autocorrelation coefficient of random data is normal with

$$\mu p_k = 0 \; \sigma p_k = 1/\sqrt{n} \qquad (3.4.1.1)$$

3.5 PARTIAL AUTOCORRELATION FUNCTION (PACF)

Partial auto correlation function is another aid to determining the order of an Autoregressive (AR) model when fitting an AP (p) model; the last coefficient αp will be denoted and measure the excess correlation at lag p which is not accounted for by an AP (p-1) model. It is called the pth partial auto correlated coefficient and when π_p is plotted against p gives the partial autocorrelation function. Partial autocorrelation coefficients measure the degree of association between the current observations of X_p and earlier observations X_{t-x} when the effects of other time lags on X are held constant.

The purpose is to help identify the best autoregressive moving average (ARMA) model for forecasting. PACF is calculated when we are aware of the appropriate order of the autoregressive process to fit the time series It is defined in terms of the last autoregressive term of an AR model of m lags. The symbol ϕ is used to denote partial autocorrelation and ϕ is the estimate of partial autocorrelation. Thus $\phi_1, \phi_2, \ldots, \phi_{m-1}, \phi_m$ are the m partial auto correlation of the AR (m) model as defined in the following equations.

$$X_t = \phi_1 X_{t-1} + e_1$$

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + e_1$$

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots \phi_{m-1} X_{t-m--1} + \phi_{m-1} X_{t-m} + e_1$$

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots \phi_{m-1} X_{t-m--1} + \phi_{m-1} X_{t-m} + e_1$$

Solving the above equations for ϕ_i, i = 1, 2, …, m is very difficult and time consuming and so we use their estimate ϕ_i, I = 1, ….m, based on their autocorrelation coefficients.

$\Phi_i = P_1$ = for i = 1, ……..,m. This is the partial autocorrelation coefficient of one time lag is ϕ_1 and r_1 is its estimator.

3.5.1 SAMPLING DISTRIBUTION OF PARTIAL AUTOCORRELATION COEFFICIENTS.

An autoregression of order p, AR (p) process has only partial autocorrelations that are statistically different from zero. Thus AR (1) model ϕ_{m-1} will not be statistically different form zero. Values which fall outside of 2 standard deviation ($\underline{1}\ 2/\sqrt{N}$) will indicate the order of the process since the partial autocorrelation function of p^{th} order of AR process is zero, for all lags greater than P. Finally, partial autocorrelations are conducted to fit only AR models and also introduce dependence from one lag to another that causes them to behave in the manner similar to that of autocorrelation for AR process. On the other hand, partial correlation decline to zero exponentially. Therefore if the partial autocorrelation do not

decline to random values after p time lags but instead decline to zero exponentially, the time generating process is concluded to be an MA (Moving Average) process.

3.6 STATIONARITY

In analyzing a time series, it is commonly assumed that the data are stationary. A stationary process has the property that the mean and autocorrelation structure do not change over time. Stationarity can be defined as a flat looking series, without trend, constant variance over time, a constant autocorrelations structure over time and periodic fluctuations (seasonality). In practice stationarity can be usually determined from one sequence plot.

3.6.1 ACHIEVING STATIONARITY

If a time series is not stationary, we can often transform it to stationarity by differencing the data. That is given the series Z_t, we create the new series.

$$Y_t = Z_t - Z_{t-1}$$

The differenced data will contain one less point than the original data. The data can be differenced more than once but this depends on the nature of the values and the non stationarity of the mean and variance. Differencing once is usually sufficient for variances which are non-constant, taking the logarithm or square root of the series may also stabilize the variance. If the data contains trends, we can fit some types of curve to the data and then model the residuals form the fit. Since the purpose of the fit is to simply remove long term trend, a simple fit, such as a straight line is typically used.

3.7 BACKSHIFT OPERATOR OR LAG OPERATOR

To express and understand the differenced ARIMA model, the concept of the Backshift operator B must be introduced. This operator has no mathematical meaning other than to facilitate the writing of different types of models that would otherwise be extremely difficult to express. The backshift operator is defined as $B^m Y_t$. For example

$BY_t = Y_{t-1}$ or $Be_1 = e_{t-1}$

$B^0 Y_1 = Y_t$ or $Be_t = e_t$

$B^2 Y_t = Y_{t-2}$ or $B^2 Y_1 = Y_t$ or $Be^2 e_1 = e_{t-2}$

$B^3 Y_t = Y_{t-3}$ or $B^3 e_t - e_{t-3}$

$B^m Y_t = Y_{t-m}$ or $B^m e_t = e_{t-m}$

Any ARIMA model can be expressed in terms of the backshift operator. An AR(1) model, ARIMA (1,0, 0) is expressed as

$Y_t = \phi_1 Y_{t-2} + e_1$ .....(1)

$Y_t - \phi Y_{t-1} + e_t$

However, since $BY_1 = Y_{t-1}$ equation (1) can be rewritten as

$Y_t - \phi_1 BY_t = e_t$

$Y_t(1 - \phi_1) = e_t$

For an AR(2) model $Y_t - \phi Y_{t-1} - \phi_2 Y_{t-2} = e_1$

ARIMA(2, 0, 0) we can write $BY_t = = Y_{t--1}$

And $B^2Y_t = Y_{t-2}$, which can be written as

$Y_t - \phi_1 BY_t - \phi_2 B^2 Yt = e_t$

And $(1 - \phi_1 B - \phi_2 B^2)\, Y_t = e_t$

If we have an AR (2) model where first differences have been used, suggest of an ARIMA, (2,1,0) written as :

$(Y_t - Y_{t-1}) = \phi_1(Y_{t-1} - Y_{t-2}) + \phi_2(Y_{t-2} - Y_{t-3}) + e_t$

3.8 TIME SERIES MODEL

Finding a suitable model for a given time series depends on a number of factors which include properties of the series which are assumed by visual examination of the data and the number of observation.

3.8.1 AUTORGRESSIVE MODEL OF ORDER P; AR(P)

An autoregressive is a regression where the right side variables are merely the values of the dependant variables in previous period. An autoregressive time series of order one(1) is denoted by AR(1) defined by;

$X_t = \mu + \phi_1 X_{t-1} + e_t$

Similarly, an autoregressive time series of order two is denoted by AR(2) and defined by $X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + e_t$

Hence in a more general sense, the series may have the form

$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + e_t$

Where $\mathbf{X_t}$: is the observed value of the series at time t, $\mathbf{X_{t-1}}$: the observed value of the series at time t -1 and $\mathbf{X_{t-p}}$:the observed value at time t-p. The constant μ is a fixed

parameter to be estimated from the least square regression analysis. The coefficient ϕ_1, ϕ_2 …, ϕ_p, are also to be estimated manually or by computerized calculation . e_t is a random term which is independent from period to period. In general, the necessary and sufficient condition for an AR (P) process to be stationary is that $-1 < \phi < 1$, thus the zeros of ϕ should all lie outside the unit circle.

3.8.2 MOVING AVERAGE MODEL OF ORDER Q; MA(Q)

Moving average models provide predictions of X_t based on linear combinations of past forecast errors. A moving average time series model of order q denoted by MA (q) is defined by

$$X_t = \mu + e_t \theta_1 e_{t-1} - \theta_2 e_{t-2} \ldots - \theta_q e_{t-q}$$

When q = 1 we have moving average of order 1 MA (1) and thus $X_t = \mu + e_t - . \theta_1 e_{t-1}$

A special feature of MA (1) process is that the autocorrelation function cuts off at a lag q

In general the MA (1) process with q = 1 is given by $X_t = . \theta_1 e_{t-}t + e_t$

Furthermore, theoretical autocorrelation function cuts off after a lag 1. It is general desirable to impose restrictions on (θ_1) to ensure that the model satisfies a condition called invertibility.

Invertibility refers to the probability of inverting an MA model and expressing it as an AR model of finite order. An MA model is invertible, example, if the deviation $X_1 - \mu$ does not depend overwhelmingly on deviations in the distant past. For an MA (1) model, the invertibility condition is $-1 < \theta < 1$.

3.8.3 ARMA MODELS

This is a general model which is a mixture of AR (p) and MA (q) called the autocorrelation moving average model (ARMA). This is used when autocorrelation patterns may require more complex models. The integrated part does not exist since there is no differencing. ARMA (p, q) is given by

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots, + \phi_p X_{t-p} + e_1 - \theta_1 e_{t-1} - \theta_2 e_{t-2} \ldots - \theta_q e_{t-q}$$

For ARMA (1, 1) model we have $X_t = \phi_1 X_{t-1} - \theta_1 e_{t-1}$ It can be seen that ARMA (1, 1) model is stationary and invertible.

3.8.4 AUTOREGRESSIVE INTEGRATED MOVING AVERAGE ARIMA(P, d q);

A process is said to be an ARIMA process of order (p,d q,) if the dth difference of X_t satisfies the ARMA operator of order (p,q). An ARIMA process with a positive integer '**d**' is generally not stationary, whereas a fractional ARIMA process can be stationary. Fractional ARIMA processes are also known as long memory processes while ARIMA, AR and MA are short memory processes.

3.9 THE YULE-WALKER EQUATION AND ITS USE IN ESTIMATING THE PARAMETERS OF THE AR (P) MODEL

The most important fact about AR models is that, it is possible to obtain a simple set of linear equations that expresses the parameters of the model in terms of the autocorrelation and variance. These equations are called the Yule-Walker Equations

If we write general order AR (p) process

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + e_1 \qquad (3.9.1)$$

We can multiply by X_{t-k}, take expectation and obtain

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \ldots + \phi_p \rho_{k-p} \qquad (3.9.2)$$

In the operator notation, we can write AR (p) process as **AR (p)X, = e_t** and these correlation functions satisfies the equation **AR(B)** $\rho_k = 0$ where k > 0. This is called the autocorrelation generating function. It also gives the familiar Yule-Walker equation.

$$\rho_k = \phi_1 \rho_{k-1} + \phi_1 \rho_{k-2} + \ldots + \phi_p \rho_{k-p} \qquad (3.9.3)$$

Substituting k = 1, 2, 3, …, p in the above equation, it gives

$$P_1 = \phi_1 + \phi 2\, p_1 + \ldots + \Phi_p\ p_{p-1} \qquad (3.9.4)$$

$$P_2 = \phi p_1 + \phi 2 + \ldots + \Phi_p\ p_{p-2} \qquad (3.9.5)$$

$$P_p = \phi_1 p_1 + \phi_2\, p_1 + \phi_2\, p_{p-2} + \ldots + \Phi_p \qquad (3.9.6)$$

These are called the Yule-Walker Equations.

