MULTIVARIATE TIME SERIES ANALYSIS OF HYPERTENSION AND HEART DISEASE: A CASE STUDY OF THE HO MUNICIPAL HOSPITAL

BY

BAETA FRANCIS DELALI

(10443480)

THIS THESIS IS SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES, UNIVERSITY OF GHANA IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE AWARD OF THE MASTER OF PHILOSOPHY DEGREE IN STATISTICS

JULY 2015

DECLARATION

Candidate's Declaration

This is to certify that, this thesis is the result of my own research work and that no part of it has been presented for another degree in this University or elsewhere

SIGNATURE:……………………… DATE:……………………………………………

BAETA FRANCIS DELALI

(10443480)

Supervisors' Declaration

We hereby certify that this thesis was prepared from the candidate's own work and supervised in accordance with guidelines on supervision of thesis laid down by the University of Ghana.

SIGNATURE : ………………………….. DATE : ……………………………………

DR EZEKIEL NII NOYE NORTEY

(Principal Supervisor)

SIGNATURE : ……………………….. DATE : ………………………………………

DR SAMUEL IDDI

(Co-Supervisor)

ABSTRACT

Concerns about the dangers of Cardiovascular Diseases (CVD) have been a major problem all over the world. Hypertension, which is the commonest CVD is gradually usurping the minds of public health workers globally. The disease is the commonest cause of morbidity and mortality all over the world. It is also the commonest cause of heart disease. In this study, we employ the Vector Autoregressive (VAR) modeling approach to model and to forecast hypertension and heart disease cases in the Ho municipality. The data was obtained from the municipal hospital and it spanned from the 1st January 2010 to 31st December 2014. The results revealed that VAR (2) was the best to model and forecast these CVDs. Diagnostic checks revealed the model is free of serial correlations and conditional heteroscedasticity. It also passed the stability test. The model was proposed for predicting hypertension and heart disease cases in the Ho municipality for the year 2015. The study reveals that there will be slight increase in cases of both conditions of hypertension and heart disease. A lot of sensitization programs must be organized by health workers and stake holders in the catchment of Ho so as to create awareness. This can prompt people to go for early checkup so that should they be positive, they can start medications early before it results into complications such as heart disease and stroke.

DEDICATION

This work is dedicated to my father Christian Kofi Baeta

ACKNOWLEDGEMENT

My sincerest gratitude goes to the Almighty God for all the travelling mercies and for his grace upon my life throughout my study in this University.

My profound gratitude and appreciation also goes to my supervisors, Dr. E.N.N. Nortey and Dr Samuel Iddi whose valuable suggestions and criticisms has helped enrich my work.

I also thank all the lecturers of Statistics Department, especially Dr. Doku Amponsah for their services and pieces of advice throughout my years of study in this University. Again my special thanks goes to my dad, Christian Baeta, my wife Irene, my daughter Deborah, my mum Vivian, and my siblings Nic, Adwoa, and Nana.

To my cousin Marcus, I say thank you for your support, encouragement, advice and for particularly being like a father to me during the period of my study here in Accra. May the good Lord continue to bless you.

Finally, I say a big thank you to Mr. Emmanuel Nyavi for his support and also to my good friend Solomon Gala for his contribution towards my work.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