Writing these in terms of matrix form, where p is the matrix equation and ϕ is the column vector, The Yule – Walker equation for AR(2) are

$$P_1 = \phi_1 + \phi_2 p_1 \qquad (3.9.7)$$

$$P_2 = \phi_1 + p_1 + \phi_2 \qquad (3.9.8)$$

Putting these in matrix form we have

$$\begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} 1 & p_1 \\ p_1 & 1 \end{bmatrix} \begin{bmatrix} \phi \\ \phi \end{bmatrix} \qquad (3.9.9)$$

Applying Crammer's rule we have

$$\begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} 1 & p_1 \\ p_1 & 1 \end{bmatrix} \begin{bmatrix} \phi \\ \phi \end{bmatrix} \qquad (3.9.10)$$

Where $1 - p_1^2$ is the determinant, Hence

$$\begin{bmatrix}\phi\\ \phi\end{bmatrix}=\begin{bmatrix}1 & -p_1\\ -p_1 & 1\end{bmatrix}\begin{bmatrix}p_1\\ p_2\end{bmatrix} \tag{3.9.11}$$

$$\varphi=\frac{p_1-p_1^{\,2}}{1-p_1^{\,2}} \text{ and } \varphi=\frac{p_2-p^2}{1-p^2_1} \tag{3.9.12}$$

By replacing the p_k by $r_k$we have Yule-Walker for R(2).

3.10 TREND REMOVAL AND STATIONARY TIME SERIES

A time series is said to be *stationary* if both its mean (the value about which it is oscillating), and its variance (amplitude) remain constant through time. Classical Box-Jenkins ARMA models only work satisfactorily with stationary time series, so for those types of models it is essential to perform transformations on the series to make it stationary. The models developed by DTREG are less sensitive to non-stationary time series than ARMA models, but they usually benefit by making the series stationary before building the model. DTREG includes facilities for removing trends from time series and adjusting the amplitude. Consider for example figure 3.4 which deals with the series trend for passengers. The figure below has both increasing mean and variance:

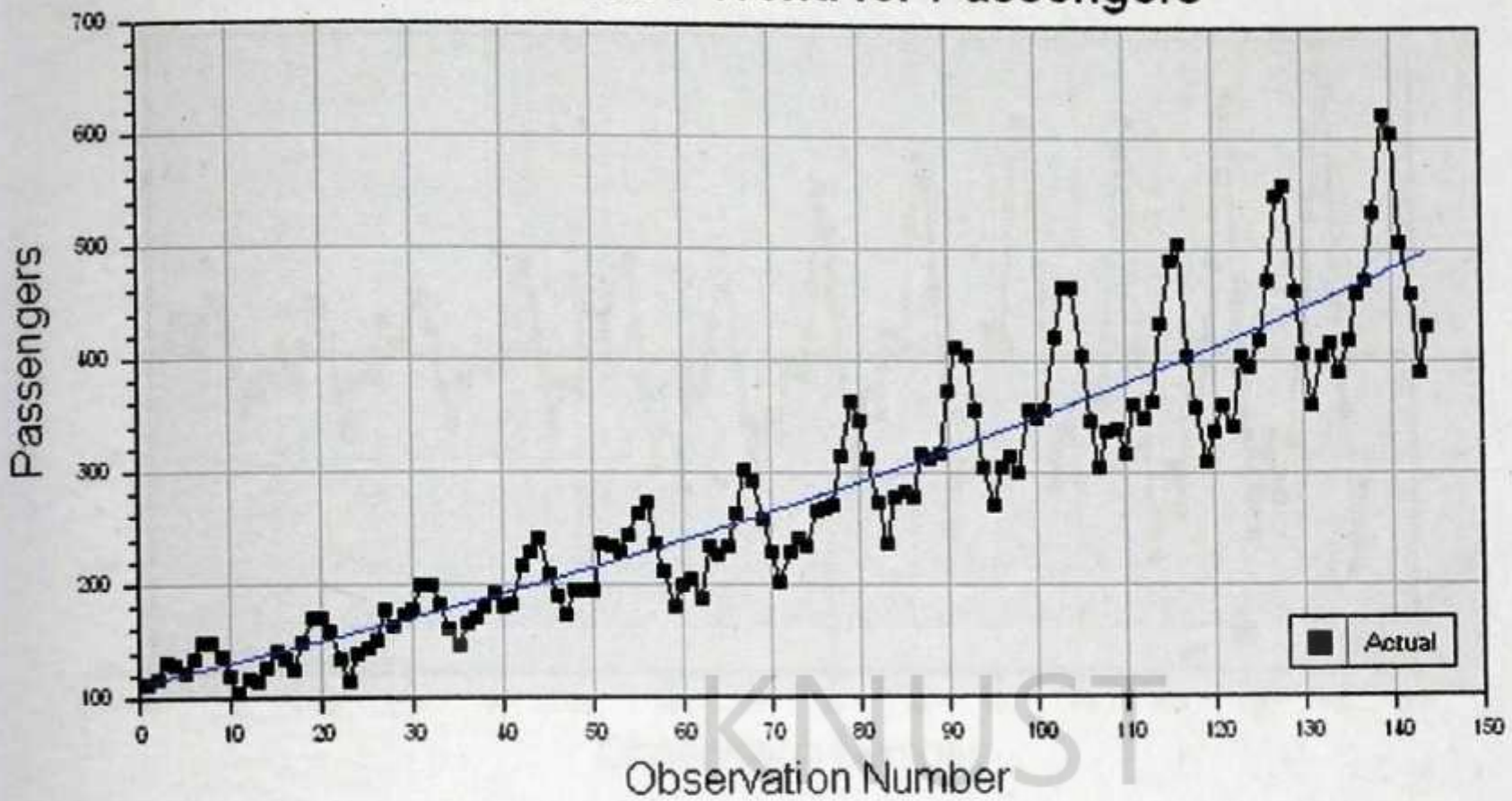


Figure 3.6 Time series graph showing Trend for Passengers

If the trend removal option is enabled on the Time Series property page see figure 3.6, then DTREG uses regression to fit either a linear or exponential function to the data. In this example, an exponential function worked best, and it is shown as the blue line running through the middle of the data points. Once the function has been fitted, DTREG subtracts it from the data values producing a new set of values as shown in figure 3.7 . The figure below is a transformed data of the time series trend for passengers. See the figure 3.7 below.

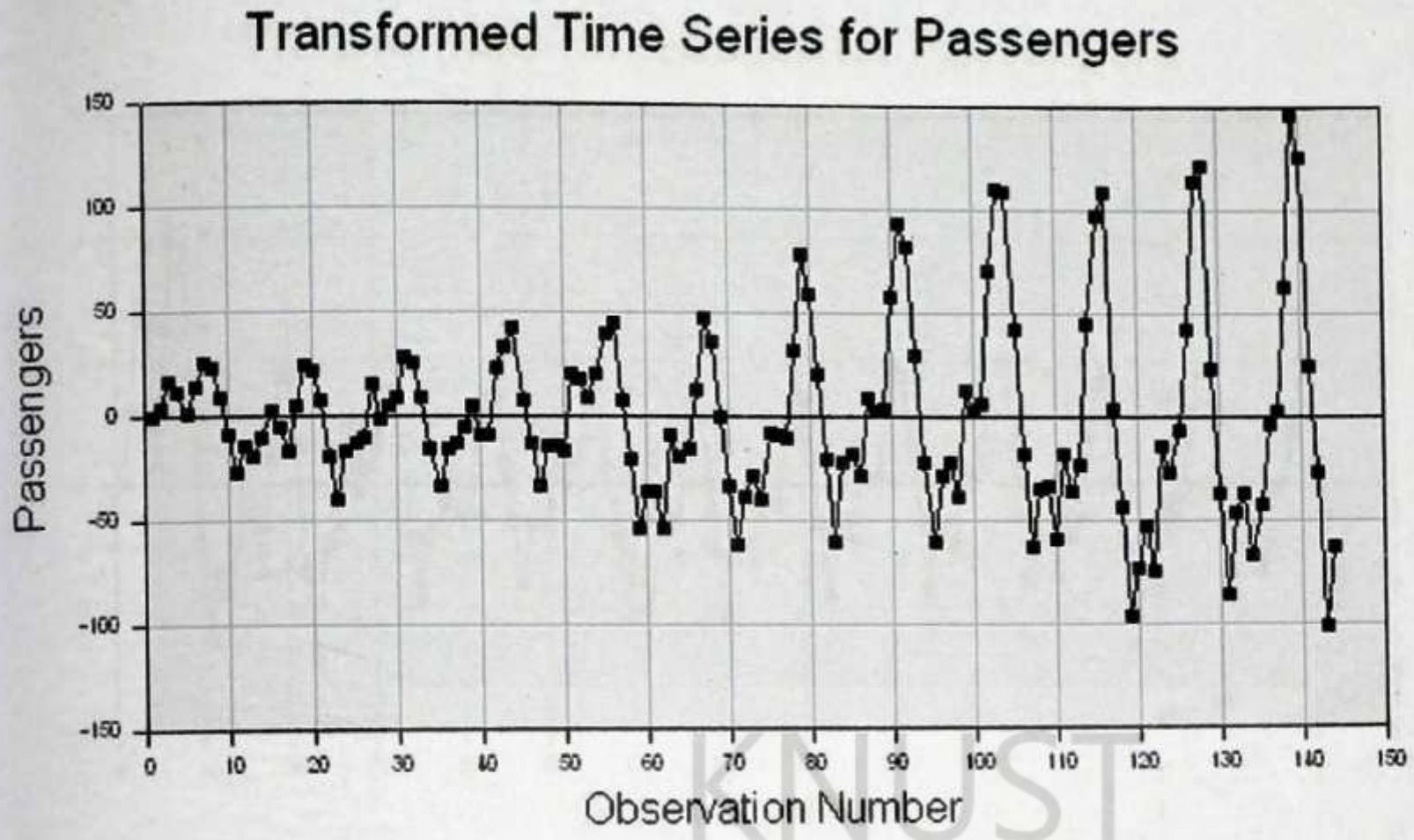


Figure 3.7 Transformed Time Series for Passengers

The trend has been removed, but the variance (amplitude) is still increasing with time. If the option is enabled to stabilize variance, then the variance is adjusted to produce this series: An increasing variance from Figure 3.7 suggest that the mean and variance are stationary. Once the data is stationary implies that a PACF and ACF test ought to be applied to determine the type of model and thereby fixing it .

Judge (1985) points out that when the PACF has a cutoff at p whiles the ACF tails of suggest an autoregressive of order p (AR, p). If on the other hand the ACF has a cutoff at q whilst the PACF tapers off gives a moving average of order q. On the contrary when both ACF and PACF tails off , indicate an Autoregressive moving average of order p and q.

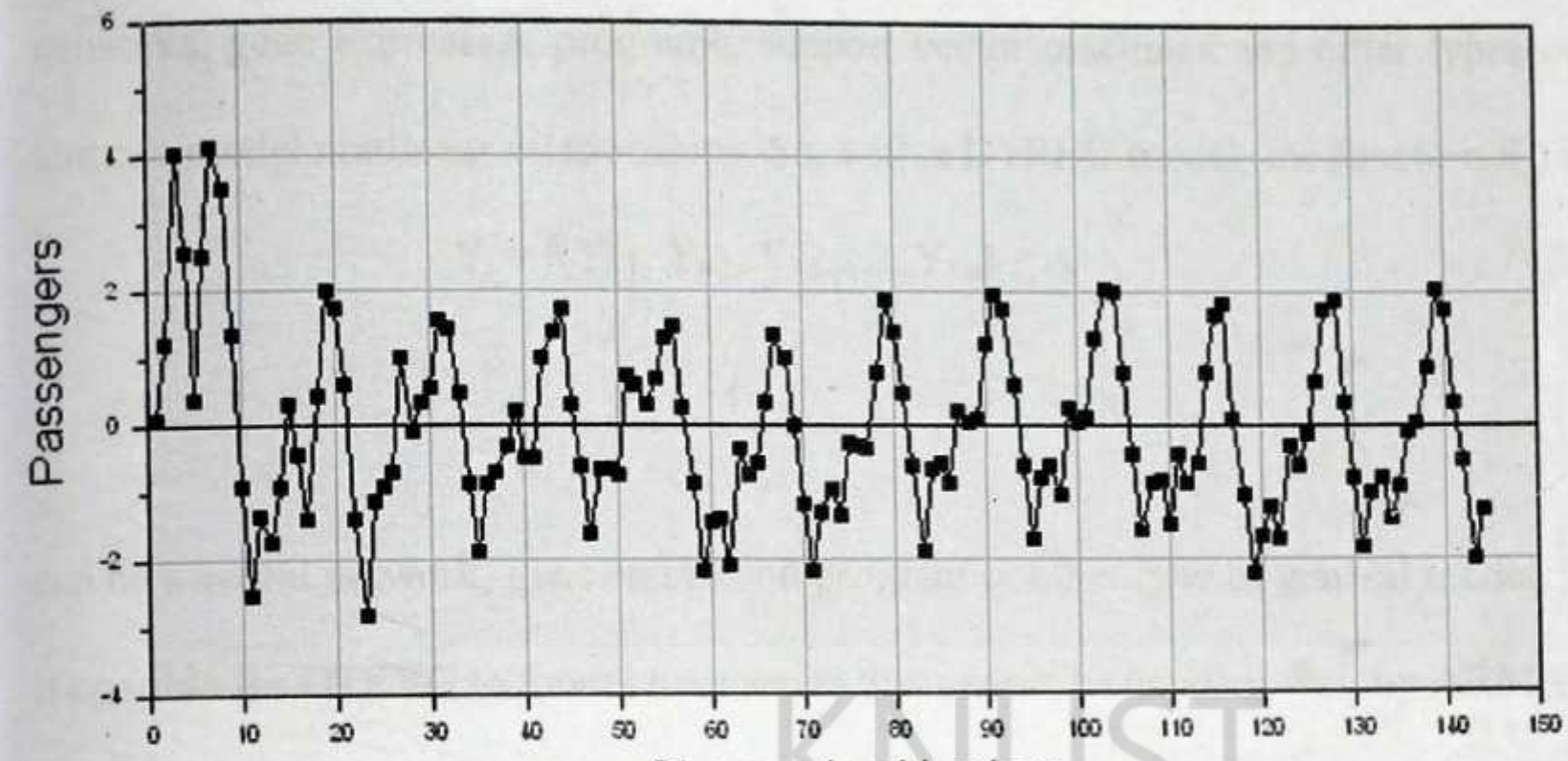


Figure 3.8 Transformed Time Series for Passengers

This transformed series is much closer to being stable. The transformed values are then used to build the model. A reverse transformation is applied by DTREG when making forecasts using the model. The ACF and PACF are used to identify a model by taking into consideration the nature of the graph . pointing where it is decaying or cutting.

3.11 BOX–JENKINS METHOD OF MODELLING TIME SERIES DATA

The box-Jenkins methodology is a statistical way of analysing and building a forecasting model which best represents a time series. It has a number of advantages over other methods of time series analysis. Traditional time series analysis uses Box-Jenkins ARMA (Auto-Regressive Moving Average) models. An ARMA model predicts the value of the target variable as a linear function of lag values (this is the auto-regressive part) plus an effect from recent random shock values (this is the moving average part). While ARMA models are widely used, they are limited by the linear basis function.

In contrast to ARMA models, DTREG can create models for time series using neural networks, gene expression programs, support vector machines and other types of functions that can model nonlinear relationships. So, with a DTREG model, the function $f(.)$ in

$$Y_t = f(Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots, Y_{t-n}) + e_t \qquad (3.11.1)$$

can be a neural network, gene expression program or other type of general model. This makes it possible for DTREG to model time series that cannot be handled well by ARMA models

3.11.1 ADVANTAGES OF THE BOX – JENKINS METHOD.

The Box-Jenkins method has number of advantage over the other methods of time series analysis.

- It is logically and statistically accurate
- The method extracts a great deal of information from the historical time series.
- Finally, the method results in an increase in forecast accuracy while keeping the number of parameters to a minimum in comparison with similar modelling process

3.11.2 STAGES IN BOX-JENKINS MODELLING

The Box- method consists mainly of four stages.

- Model Identification
- Model Estimation
- Model Validation
- Forecasting

The Box Jenkins method hopes, from inspection of the estimated ACF and PACF obtained from an observed time series, to recognize patterns which could be explained by some such model. After this identification, it proceeds to efficiently estimate the values of the various parameters, the estimation stage, and then perform verification test to determine whether the 'fitted' model is adequate;

If it is not, these very tests should indicate how the model ought to be modified, and further cycle of identification, estimation and verification is instigated. To effectively fit the Box-Jenkins model, it requires at least a moderately long series. Chatfield (1996) recommends at least 50 observations. Many others recommend at least 100 observations. Found below in figure 3.7 is a chart underlining the stages in the box –Jenkins approach such as data collection for forecasting, identifying the model, estimating the parameters of the tentative model, diagnostic checking and finally using the model to forecast.

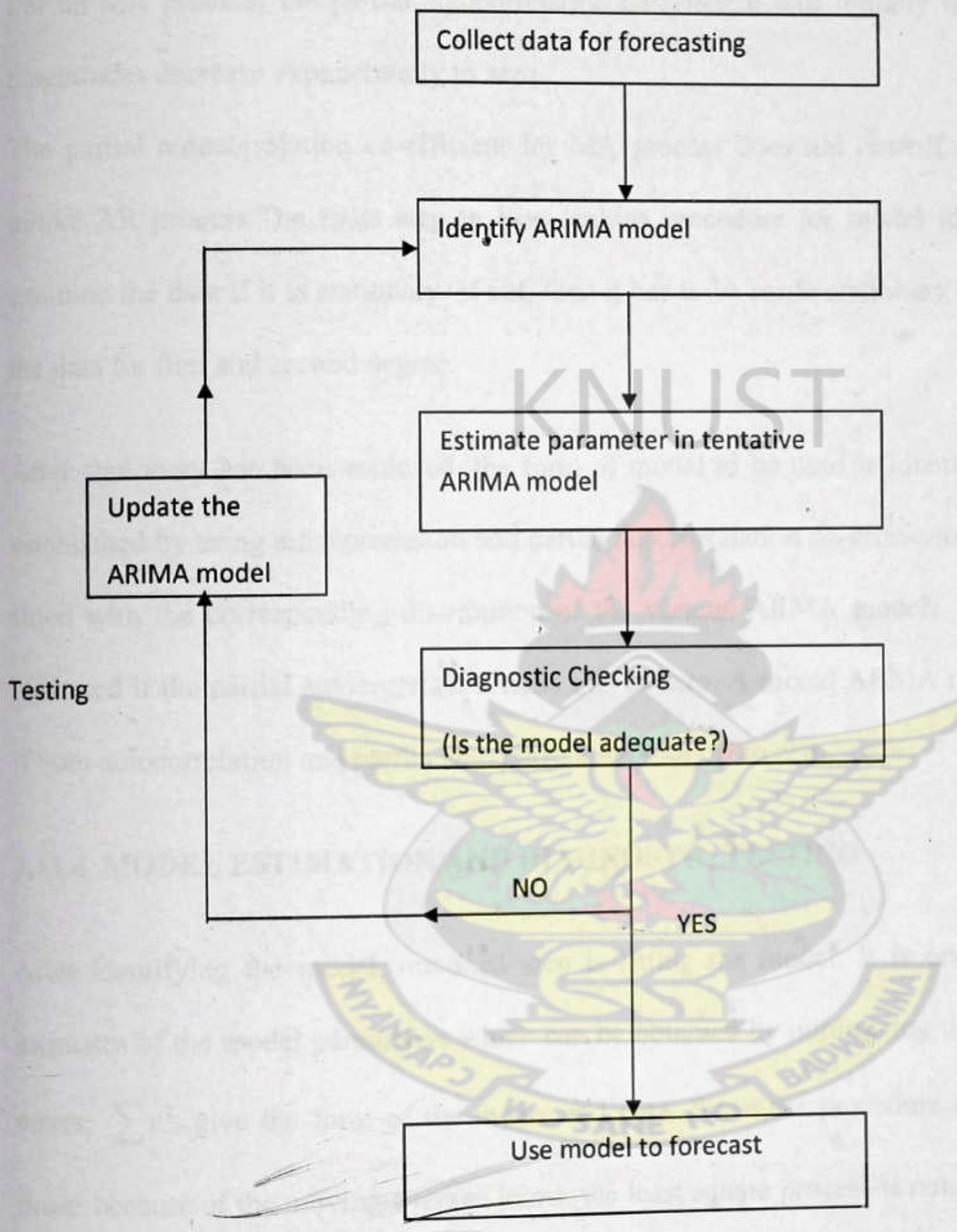


Figure 3.9 Chart of the box-Jenkins process

In this phase, we are to choose an appropriate model from the general class of ARMA (p,q) model denoted as

$$X_1 = \phi_1 X_{t-1} + \phi_2 X_{t-2} + .. \phi_p X_{t-p} + e_1 - \theta_2 e_{t-2} - ... - \theta_q e_{t-q} \qquad (3.11.3.1)$$

There is the need to examine the autocorrelation and partial autocorrelation co-efficient calculated for the given data before an appropriate values of p and q can be selected.

For an MA process, the partial autocorrelation co-efficient will initially be large and their magnitudes decrease exponentially to zero.

The partial autocorrelation co-efficient for MA process does not cut off after p time lags unlike AR process.The firsts step in Box-Jenkins procedure for model identification is to examine the data if it is stationary. If not, then it has to be made stationary by differentiating the data for first and second degree.

After stationary has been achieved, the form of model to be used is identified. The form is established by using autocorrelation and partial autocorrelation co-efficients of the data to be fitted with the corresponding distribution of the various ARMA models. An AR model is indicated if the partial autocorrelation trails off to zero. A mixed ARMA model is indicated if both autocorrelation and partial auto correlation trail off to zero.