HBP ……………………………….High Blood Pressure

BP ……………………………… Blood Pressure

CVD …………………………… Cardiovascular Disease

VAR ……………………………. Vector Auto Regressive

AR ……………………………… Auto Regressive

MA …………………………..… Moving Average

IHD ……………………………… Ischemic Heart Disease

NCD ……………………………. Non Communicable Disease

WHO ………………..………….. World Health Organization

GHS ……………..……………… Ghana Health Service

WHS ……………………………. World Health Survey

ARIMA ………………………..…. Auto Regressive Integrated Moving Average

ADF ……………………...……. Augmented Dickey Fuller

PP ……………………………….. Phillip Perron

AIC ………………….………… Akaike Information Criteria

BIC ……………………….....….. Bayesian Information Criterion

HQ ……………………………….. Hannan Quin

FPE …………………………….. Final Prediction Error

ARCH …………………………… Auto Regressive Conditional Heteroscedasticity

GARCH …………………..……… General Auto Regressive Conditional Heteroscedasticity

VARMA ………………………….. Vector Auto Regressive Moving Average

ACF ……………………………….. Auto Correlation Function

PACF ……………………………… Partial Auto Correlation Function

VMA …..………………………….. Vector Moving Average

LM …………………….………….. Lagrange Multiplier

CUSUM ……………..………….. Cumulative Sum of Recursive Residuals

KPSS ……………..…….………….. Kwiatkowski Phillip Schmidt Shin

CHAPTER ONE

INTRODUCTION

1.0 Background of the Study

A time series is defined as a set of quantitative observations that has been arranged in a sequential or in a chronological order.

Data obtained from observations collected sequentially or chronologically over time are extremely common. For example, in business weekly interest rates, daily closing stock prices, monthly price indices, year sales figures and so forth, are observed. In meteorology daily high and low temperatures, annual precipitation and drought indices, and hourly wind speeds, are observed. Also in agriculture, annual figures for crop and livestock production, soil erosion, and export sales are recorded. The list of areas in which time series is applied is endless.

The purpose of time series analyses is generally two folds: to understand or model the stochastic mechanism that gives rise to an observed series and to predict or forecast the future values of a series based on the history of that series and, possibly, other related series or factors.

A time series may be univariate or multivariate. It is univariate when it involves only one variable and multivariate if it simultaneously studies more than one variable. For example in this study two variables shall be considered which are hypertension and heart disease.

Multivariate models predict outcomes of situations that are affected by more than one variable. A somewhat unique feature of time series and their model is that one usually cannot assume that the observations arise independently from a common population (or from populations with different means, for example). Studying models that incorporate dependence is a key concept in time series analysis. As already mentioned above this study involves two variables: hypertension and heart disease.

1.0.1 Definition of variables (diseases)

Hypertension is the term used to describe High Blood Pressure (HBP). Throughout this work the words hypertension and high blood pressure will often be used interchangeably.

When the heart beats it contracts to pump blood and the pumped blood circulates throughout the body in blood vessels or arteries. As blood circulates through the blood system it exerts pressure on the walls of the vessels or arteries and it is this pressure, however strong or weak, that is referred to as blood pressure. Blood Pressure (BP) is not constant at all times. It varies based on activities being carried out by an individual at a particular point in time or based on certain health conditions. But when this pressure is most of the time high in an individual, then the fellow is said to be suffering from HBP, or more formally hypertension. “Hypertension occurs as a result of long duration of abnormal pressure in the main arteries”, (Cunha, 2011). Hypertension is the single most important risk factor for all other Cardiovascular Diseases, especially stroke and heart disease.

BP readings are measured in millimeters of mercury (mm HG) with an instrument known as sphygmomanometer. This consist of an inflatable rubber cuff, an air pump and a column of mercury, which reflects the blood pressure of an individual. The blood pressure readings are usually given in two numbers which appears like a fractional expression. For example 130 over 70, and this is professionally written as 130/70 mmHg. The number at the top represents the systolic blood pressure and the one below is the Diastolic Blood Pressure (DBP). Systolic Blood Pressure (SBP) is the pressure in the blood vessels or arteries during a heartbeat, whilst diastolic blood pressure is the pressure exerted by blood when the heart relaxes. (Zareian 2004; Cunha 2011). The systolic blood pressure is considered to be high if it is 140 mmHg or above, but readings below 120 mmHg are considered as normal pressure. On the other hand a diastolic pressure is considered as high if it is 90 mmHg and above but readings below 80 mmHg are considered normal. And so clearly, a person can be said to be hypertensive if his or her blood pressure is consistently above the normal threshold of 140/90 mmHg. Readings between 120/80 and 139/89 are termed pre-hypertension (i.e. the individual is likely to develop HBP). Hypertension can be classified into primary hypertension and secondary hypertension. The primary hypertension which is also known as essential hypertension contributes about 90% to 95% of all hypertension cases. A condition is described as primary hypertension when there is no obvious medical cause for the elevation of the blood pressure. High blood pressure occurring as a result of a consequence of another disorder or a side effect of medication is classified as secondary hypertension. Such disorders may include renal failure or reno vascular disease. This type of blood pressure is evident in about 5% to 10% of hypertension cases (Cunha et al., 2011).

Heart disease (also known as Cardiac disease) is the term used to describe any inability of the heart to function properly and they include:

(a) Coronary artery disease: Is a disease of the blood vessels supplying the heart muscles with blood. This does not allow blood to flow with ease and when this happens it leads to chest pains, a heart attack, or irregular heart rhythms (arrhythmias).

(b) Enlarged left heart: High blood pressure forces the heart to work harder than necessary in order to pump blood to the rest of the body. This causes the left ventricle to thicken and stiffen (left ventricular hypertrophy). These changes limits the ventricles ability to pump blood to the body. This condition increases the risk of heart attack, heart failure and sudden cardiac death.

(c) Heart failure: Over time, the strain on the heart caused by high blood pressure can cause the muscles of the heart to weaken and work less efficiently. Eventually, the overwhelmed heart simply begins to wear out and fail.

1.0.2 Relationship between hypertension and heart disease

Heart diseases are caused by unhealthy arteries or unhealthy blood vessels. An artery is unhealthy when it is narrowed or blocked. Healthy arteries are flexible, strong, and elastic. Their inner lining is smooth so that blood flows freely, supplying vital organs and tissues with adequate nutrients and oxygen. If a person has high blood pressure, the increased pressure of blood flowing through the arteries can gradually cause a damage to the arteries by rupturing it completely. Fats from consumed diets entering the blood stream are collected at the damaged points to form clots and this can totally block blood flow. Cases where the flow of blood is not totally blocked, the vessel is narrowed allowing just a little amount of blood to flow through. These damages can affect arteries throughout the body. If as a result of these happenings the vessels or arteries in the heart are damaged, it leads to heart disease. And so quite clearly hypertension can lead to heart disease. According to the World Health Organization report in the year (2012), one in every three adults suffer from hypertension. The report also stated that about 50% of all deaths from stroke and heart disease could be attributed to high blood pressure alone.

As part of many other aims, this study looks at the relationship between hypertension and heart diseases in the Ho municipality and its environs.

1.1 Statement of Problem

In the year 2002, hypertension accounted for 8.9% of institutional deaths in Ghanaian hospitals (excluding teaching hospitals), compared to malaria, which accounted for 17.1% of the deaths. But in the year 2008, hypertension was the leading cause of reported institutional deaths accounting for 14.5% of institutional deaths compared to malaria which accounted for 13.4% of all deaths, Ghana Health Services (2009). Quite clearly, hypertension cases are on the increase and their impact is even worse than those of malaria. The danger about hypertension is that there are no clear signs associated with it until it results into a cardiovascular disease (CVD), especially stroke or heart disease. According to a 2002 study by Ezzati et al, about two-thirds of all stroke cases for example and about half of all heart diseases globally could be attributed to hypertension. Several studies has pointed to the fact that awareness of hypertension in African countries is very low. And so by extension awareness of its associated diseases may also be lacking. For example in a study conducted in South Africa by (Hale et al.,1998) on a group of CVD patients, only one out of five of the total group understood that hypertension had probably caused their condition. Another issue to worry about is that most people suffering from hypertension are people that are in their prime. From the findings of Lawes et al (2000), most high blood pressure related diseases in Africa and for that matter Ghana occur among individuals aged 40 to 50 years and clearly these are people that are still in their productive stages. This will obviously have adverse effects on most economies of African countries and for that matter Ghana. It will create a lot of social problems as well since many bread winners could be unhealthy to work.

Despite such gloomy pictures, high blood pressure and its related diseases have not been accorded the needed priority (Perkovich et al., 2007).