3.11.4 MODEL ESTIMATION AND DIAGNOSTIC TESTING

After identifying the model, our next step is fitting the model. It is one by least square estimates of the model parameters which can be obtained by minimizing the residual sum of errors; $\sum e^2_t$ give the form of the model and data. Since the procedure is in general non-linear because of the moving-average terms, the least square process is non-linear. Before the model is used for forecasting, it should be checked for adequacy. This diagnostic checking is done by examining the error e_t. to be sure that they are random.

If the error terms are statistically different from zero, the model is not considered adequate. If several autocorrelations are large, we should return to the initial stage select an alternative model and then continue the analysis. To check for adequacy, the autocorrelations of the

residuals are diagnostically examined by calculating a Chi square statistic (X^2). The test statistic is the Q-STATISTIC.

$$Q = n(n+2)\sum_{i=1}^{k}(r_i^2 / n-k) \qquad (3.11.4.1)$$

This is approximately distributed as X^2. With $k-p-q$ degrees of freedom.

Here,

- n is the length of the time series.
- k is the first k autocorrelations being checked
- p is the order of the AR series
- q is the order of the MA process.
- R is the estimated autocorrelation co-efficient of the i^{th} residual term.

If $Q>X^2$ for $k-p-q$ degrees of freedom, then the model is considered inadequate. In this case, we return to selecting an alternative model and continue the Box-Jenkins analysis until a satisfactory model is found.

3.11.5 FORECASTING

The forecasting can be made when a satisfactory model is found. Forecasts are conditional statements about the future based on specific assumptions. Forecasting methods may be broadly classified into three groups as follows;

3.11.6 SUBJECTIVE

In this case, forecasting is made on a subjective basis using judgment, intuition, commercial knowledge and any other relevant information.

3.11.7 UNIVARIATE

Forecasts of a given variable are based on a model fitted only to past observation of the given time series. Univariate forecasts are particulafly suitable when there is large number of series to be forecast.

3.11.8 MULTIVARIATE

Here forecast of a given variable depends at least partly on values of one or more other series called prediction or explanatory variable.

3.12 CRITERIA FOR SELECTING COMPOSITION MODELS

The following criteria exit and are crucial in selecting models. They for example guide in selecting the required model among the lot. They include Q-statistics, Residual Variance, Akaike Information Criteria (AIC) and Parsimony.

3.12.1 Q-STATISTIC

This is used in conjunction with X^2. Statistic to check for the adequacy of the model to be used. It is given by;

$$Q = n(n+2)\sum_{i}^{k}(r_i^2/n-k) \qquad (3.12.1.1)$$

Where n is the length of the time series, k is the first k autocorrelation being checked, r is the estimated autocorrelation co-efficient of the ith residential term. This statistic is approximately distributed as X^2 with k-p-q degrees of freedom with p as the order of the AR, q the order of the MA and k the first k autocorrelation co-efficient being checked. If $Q > X^2$, then the model is not adequate. The whole process is repeated until a satisfactory model is found.

3.12.2 RESIDUAL VARIANCE

When calculating the variance σ^2 we square the deviation from the mean sum and divide by N. The square root of the variance provides our standard deviation.

If we are now to square and sum the squared deviation of scores from the regression line $\sum(Y-Y^1)^2$. We would have a basis for calculating another variance and standard deviation.

The variance around the regression line is known as residual and is defined as

$$\sigma^2 VarY = \frac{\sum(Y-Y^1)^2}{N-2} \qquad (3.12.2.1)$$

3.12.3 AKAIKE INFORMATION CRITERIA (AIC)

This is given by; AIC (p,q) = -2 In (maximum likelihood) +2r (3.12.3.1)

$$\mathbf{In}(\sigma^2 e) + r2n + cons\tan t \qquad (3.12.3.2)$$

Where (σ^2 e) denotes the maximum likelihood estimate and r = p + q+ 1 denotes the number of parameters estimated in the model including a constant term.

The first term corresponds to minus of 2/n times the natural log of the maximum likelihood, while the second is a 'penalty factor' for inclusion of additional parameters into the model.

3.12.4 PARSIMONY

If two different models adequately describe a time series equally well, the model with few parameters is used because there are fewer parameters to be estimated and the estimation process is simpler to interpret. Models with fewer parameters are called parsimonious models.

3.12.5 SUMMARY

In this chapter, we reviewed the general literature of the models of the study, explained some of the terms and the methodology used to achieved the results of the study.

In the next chapter, we shall presents data collection and analysis of the study.

CHAPTER 4

4.0 INTRODUCTION

This chapter presents data collection analysis of the study. The information gathered concerned the malaria cases at the (OPD) of (KATH). The data is monthly data for the years 2000 to 2010.

4.1 OPD Malaria Cases Data

Data collected for the analysis consists of a eleven years data from 2000 to 2010 of the various months, making a total of one hundred and thirty two data observations.

Table 4.1 shows the month and corresponding malaria cases.

MONTH	MALARIA CASES	MONTH	MALARIA CASES	MONTH	MALARIA CASES	MONTH	MALARIA CASES
1	2247	13	2574	25	2649	37	2800
2	2549	14	2130	26	2559	38	1769
3	3151	15	2899	27	2311	39	1567
4	1945	16	2250	28	2431	40	2001
5	1542	17	2512	29	1461	41	1533
6	1563	18	3112	30	2000	42	1459
7	1618	19	3211	31	1949	43	3245
8	1931	20	2245	32	2146	44	2031
9	1129	21	1234	33	1459	45	1432
10	1673	22	1324	34	2314	46	2321
11	3521	23	2321	35	2432	47	3241
12	1949	24	3121	36	2451	48	3214

Con't: Table 4.1 shows the month and corresponding malaria cases.

MONTH	MALARA CASES	MONTH	MALARIA CASES	MONTH	MALARIA CASES	MONTH	MALARIA CASES
49	2770	71	2696	92	2187	113	1637
50	2427	72	2843	93	2178	114	1673
51	3058	73	2489	94	2731	115	2211
52	1836	74	1867	95	2365	116	2948
53	1900	75	1733	96	2531	117	1030
54	2347	76	737	97	2643	118	1400
55	1418	77	1470	98	1678	119	2254
56	2281	78	1060	99	1919	120	1878
57	1496	79	1890	100	2808	121	2130
58	1489	80	2362	101	2771	122	2006
59	4071	81	1812	102	3461	123	2233
60	3411	82	1938	103	2913	124	2363
61	3403	83	1529	104	2554	125	1541
62	2650	84	953	105	2577	126	3000
63	2872	85	1498	106	2620	127	2884
64	2284	86	1638	107	1807	128	2029
65	2028	87	1533	108	1988	129	1032
66	2347	88	2014	109	2119	130	1979
67	3274	89	2943	110	1542	131	1409
68	3136	90	2859	111	1528	132	1527
69	1736	91	2520	112	1353		
70	3126						

Transformation of Data into a time series data

Since data were raw taken from the OPD, it was transform into a Time Series Data by the use of R command, starting from 2000 with a frequency of 12, given the result shown in Table 4.2

Table 4.2 Data into a time series data

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2000	2247	2549	3151	1945	1542	1563	1618	1931	1129	1673	3521	1949
2001	2574	2130	2899	2250	2512	3112	3211	2245	1234	1324	2321	3121
2002	2649	2559	2311	2431	1461	2000	1949	2146	1459	2314	2432	2451
2003	2800	1769	1567	2001	1533	1459	3245	2031	1432	2321	3241	3214
2004	2770	2427	3058	1836	1900	2347	1418	2281	1496	1489	4071	3411
2005	3403	2650	2872	2284	2028	3274	3274	3136	1736	3126	2696	2843
2006	2489	1867	1733	737	1470	1060	1890	2362	1812	1938	1529	953
2007	1498	1638	1533	2014	2943	2859	2520	2187	2178	2731	2365	2531
2008	2643	1678	1919	2808	2771	3461	2913	2554	2577	2620	1807	1988
2009	2119	1542	1528	1353	1637	1673	2211	2048	1030	1400	2254	1878
2010	2130	2006	2233	2363	1541	3000	2884	2029	1032	1979	1409	1527