Awareness of the disease, its related complications and the rate at which they are increasing must be created and that will catalyze initiatives aimed at improving access to early detection and treatment before it leads to irreversible complications. Statistical forecasts are required to help government make prudent and accurate budget allocations for these cardiovascular diseases in these days of economic hardships. Specific forecasts are needed for different areas (or regions)

1.2 Objectives of the study

1.2.1 General Objective

The main objective is to investigate the relationship between hypertension and heart disease in the Ho municipality and its environs.

1.2.2 Specific Objectives

The specific objectives of the study are;

i. to investigate the direction of causality between cases of hypertension and heart diseases,
ii. to study the association between the variables i.e. hypertension and heart disease,
iii. to identify an appropriate VAR model and
iv. to use this identified model to forecast the prevalence of hypertension and heart diseases in the municipality of Ho.

1.3 Significance of the study

The results of this study will benefit academia and researchers by contributing to existing literature. There has been little multivariate analysis of these variables in Ghana and clearly this research work will significantly bridge the gap. Appropriate multivariate models will be estimated to forecast both hypertension and heart disease prevalence and this will be useful in government budgeting and planning as far hypertension and its complicated issues are concerned in the Ho municipality. This work will also serve as basis for any future research in academia.

1.4 Scope of the study

Monthly data spanning from January 2010 to December 2014 will be used in this study. This implies that 60 data points will be considered for each variable. The source of the data is the Ho Municipal Hospital.

1.5 Methodology

The data will be modeled with the Vector Autoregressive (VAR) model. The VAR model is the multivariate form of the univariate Autoregressive (AR) model and so the theory of AR processes is somehow the basis of the VAR process. The AR model shows that there is a relationship between present and past value, a random value.

In dealing with several variables, often the values of one variable may not only relate to its predecessors in time but in addition it may depend on past values of other variables. In response one way to incorporate the influence of one variable on other variables over time is to use the VAR model which was popularized by Box and Tiao (1981). The VAR model interrelate multiple variables to each other in both the contemporary and historically. It is mainly used for structural analysis and forecasting. VAR forecasting treats future cases as a function of past values of the variables included in the VAR model.

The modeling of a VAR process consist of the following 3 key steps:

(i) Identification of the model

(ii) Parameter estimation

(iii) Diagnostic checking

1.6 Thesis Organization

This study will be organized into five chapters. Chapter one introduces the entire research work. It expounds the background of the study, statement problem, objectives of the study, the significance, scope and a brief methodology of the study. Chapter two presents the literature review, which looks at work done by other researchers. Chapter three is concerned with the detailed methodology used in the study. Chapter four is where the data will be analyzed. It consists of graphical and tabulation of results for discussion. Chapter five, looks at the summary of findings, conclusions, limitations, recommendations, and future research proposal of the study.

CHAPTER TWO

LITERATURE REVIEW

2.0 Overview of this chapter

This chapter reviews relevant works associated with hypertension and heart disease that has been done by other researchers. Because this work is more geared towards the statistics of hypertension than the biological or the medical aspect of it, various literature on the anthropology of hypertension will not be considered.

2.1 General literature on Cardiovascular Diseases

Hypertension, stroke and heart diseases fall under a class of diseases known as cardiovascular diseases. Cardiovascular diseases generally refers to conditions that involves narrowed or blocked blood vessels. It involves the heart, the blood vessels (arteries, capillaries, and veins) or both. Hypertension is the leading cause of death globally responsible for over 7.5million deaths annually, which is about 12.8% of the total number of all deaths. It is a major risk factor for coronary heart disease and for both ischemic and hemorrhagic stroke. In their study, Lewington et al., 2002 demonstrated the enormous impact of blood pressure levels on the risk of heart diseases when they demonstrated that the elevation of blood pressure is positively related to increases of heart diseases. The authors did a meta-analysis of data from 61 prospective observational studies that involved almost 1,000,000 individuals with no vascular disease at baseline carried out. This analysis calculated the effect of a 20 mmHg difference in usual Systolic Blood Pressure (SBP) on the risk of stroke and Ischemic Heart Disease (IHD). The authors found that for individuals between the ages of 40 and 89 years, a usual SBP value that was lower by 20 mmHg was associated with significantly lower risk of death from stroke (hazard ratios, 0.36–0.67) and IHD (0.49–0.67).