4.2 Trend Analysis

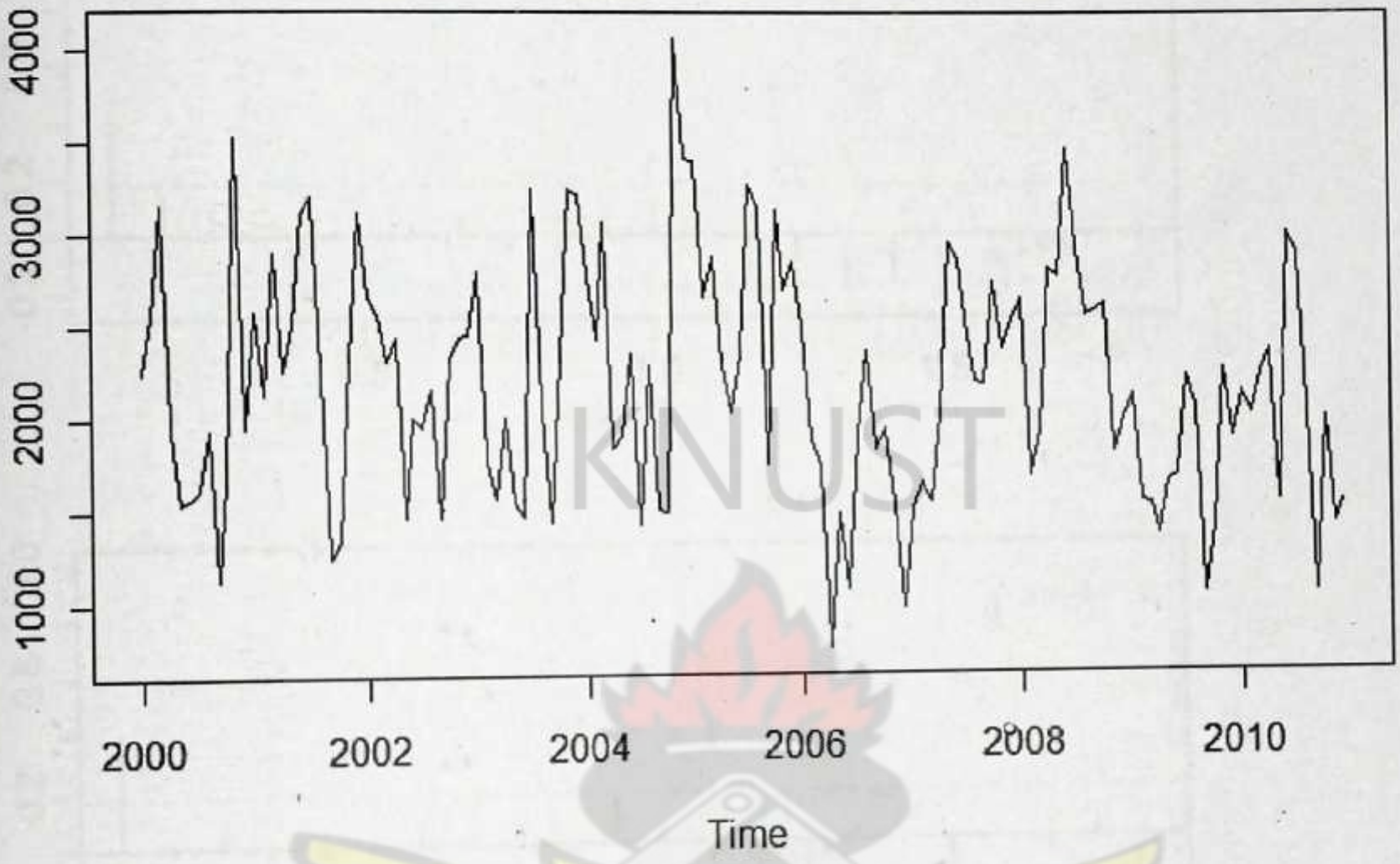
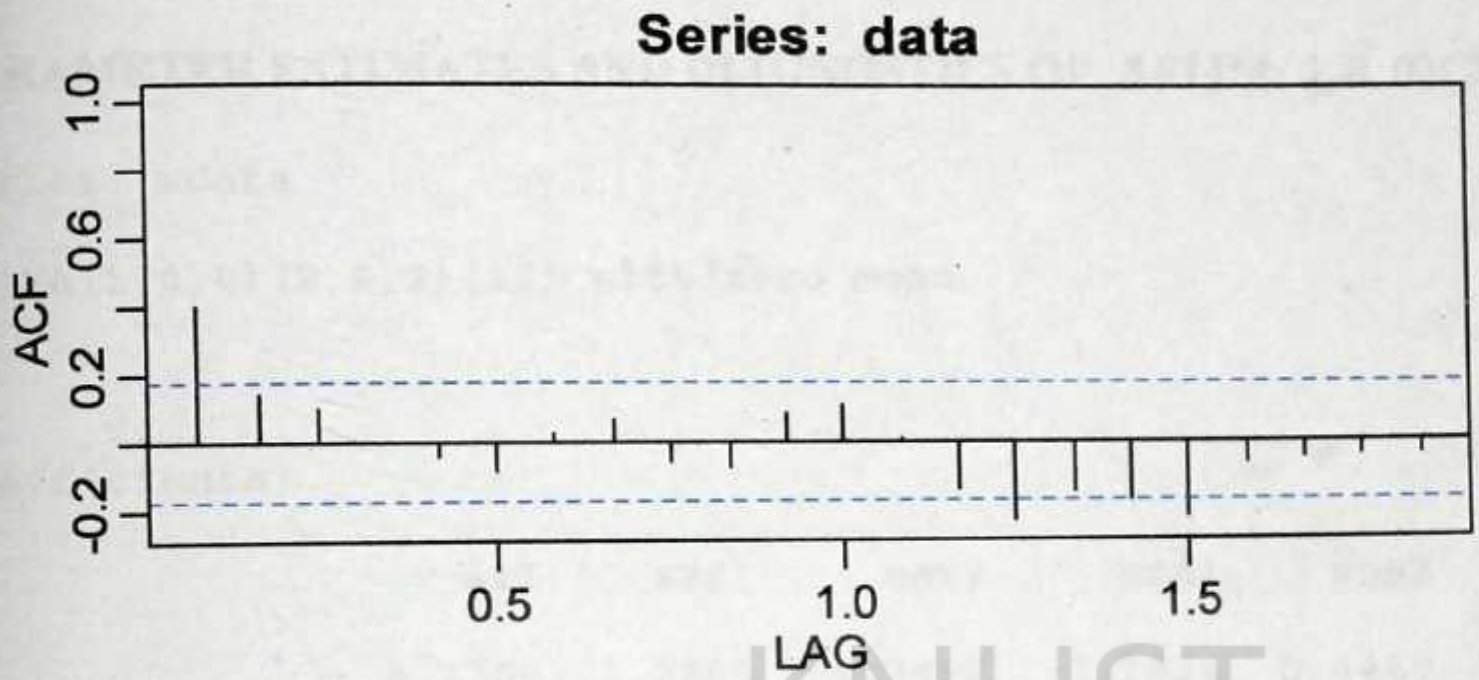


Figure 4.1 Stochastic Time series Graph of the OPD Malaria Data

Figure 4.1 depict a trend in the OPD malaria data. The exist a seasonal component in the time plot. The OPD malaria data was stationary using the KPSS test with mean zero and a constant variance.

MODEL IDENTIFICATION



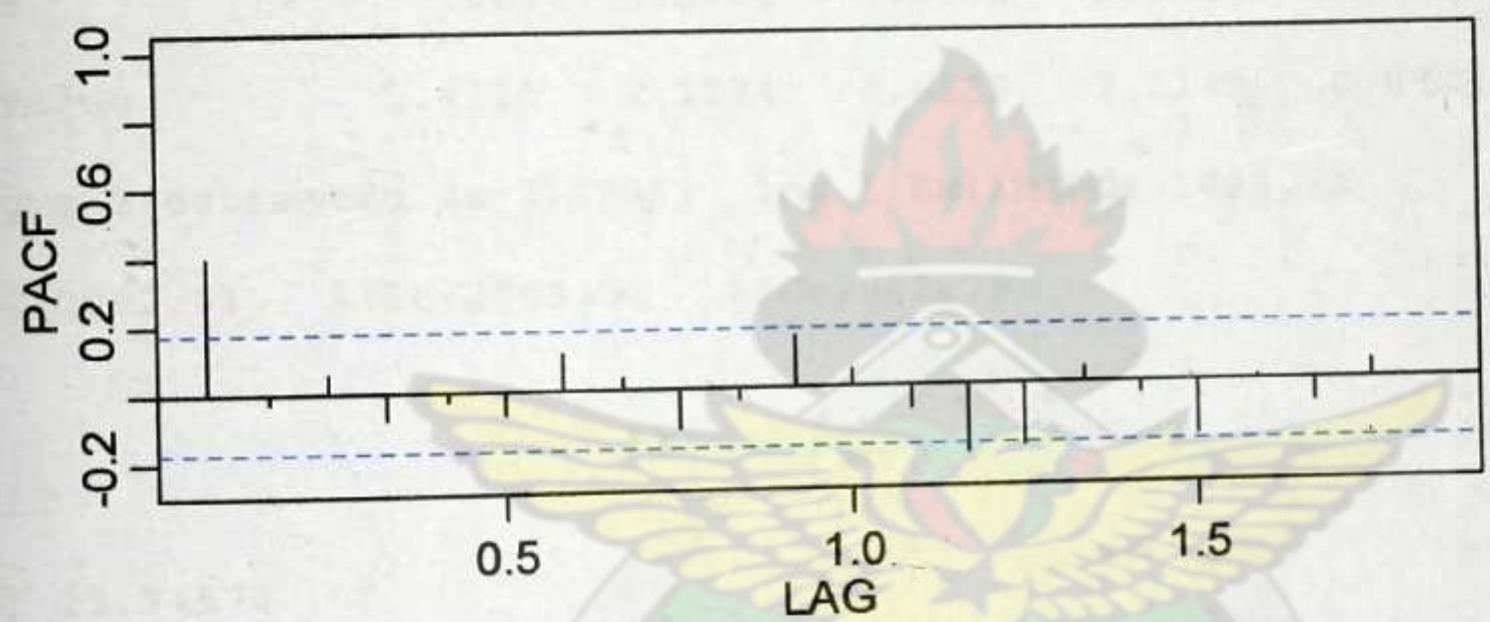


Figure 4.2 Autocorrelation function (ACF) and the Partial Autocorrelation function of OPD data

Figure 4.2 shows the Autocorrelation function (ACF) and the Partial Autocorrelation function of the malaria data. The PACF cuts at lag one and the ACF also cuts at lag one. Inspecting the PACF and the ACF the following models were suggested.

- $ARIMA(1,0,0)(2,0,2)_{[12]}$
- $ARIMA(1,0,0)(2,0,0)_{[12]}$
- $ARIMA(1,0,0)(0,0,2)_{[12]}$

4.3 MODEL SELECTION

PARAMETER ESTIMATES AND DIAGNOSTICS OF $ARIMA(1,0,0)(2,0,2)_{[12]}$

```
Series: xdata
ARIMA(1,0,0)(2,0,2)[12] with zero mean


Coefficients:
          ar1     sar1     sar2     sma1     sma2       xmean
       0.4367   1.2286  -0.3492  -1.1934   0.4460  2168.9010
s.e.   0.0804   0.5786   0.5693   0.5643   0.5112   144.9215
t-value 5.4316   2.1234   0.6133   2.1148   0.8725

sigma^2 estimated as 313286:  log likelihood=-1024.31
AIC=2062.61   AICc=2063.51   BIC=2082.79


$AIC
[1] 13.74578


$AICc
[1] 13.76778


$BIC
[1] 12.87682
```