In the developed countries evidence from large clinical trials has shown a 40% reduction in stroke is associated with treatment and control of hypertension and that was according to Neal (2000).

According to Gaziano (2007), these CVDs are the leading cause of death in developing countries and they cause nearly as many deaths as HIV, malaria and tuberculosis. Bertrand (1999), said hypertension has emerged as a major public health concern, and hypertensive disease accounts for the dominance of stroke.

The proportions of the global burden of diseases due to hypertension has significantly increased from about 4.5% in the year 2002, Kearney et al. (2005), to 7% in the year 2010, Lim et al. (2012).

Hypertension was formerly associated with the developed world but currently, the world wide burden of hypertension is greatest in developing countries where it affects 20% of the adult population. Studies conducted by Seedat (2000), forecasts that by the year 2025 almost 75% of people with hypertension will be in developing countries. Opie and Seedat (2005) attributes these expected increases to globalization and economic advancement which will usher in urbanization and long life expectancies. As the prevalence of hypertension increases it will invariably lead to dramatic increase in the incidence of other CVDs as well and that has the potential to overwhelm the already fragile health systems in poorer nations (WHO, 2005). This will also have financial implications on developing nations because there is increasing evidence from studies done by (Twagirumukiza et al., 2011), in which they found that majority of patients with hypertension will require 2 or 3 drugs to enable them control blood pressure. Despite these gloomy projections, cardiovascular disorders currently receive little or no attention in most African countries (Muna, 1993b). Projections based on recent studies suggest that the management of these disorders will present a major challenge for the overextended and shrinking health budgets of poorer nations in the near future.

According to Brundtland (2002), hypertension is the driver of the CVD epidemic in Africa where it is the major, independent risk factor for heart failure, and stroke. Hypertension is an important cause of heart failure in Ghana, and this is according to (Owusu, 2007).

In Ghana between 28% and 40% of all adults have hypertension. That was according to the Ghana Health Service 2012 report on Non-Communicable Diseases (NCD). It added that in the year 2012, stroke and hypertension were respectively second and third causes of death in Ghanaian hospitals. Analysis of institutional data in Ghana suggests that CVDs have been increasing in both absolute and relative terms. The reported outpatient cases of hypertension in public and mission facilities other than teaching hospitals increased from about 60,000 cases in 1990 to about 600,000 cases in 2009. Hypertension has ranked in the top five outpatient diseases for more than 15 years, accounting for 3.0%-5.0% of all new outpatient diseases across all ages. It ranks as the third most common newly diagnosed outpatient disease among adults. In 2002, hypertension accounted for 8.9% of institutional deaths (excluding teaching hospitals), compared to malaria which accounted for 17.1% of the deaths. But in 2008, hypertension was the leading cause of reported institutional deaths accounting for 14.5% of institutional deaths compared to malaria which accounted for 13.4% of all deaths, Ghana Health Services (GHS), (2009). Quite clearly, hypertension cases are on the ascendency.

Presently in the whole of Africa, the age-specific mortality rates from CVDs are much higher in younger age groups for both men and women, than in the developed world (Unwin, 2001). CVDs are the second leading overall cause of death in Africa, after Human Immune Virus (HIV)/Acquired Immune Deficiency Syndrome (AIDS), and they are the leading causes of mortality among individuals over the age of thirty and this is according to Bertrand (1999).

The WHO (2011) and Alwan (2010) projected that over the next ten years Africa will experience the largest increase in death rates from CVDs and therefore the negative economic impact of CVDs will be more felt on the continent.

In the first half of the 20th century hypertension was virtually nonexistent in African countries. However in their studies Addo, Smeeth, and Leon (2007) noted that in some African settings more than 40% of adults now have hypertension. Traditionally in Africa, communicable diseases and maternal causes of morbidity and mortality used to be the greatest burden of morbidity and mortality. However according to Lopez et al. (2006) this burden is being taken over by non-communicable diseases and by extension CVDs. For example in Ghana the number of reported new cases of hypertension in outpatient public health facilities increased more than ten-fold from 49,087 in 1988 to 505,180 in 2007. Over the same period, hypertension relative to the total reported outpatient diseases increased from 1.7% to 4.0% in all ages. In most regions of Ghana, hypertension ranks as the fifth commonest cause of outpatient morbidity. However, in the Greater Accra Region of Ghana, hypertension moved from fourth to become second to malaria as the leading cause of outpatient morbidity in 2007, GHS (2008). Studies conducted by the imperial college of London, in the year 2013, indicates that the prevalence of hypertension has increased significantly over the past 2 to 3 decades. According to the study there were approximately 80 million adults with hypertension in sub Saharan Africa in the year 2000 and projection based on current epidemiological data suggests that those will rise to 150 million by 2025. These trends have been strongly linked to individual lifestyle and social lifestyle changes such us excessive drinking of alcohol, excessive use of tobacco, adoption and consumption of diets high in salt and low in fiber. Steyn and Fourie, 1991 did such a study in Sub- Saharan Africa among urban dwellers and also found that intakes of food, especially fat, have risen, and intakes of high-fiber foods have fallen. The mean serum cholesterol level among urban dwellers is significantly higher than those of rural populations living traditionally.

From most of the literature, hypertension has a high prevalence rate in Africa and for that matter Ghana. In his study on the changing patterns of hypertension in four rural communities in Ghana, Amoah (2003) said the prevalence rate in Ghana was 25.4%. Cappucio et al. (2004) in their study of the Ashanti region of Ghana realized a prevalence rate of 28%. According to the WHO African Regional Consultation meeting report on global strategy for diet, physical activity and health, the risk for non-communicable diseases appear to be gaining importance in Ghana, with prevalence of high blood pressure estimated at 30-40% although prevalence data for survey are generally inadequate.

Relative to gender (Addo et al., 2007), did a study and realized that hypertension is usually more pronounced in males than in females and that confirms a WHO survey of 20 African countries between 2003 and 2009 in which males in Seychelles had the highest prevalence. A similar gender based survey conducted by (Burket, 2006) on blood pressure levels in the Volta Region of Ghana reported a prevalence of 32.8% for males and 30.7% for females.

However in countries such as Algeria, Botswana, and Mali, females had higher prevalence than males.

According to research done by the World Health Study (WHS, 2002) more people generally have hypertension in urban areas than in the rural areas of Africa. In Ghana for example, the prevalence of hypertension in urban Accra was estimated to be 28.3% (crude) and 27.3% (age-standardized) and this according to (Amoah, 2003). Interestingly, in countries such as South Africa and Democratic Republic of Congo (DRC), some rural settings have higher prevalence than the urban population in Ethiopia and Tanzania. Even some rural areas in Ghana also have a higher prevalence rate than some urban areas in Ethiopia and Tanzania. This interesting prevalence pattern was attributed to the fact that countries are at different stages of the epidemiological transition. Van de Vijver et al. (2013), did intra-urban studies in the cities of some African countries on hypertension and found that HBP was more prevalent in the slums of those cities than in the elite societies. Such a data dispels the notion that hypertension is a disease of the affluent.

De-Graft et al. (2010) also conducted studies on the levels of awareness, treatment and control of hypertension in most African countries and realized that there are low rates for all these indicators. They found out that even many of those that are aware of their condition are not on treatment. Hendricks et al (2011) confirmed these findings of De-Graft et al (2010), that many of those that are aware of their condition are not on treatment and attributed this situation to high cost of health care. In their study of four rural communities in Ghana, Addo et al, 2006, found that of those with hypertension only 32.3% had prior knowledge of their condition and less than half of these were on treatment (Addo et al 2006). These low levels for all these indicators is crucial since the implication is that there is a large population of people with hypertension that are unaware of their increased risk of hypertension related complications such as heart diseases and stroke, in the coming years. In a 1998 South African study of CVD patients for instance, only 20 percent of the total group understood that hypertension had probably caused their condition (Hale et al. 1998). There could be various possible explanations for the low treatment and poor control of hypertension in Africa. These include scarce resources, lack of patient education, and poor organization of the health care systems. Unaffordable drug prices have been reported to be the major cause of non-compliance with hypertension medication in Ghana. Ghana has a National Health Insurance Scheme (NHIS) that in theory, offers exemptions for some medications and treatment for all service users and cross-cutting medical support for a particularly vulnerable group such as the extremely poor and the aged. But according to Awuah (2009), means-testing for the deserving poor has not been robust and poor communities lack access to the full range of health services, particularly complex services involving long-term chronic disease care.