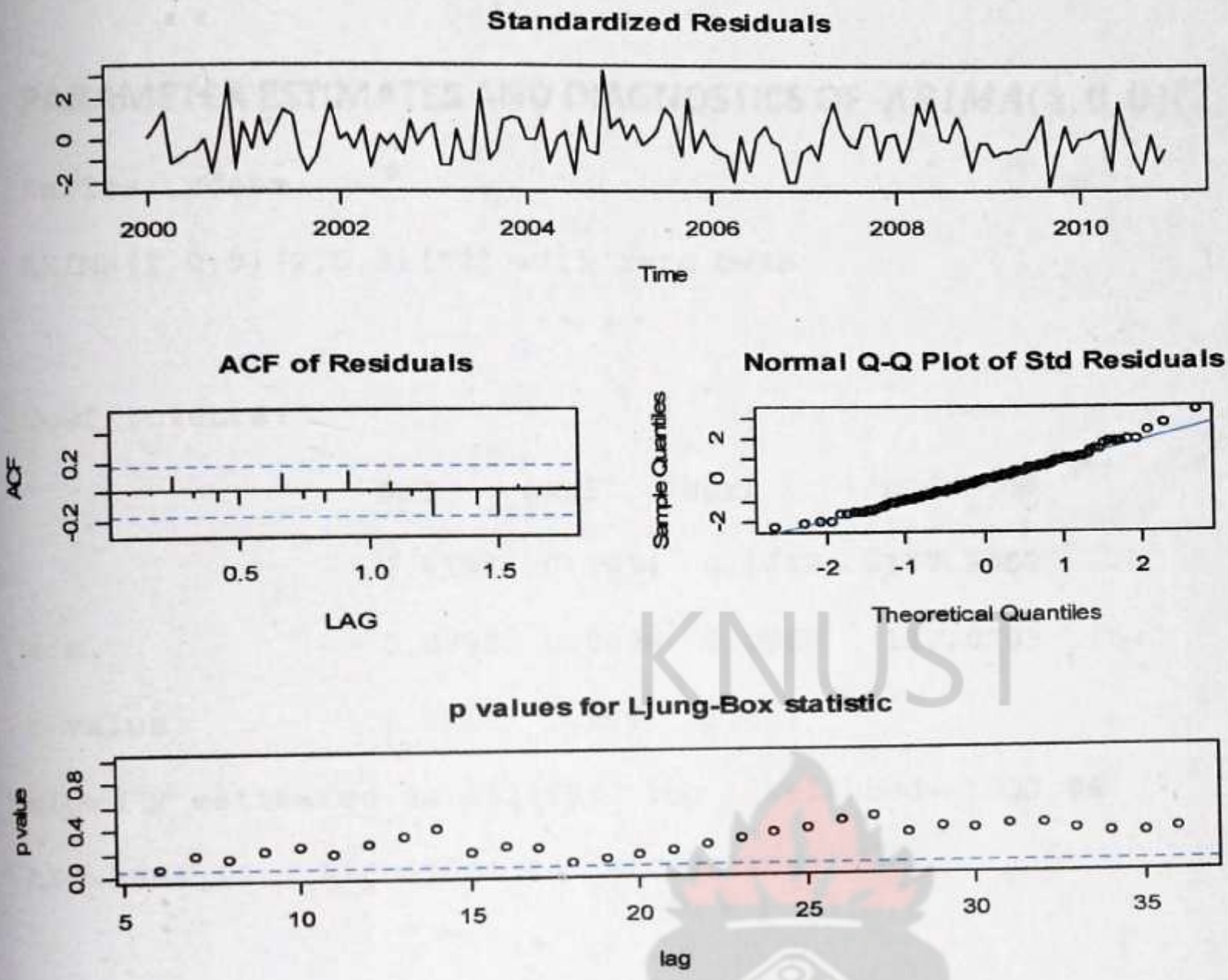


FIGURE 4.3 DIAGNOSTICS OF $\boldsymbol{ARIMA(1,0,0)(2,0,2)_{[12]}}$

a. The standardized residuals pot in the figure 4.4 above supports the model, since there is no evidence of trend thus the standardized plot exhibits no known pattern.

b. The plot of the ACF of the residuals the lags shows no statistically significant evidence of non – zero autocorrelation in the residuals.

c. With the normal q- q plot of the standardized residual, most of the residuals seems to follow the line of best fit fairly closely except for some few residuals deviating from the normality. Since most of the residuals are located on the straight line, we conclude that the normality assumption has also been satisfied.

d. The plot of the Ljung – Box statistic shows that the Ljung – Box p – values are all greater than 0.05, thus the Ljung – Box statistic is not significant at any positive lag.

PARAMETER ESTIMATES AND DIAGNOSTICS OF $ARIMA(1,0,0)(2,0,0)_{[12]}$

```
Series: xdata
ARIMA(1,0,0)(2,0,0)[12] with zero mean

Coefficients:
                ar1     sar1     sar2       xmean
              0.4198   0.1034   0.1643   2177.8057
s.e.          0.0795   0.0896   0.0968    112.0703
t-value       5.281    1.154    1.697

sigma^2 estimated as 332965:  log likelihood=-1027.06
AIC=2064.12   AICc=2064.6   BIC=2078.53

$AIC
[1] 13.7764

$AICc
[1] 13.79516

$BIC
[1] 12.86376
```

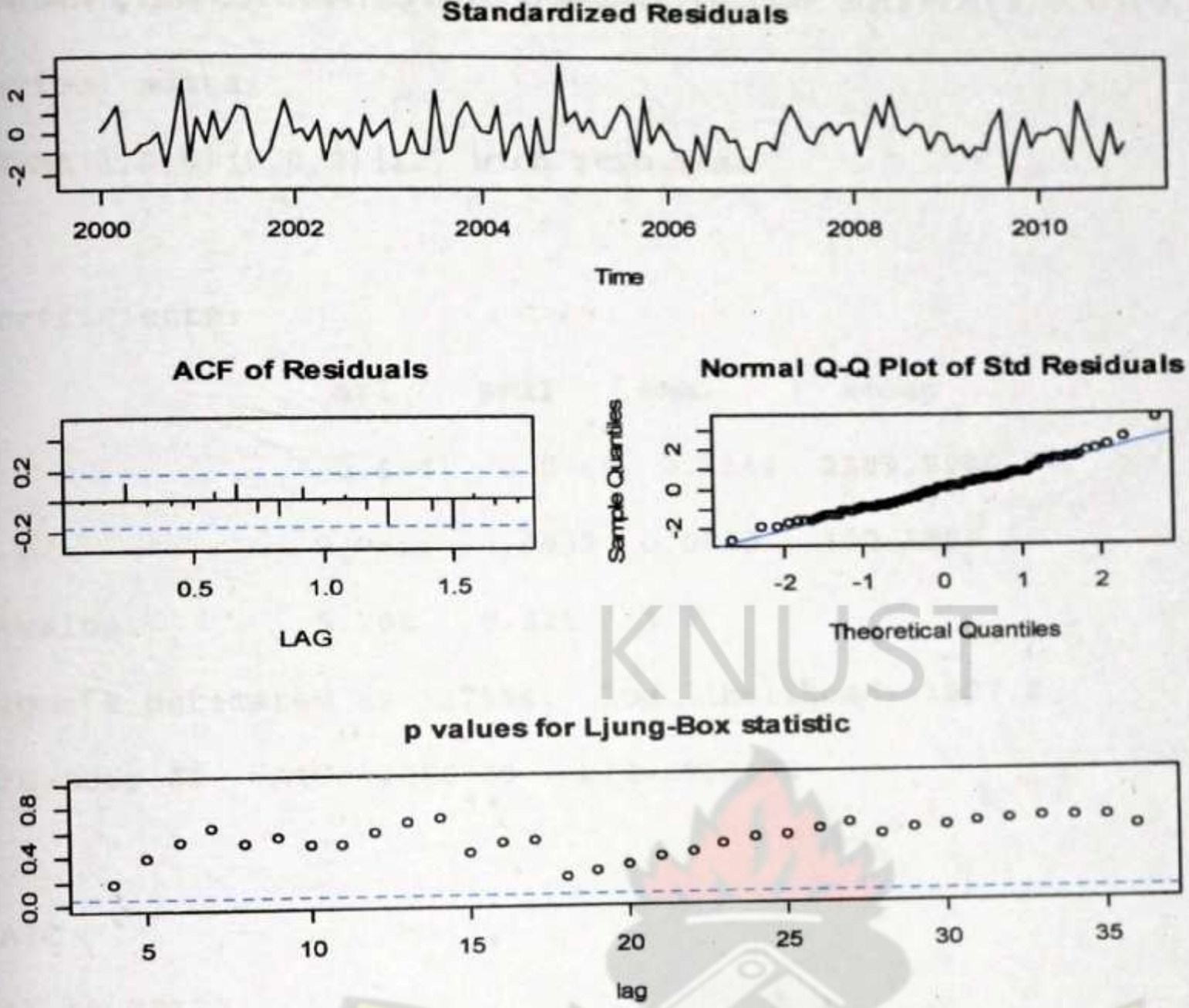


FIGURE 4.4 DIAGNOSTICS OF $\boldsymbol{ARIMA(1,0,0)(2,0,0)_{[12]}}$

a. The standardized residuals pot in the figure 4.5 above supports the model, since there is no evidence of trend thus the standardized plot exhibits no known pattern.

b. The plot of the ACF of the residuals the lags shows no statistically significant evidence of non – zero autocorrelation in the residuals.

c. With the normal q- q plot of the standardized residual, most of the residuals seems to follow the line of best fit fairly closely except for some few residuals deviating from the normality. Since most of the residuals are located on the straight line, we conclude that the normality assumption has also been satisfied.

d. The plot of the Ljung – Box statistic shows that the Ljung – Box p – values are all greater than 0.05, thus the Ljung – Box statistic is not significant at any positive lag.

PARAMETER ESTIMATES AND DIAGNOSTICS OF $ARIMA(1,0,0)(0,0,2)_{[12]}$

```
Series: xdata

ARIMA(1,0,0)(0,0,2)[12] with zero mean


Coefficients:

                 ar1     sma1     sma2       xmean

                0.4141  0.0769   0.1144  2183.5989

s.e.            0.0796  0.0937   0.0880   100.1829

t-value         5.202   0.821    1.3

sigma^2 estimated as 337954:  log likelihood=-1027.8

AIC=2065.61   AICc=2066.08   BIC=2080.02


$AIC

[1] 13.79127


$AICc

[1] 13.81003


$BIC

[1] 12.87863
```

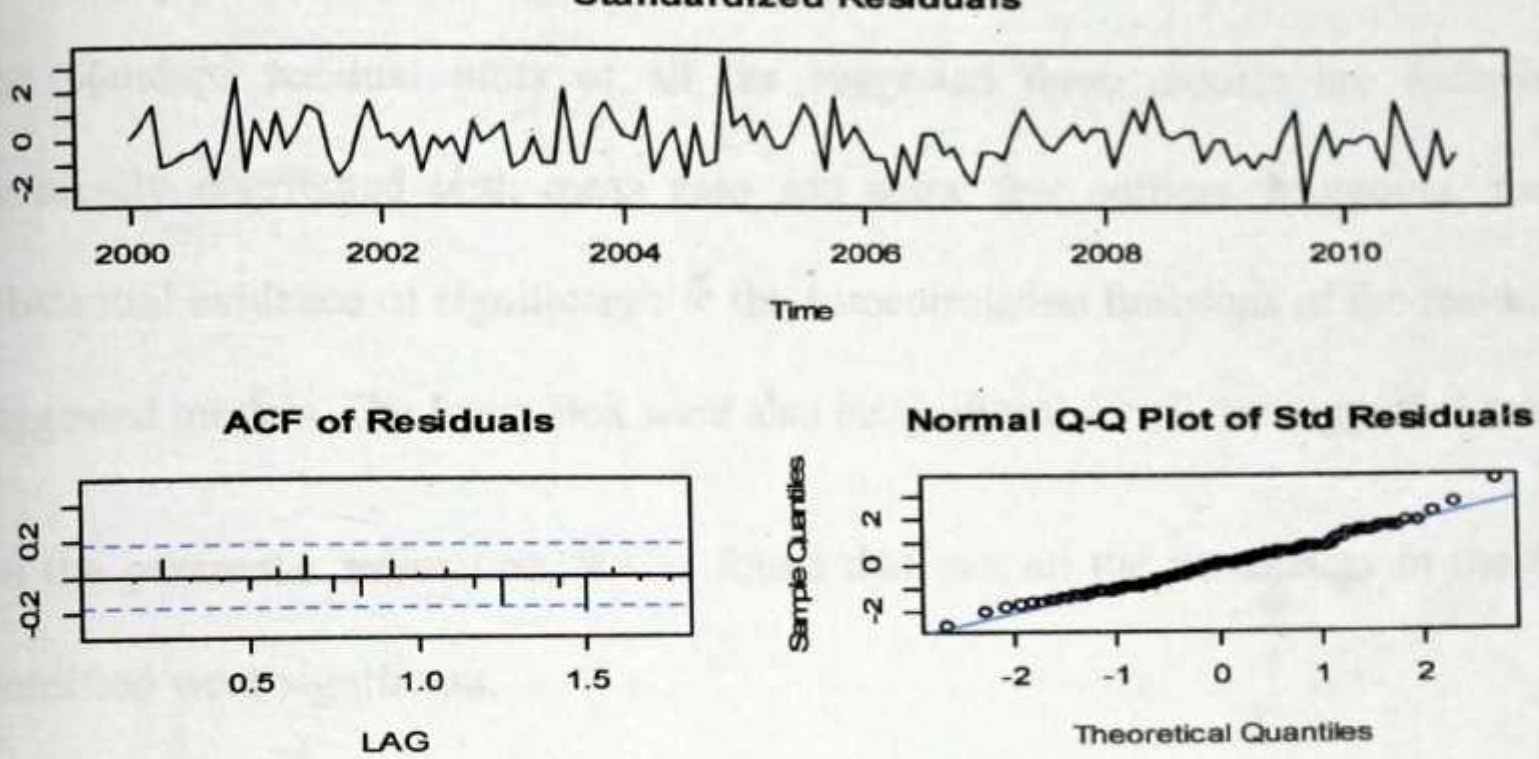


FIGURE 4.5 DIAGNOSTICS OF $\boldsymbol{ARIMA(1,0,0)(2,0,0)_{[12]}}$

a. The standardized residuals pot in the figure 4.6 above supports the model, since there is no evidence of trend thus the standardized plot exhibits no known pattern.