According to (Whelton et al., 2004), increasing awareness, treatment and control rate of hypertension will have a huge impact on CVD prevention in Africa. People need to be educated about hypertension and its attendant complications. In North America and Europe for instance, reviews have shown an improvement in awareness and this improved awareness has been attributed to rigorous education programs on hypertension after the realization that hypertension was a major player in morbidity and mortality in those continents some decades ago and this according to (McAlister et a., 2011). In these developed countries, the improved control of hypertension has led to considerable reduction in overall morbidity and mortality according to (Gu et al., 2010).

On the average, 50 percent of patients in developed countries do not take their prescribed medicines after one year despite having full access to medicines. In developing countries, this poor adherence is made worse by poor access to health services and drugs, lack of education, and to other factors (Bovet et al., 2002; WHO, 2003a).

According to studies conducted by (Twagirumukiza et al., 2011) in African countries, providing medication is considered an important and cost effective way to the reduction of hypertension but accessibility and cost of treatment, are very often forgotten. Management of hypertension require lifestyle changes as well as the preferred drug treatment. The autors also found in their studies that currently African countries are 80% below the global average of pharmacological spending and 20% below the global average of behavioral risk factors as far as hypertension is concerned. The failure of African countries to adopt the necessary steps early in this twenty first century, to promote healthy lifestyles, however difficult such decisions might be, will inevitably lead to increasing levels of hypertension and CVDs in African countries. This in turn, will be the cause of an unavoidable chronic disease epidemic within the next few decades. The management of these complications is difficult to sustain in sub-Saharan countries where resource-intensive care is not very feasible. Insufficient diagnosis of hypertension and suboptimal blood pressure control in the diagnosed patients increases morbidity and mortality with an increased burden to health care resources. Because of limited resources in most African countries, the best way of controlling the hypertension epidemic lies in the prevention, or at least early detection and adequate control. According to (Beaglehole et al., 2011), the major obstacle to the control of BP is the absence of appropriate services at the primary health care levels of the health services delivery systems. The challenge is to introduce primary and secondary prevention measures now, before the epidemic of CVDs accelerates, particularly as such strategies may be cost-effective than angioplasty and cardiac surgery in the cash-trapped economies of Sub-Saharan Africa. Up to 22 percent of premature all-cause mortality and 45 percent of stroke could be reduced by appropriate detection and treatment (Yusuf et al., 2001). Regular screening programs to detect hypertension in the community may be appropriate, as the disease does not become clinically manifest until it causes end organ damage, such as heart failure, heart attack, stroke, and kidney failure.

2.2 Empirical research with time series methods.

Time series modeling traditionally were developed and used in the area of econometrics. However of recent some works in other areas such as biomedicine, meteorology, health, just to mention a few, employed time series methods for modeling and forecasting.

In the area of health for example, Suleman and Sarpong (2011), used time series to model and to forecast hypertension cases in Navrongo, Ghana. They employed the Box-Jenkins approach to model and forecast hypertension cases in Navrongo. The data they employed spanned from the year 2000 to the year 2010. The data was modelled using the Autoregressive Integrated Moving Average (ARIMA) stochastic model popularized by Box and Jenkins (1976). They concluded that ARMA (3,2) was adequate for modelling and forecasting hypertension in the catchment of Navrongo and their forecasts showed a slight decline in monthly hypertension cases for the year 2011.