b. The plot of the ACF of the residuals the lags shows no statistically significant evidence of non – zero autocorrelation in the residuals.

c. With the normal q- q plot of the standardized residual, most of the residuals seems to follow the line of best fit fairly closely except for some few residuals deviating from the normality. Since most of the residuals are located on the straight line, we conclude that the normality assumption has also been satisfied.

d. The plot of the Ljung – Box statistic shows that the Ljung – Box p – values are all greater than 0.05, thus the Ljung – Box statistic is not significant at any positive lag.

4.4 SELECTION OF BEST MODEL FOR FORECASTING

The standard residual plots of all the suggested three models are independently and identically distributed with mean zero and some few outliers. Moreover, there were no substantial evidence of significance in the autocorrelation functions of the residuals of all the suggested models. The Lung-Box were also insignificant for all the suggested models.

On the parameter estimation, it was found that, not all the parameters in the three models identified were significant.

Comparing the AIC, BIC and AIC_C the model $ARIMA(1,0,0)(2,0,2)_{[12]}$ is the best model.

4.5 FORECASTING

Forecasting for 24 months

```
$pred
          Jan       Feb       Mar       Apr       May       Jun
Jul       Aug
2011 1945.609 1816.855 1907.004 2013.872 2198.031 2327.150
2352.204 2381.660
2012 2159.845 1935.169 1979.216 2084.009 2118.780 2383.731
2383.851 2336.962
          Sep       Oct       Nov       Dec
2011 2014.210 2209.046 2090.894 2050.250
2012 1894.947 2131.144 1998.992 1976.092

$se
          Jan       Feb       Mar       Apr       May       Jun
Jul       Aug
2011  559.7276  610.7639  620.0190  621.7681  622.1011  622.1646
622.1767 622.1790
```

2012 622.4882 622.5479 622.5593 622.5615 622.5619 622.5620
622.5620 622.5620

	Sep	Oct	Nov	Dec
2011	622.1794	622.1795	622.1794	622.1787
2012	622.5620	622.5620	622.5620	622.5619

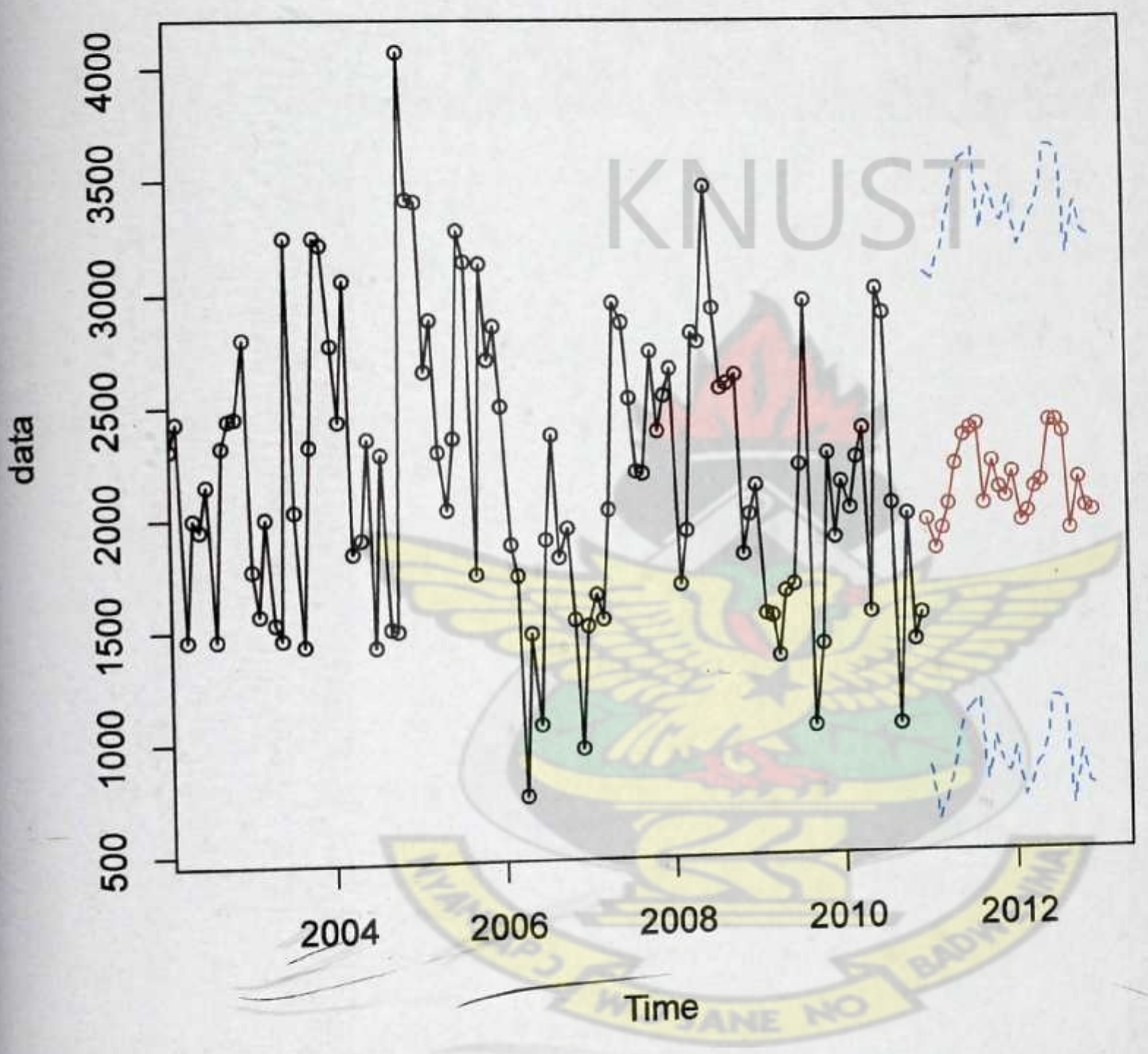


FIGURE 4.6 FORECAST TREND FOR 24 MONTHS PERIOD

Figure 4.6 above gives the visual representation of the OPD Malaria cases over the years, its forecast in the middle (red) and confidence interval (up and down in blue lines).

4.6 SUMMARY

In this chapter, we considered data collection analysis of the study and forecasted by using the best model.

In the next chapter, we shall present the conclusion and recommendation of the study.

CHAPTER FIVE

5.0 CONCLUSION

This study is an attempt to select the best and accurate model among various ARIMA models which posse's high power of predicting the forecast of Malaria rate. Malaria has always been the subject of research for medical practitioners from time immemorial. The notion malaria cannot be minimized should be the thing of the past. This study was carried out to develop forecasting and predication models of malaria cases in the Out Patient Department (OPD) of Komofo Anokye Teaching Hospital (KATH). By the use of Box-Jenkins methodology, $ARIMA(1,0,0)(2,0,2)_{12}$ was the best suggested model since most of the parameters were all significant. It was concluded that the model bases on 24 steps forecast into the future shows that the OPD malaria cases would have a random pattern thus there it will increase and decrease.

5.1 RECOMMENDATION

Malaria is one of the dreadful diseases in the world. Malaria is ranked second after HIV/AIDS in sub-Saharan African. In Ghana every three hours somebody die of malaria(MOH). Again, diseases in children, malaria are rank third in Africa (WHO). But the notion that malaria cannot be minimized should be thing of the past.

Malaria can be controlled if:

Effective antimalarial drugs should be given to every pregnant woman to prevent them from stillbirths.

The pharmaceutical company of Ghana should come out with vaccine which may be given to pregnant women and infants to prevent them from morbidity and mortality.

There should be free given insecticides treated mosquito nets to pregnant women and young children.

There should be education on malaria and it effects on the nation.

5.2 SUMMARY

The final chapter considered the summary of the study, conclusion and recommendation of the study.

REFERENCES

1. Acemoglu, D., and Simon J., (2006). Disease and Development: The Effect of Life Expectancy on Economic Growth. Working Paper 12269, National Bureau of Economic Research,Cambridge, Mass. June. Pg 1-123

2. Barreca, Alan (2007). "The Long-Term Economic Impact of In Utero and Postnatal Exposure to Malaria." Technical Report, UC-Davis. Pg 1-150

3. Bleakley, C. Hoyt (2007a). "Disease and Development: Evidence from Hookworm Eradication in the American South." The Quarterly Journal of Economics. February 122 (1), 73{117. Pg 1-200

4. Biritwum R. B., Welbeck J. and Barnish G. (2000) "Incidence and management of malaria in two communities of different socio-economic level, in Accra, Ghana". *Ann Trop Med Parasitol"* **94**: 771-8. Pg 1-220

5. Bosello F., Roson R. , Tol R.S.J. (2006), Economy-wide estimates of the implications of climate change: human health, *Ecological Economics*, 58, 579– 591. Pg 1-230

6. Brinkman U. and Brinkman A. (1991) Malaria and health in Africa: The present situation and epidemiological trends. *Tropical Medicine and Parasitology* **42**. Pg 1-240

7. Bryan RT, Balderrama F, Tonn RJ, Pinto Dias JC. 1994. "Community participation in vector control: lessons from Chagas' disease". Am. J. Trop. Med. Hyg. 50(Suppl):61–71. Pg 1-250

8. Cutler, David, Winnie Fung, Michael Kremer, and Monica Singhal (2007). Mosquitoes: "The Long-Term Effects of Malaria Eradication in India." May. Unpublished manuscript. Pg 1-260

9. Conly, Gladys N. (1975). The Impact of Malaria on Economic Development: A Case Study. Scientific Publication 297, Pan American Health Organization, Washington, D.C. Pg 1-270

10. Curriero F.C., Heiner K.S., Samet J.M., Zeger S.L., Strug L., Patz J.A. (2002), Temperature and mortality in 11 cities of the Eastern United States, *American Journal of Epidemiology*, 155, 80-87. Pg 1-280

11. Chavasse DC, Yap HH, eds. 1997. Chemical methods for the control of vectors and pests of public health importance. WHO/CTD/WHOPES/97.2. WHO, Geneva. Pg 1-290

12. Chaves L.F., Pascual M. (2007), Comparing models for early warning systems of neglected tropical diseases, *PLoS Neglected Tropical Diseases*, 1, 1-6, doi:10.1371/journal.pntd.0000033. Pg 1-300

13. Chaves L.F., Pascual M. (2007), Comparing models for early warning systems of neglected tropical diseases, *PLoS Neglected Tropical Diseases*, 1, 1-6, doi:10.1371/journal.pntd. Pg 1-310

14. De Savigny D., Mwageni E., Masanja H., Juma Z., Momburi D., Mkilindi Y., Mbuya C., Kasale H. and Reid G. (2002). Household wealth ranking and risks of malaria mortality in rural Tanzania. Third MIM Pan-African Malaria Conference, ICC, Arusha, Tanzania.

15. McCombie S. C. (1996) Treatment seeking for malaria: a review of recent research. *Social Science and Medicine* **43**: 933-945. Pg 1-320

16. Fletcher M, Teklehaimanot A, Yemane G. 1992. Control of mosquito larvae in the port city of Assabby an indigenous larvivorous fish, Aphanius dispar. Acta Tropica 52:155–166. Pg 1-330

17. Farida A., Radina P., Najibullah S and Richard K. (2010) "Towards malaria risk prediction in Afghanistan using remote sensing": Malaria journal. Doi. 1186/1475. Pg 1-340

18. Gratz NG, Pal R. 1988. Malaria vector control: larviciding. Pp. 1213–1226, In: Malaria: Principles and practice of malariology. Wernsdorfer WH, McGregor I, eds. Churchill Livingstone, Edinburgh, UK. Pg 1-350

19. Guerin, PJ., P. Olliaro, F. Nosten, P. Druilhe, R. Laxminarayan, F. Binka, WL. Kilama, N. Ford, and NJ White (2002). Malaria: current status of control, diagnosis, treatment, and a proposed agenda for research and development. *The Lancet Infectious Diseases* 2(9). Pg 1-360

20. Hong, Sok Chul (2007). Health and Economic Burden of Malaria in Nineteenth-Century America. Doctoral dissertation, University of Chicago. (proposed June 2006). Pg 1-370

21. Hanson, K., C. Goodman, J. Lines, S. Meek, D. Bradley and A. Mills (2004) "The Economics of Malaria Control Interventions" World Health Organization. Pg 1-380

22. Hijioka Y., Takahashi K., Matsuoka Y., Harasawa H. (2002), "Impact of global warming on waterborne disease", *Journal of Japan Society on Water Environment*, 25, 647-652. Pg 1-390

23. Kovats R.S., Campbell-Lendrum D., Matthies F. (2005), "Climate change and human health: estimating avoidable deaths and disease", *Risk Analysis*, 25, 1409–1418. Pg 1-400

24. Kazembe L.N., Kleinschmidt I., Sharp B.L. (2006), "Patterns of malaria-related hospital admissions and mortality among Malawian children: an example of spatial modelling of hospital register data", *Malaria Journal*, 5, 93. Pg 1- 410

25. Keatinge W.R., Donaldson G.C., Cordioli E., Martinelli M., Kunst A.E., Mackenbach J.P., Nayha S., Vuori I. (2000), "Heat related mortality in warm and cold regions of Europe: observational study". *British Medical journal*, 321, 670-673. Pg 1-420

26. Lopez Alonso, Moramay and Ra_ul Porras Condey (2003). "The Ups and Downs of Mexican Economic Growth". The Biological Standard of Living and Inequality, 1870-1950." Journal of Economics and Human Biology. 1 (2) . Pg 1-430

27. Lindsay S.W., Martens W.J.M. (1998), Malaria in the African highlands: past, present and future, *Bulletin of the World Health Organization*, 76, 33–45. Pg 1-440

28. Lacey LA, Lacey CM. 1990. The medical importance or riceland mosquitoes and their control using alternatives to chemical insecticides. J. Am. Mos. Control Assoc. 6(Suppl.):1–93. Pg 1-450

29. Llanos-Zavalaga, F, E. Huayta, and L. Lecca (2005) "Factores Asociados al Uso de Mosquiteros en el Departamento de Piura, Perú" *Revista Medica Herediana*, 16(2). Pg 1-460

30. Lindsay SW, Snow RW. 1988. The trouble with eaves: house entry by vectors of malaria. Trans Royal Soc. Trop. Med Hyg. 82:645–646. Pg 1-470

31. Malaney, Pia (2003). \Micro-Economic Approaches to Evaluating the Burden of Malaria.Working Paper 99, Center for International Development at Harvard University, Cambridge, Mass. April.

32. Meltzer, David (1992). Mortality decline, the Demographic Transition, and Economic Growth. Doctoral dissertation, University of Chicago December. Pg 1-480

33. Miller, N. Grant (2005). Contraception as Development, New Evidence from Family Planning in Colombia. Working Paper 11704, National Bureau of Economic Research, Cambridge, Mass. Pg 1-490

34. Malaney, P., Spielman, A. & Sachs, J. (2004), 'The malaria gap', Am. J. Trop. Med. Hyg. 71. Pg 1-500

35. Medina D.C., Findley S.E., Guindo B., Doumbia S. (2007), "Forecasting non-stationary diarrhea, acute respiratory infection, and malaria time-series in Niono, Mali", *PLoS ONE*, 2, 1-13. doi: 10.1371/journal.pone.0001181. Pg 1-510

36. Najera J. A. (1994) The control of tropical diseases and socioeconomic development (with special reference to malaria and its control). *Parassitologia* **36**: 17-33. Pg 1-520

37. Patz J.A., Strzepek K., Lele S., Hedden M., Greene S., Noden B., Hay S.I., Kalkstein L., Beier J.C. (1998), "Predicting key malaria transmission factors, biting and entomological inoculation rates, using modeled soil moisture in Kenya", *Tropical Medicine and International Health*, 3, 818-827. Pg 1-530

38. Patz J.A., Lindsay S.W. (1999), New challenges, new tools: the impact of climate change on infectious diseases, *Current Opinion in Microbiology*, 2, 445-451. Pg 1-540

39. Pattanayak, S., Dickinson, K., Corey, C., Murray, B., Sills, E., & Kramer, R. 2006.

40. Deforestation, malaria and poverty: a call for transdisciplinary research to support the design of cross-sectoral policies". Sustainability: Science, Practice, & Policy 2(2):45–56. Pg 1-550 http://sspp.proquest.com/archives/vol2iss2/0512-022.pattanayak.html

41. Ribero, Roc__o and Jairo Nu~nez (2000). Adult Morbidity, Height, and Earnings in Colombia. In W. D. Savedo_ and T. P. Schultz, eds., Wealth from Health: Linking Social

Investments to Earnings in Latin America, Washington, D.C.: Inter-American Development Bank.

42. Sanchez Torres, Fabio and Jairo Nu~nez (2000). Geography and Economic Development: A Municipal Approach for Colombia." Revista de Planeaci_on y Desarollo. July-December XXXI (3-4). Pg 1-560

43. Rashed S., Johnson H., Dongier P., Moreau R., Lee C., Lambert J. and Schaefer C. (2000) "Economic impact of febrile morbidity and use of permethrin-impregnated bed nets in a malarious area II. Determinants of febrile episodes and the cost of their treatment and malaria prevention". *Am J Trop Med Hyg* **62**: 181-6. Pg 1-570

44. Rozendaal, J.A. 1997. ''Vector control: Methods for use by individuals and communities". World Health Organization, Geneva. Pg 1-580

45. Rodó X., Pascual M., Fuchs G., Faruque A.S.G. (2002), "ENSO and cholera: a nonstationary link related to climate change?", *PNAS*, 99, 12901-12906. Pg 1-590

46. Shanks G.D., Hay S.I., Stern D.I., Biomdo K., Snow R.W. (2002), "Meteorologic influences on Plasmodium falciparum malaria in the highland tea estates of Kericho, Western Kenya", *Emerging Infectious Diseases*, 8, 1-10. URL: http://www.cdc.gov./ncidod/EID/vol8no12/02-0077.htm. Pg 1-600

47. Sachs J. and Malaney P. (2002) . The economic and social burden of malaria. *Nature* **415**: 680-5. Pg 1-610

48. Sychareun V., Phengsavanh A., Kitysivoilaphanh B., Prabouasone K., Viriyavejakul P., Krudsood S., Phophak N. and Looareesuwan S. (2000) A study of anaemia in pregnant women with Plasmodium falciparum at district hospitals in Vientiane, Lao PDR. *Southeast Asian J Trop Med Public Health* **31 Suppl 1**: 91-8. Pg 1-620

49. Shakoor H., Armstrong B., Baccini M., Biggeri A., Bisanti L., Russo A., Paldy A., Menne B., Kosatsky T. (2006), Impact of high temperature on mortality: is there an added heat wave effect?, *Epidemiology*, 17, 6, 632-638. Pg 1-630

50. Thomas, Duncan, Elizabeth Frankenberg, Jed Friedman, Jean-Pierre Habicht, Mohamme Hakimi,

51. Jaswadi, Nathan Jones, Gretel Pelto, Bondan Sikoki, Teresa Seeman, James P. Smith, Cecep Sumantri, Wayan Suriastini, and Siswanto Wilpo (2003). \Iron de_ciency and the well-being of older adults: Early results from a randomized nutrition intervention." April. Unpublished manuscript. Pg 1-640

52. Tol R.S.J. (2002), "Estimates of the damage costs of climate change. Part 1: Benchmark estimates", *Environmental and Resource Economics,* 21, 47–73. Pg 1-650

53. Tian, L., Bi, Y., Ho, S and Sung, J., (2008) "One year delayed effect of of fog on malaria transmission: a time series analysis in the rain forest area of Mengla county, South-west China". Malaria journal. (2008).PMC2441628, Lincence BioMed central Ltd pgs 1 214. Pg 1-660

54. Utzinger, J urg, Yesim Tozan, Fadi Doumani, and Burton H. Singer (2001)."The economic payo_s of integrated malaria control in the Zambian copperbelt between 1930 and 1950. Tropical Medicine and International Health. August 7 (8), 657{677. Pg 1-670

R - codes

```
data=ts(b,start=2000,freq=12)

plot(data)

acf2(data)

sarima(data,1,0,0,2,0,2,12)

 sarima(data,1,0,0,0,0,2,12)

sarima(data,1,0,0,2,0,0,12)

sarima.for(data,,24,1,0,0,2,0,2,12)
```