Javorka et al. (2005) used the Vector Auto Regression (VAR) theory to study the association between obesity and arterial hypertension in Slovakia. They examined forty obese children and adolescents between the ages of 9.9 years and 18.4 years. They analyzed 2000 heart beats long time series in baroflex intervals with systolic blood pressure (SBP) values starting a stabilization period. They then studied the interactions between the baroflex intervals and the SBP in a closed loop as a result of the simultaneous presence of the feedback baroflex mechanism and the feedforward mechanism operating in opposite directions from the baroflex mechanism and systolic blood pressure. The Granger causality approach offered the possibility to study the effects of both paths in the closed loop VAR system.

Schwandt (2014) in his work captioned "Wealth Shocks and Health Outcomes" which studied the association between wealth status and cardiovascular outcomes in the United States of America, employed time series methods. Using data from the Health and Retirement Study, he used the theory of impulse response functions of VAR processes to construct wealth shocks from the interactions between stock holdings and stock market changes. These constructed wealth shocks were highly predictive of changes in reported wealth and they strongly affected health outcomes. A 10% wealth shock was associated with an improvement of 2-4% of a standard deviation in health. The effects were heterogeneous across cardiovascular health conditions with most pronounced effects for the incidence of high blood pressure, and smaller effects for heart disease problems and stroke. The Granger causality test was also used to investigate the directions between good health and wealth. Their findings which were contrary to findings of similar studies in Russia and South Africa, suggested that wealth affected cardiovascular health over time. The data used in their study was employed from the first 9 waves of Health and Retirement Survey (HRS), covering the years 1992 to 2009. The HRS also contained detailed information on income and wealth holdings.

Cristina et al. (1998) presented a study that describes the use of time series analysis in the evaluation of the incidence of nosocomial infection at the Guadalajara general hospital, Spain. A monthly data comprising time series incidence of nosocomial infections was sourced from a nosocomial infection surveillance system of a primary-care general hospital. The data was analyzed by curve fitting, autoregressive integrated moving average (ARIMA) modeling and intervention and dynamic regression analysis. Their results indicated that, the imposed control and training of personnel by the surveillance system was associated with a 3.63% decrease in accumulated monthly incidence of nosocomial cases. There was also a strong indication that an increase of infection incidence of 4.34% corresponded to a medical strike. An increase of 0.18% was associated with each new nursing contract. Evidence was then obtained from the possible relationship between incidence of nosocomial infection and vacation periods.

Goka (2006) applied time series analysis on the various diseases reported at the Outpatient Departments (OPD) of the Greater Accra region, to forecast 2007 reported diseases and finally, identified the most reported disease(s) within the period. Data was obtained from the Adabraka Polyclinic which is the coordinating center for the Greater Accra regional health statistics. The data covered the period 1996 to 2006. Trend analysis was the main statistical technique used in the study. It was found that malaria constituted half of all cases reported at OPD each year. Malaria, upper respiratory tract infection and skin disease formed an overwhelming majority (about 52.4%) of diseases reported at the OPD each year. Trend analysis of these diseases yielded various forecasted values for 2007.

In a multivariate time series analysis of the cardiovascular system, Runge et al (2014), used the vector autoregressive methods by employing the theory of Granger causality to study the relationship between Heart Rate and High blood Pressure in patients.

Miyake et al. (2009) employed time-series analysis on the seasonal variation in liver function tests. In their study, they examined the seasonal variation in liver function tests using recently described data-mining methods. The latent reference values of aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase (ALP), gamma-glutamyltransferase (gamma GT), cholinesterase (ChE) and total bilirubin (T-Bil) were extracted from a seven-year database of outpatients (aged 20-79 year; comprising approximately 1,270,000 test results). After calculating the monthly means for each variable, the time-series data were separated into trend and seasonal components using a local regression model (Loess method). Then, a cosine function model (cosinor method) was applied to the seasonal component to determine the periodicity and fluctuation range. A two-year outpatient database (215,000 results) from another hospital was also analyzed to confirm the reproducibility of these methods. The authors found that the serum levels of test results tended to increase in the winter. The increase in AST and ALT was about 6% in men and women, and was greater than that in ChE, ALP (in men and women) and gamma GT (in men). In contrast, T-Bil increased by 3.6% (men) and 5.0% (women) in the summer. The total protein and albumin concentrations did not change significantly. AST and ALT showed similar seasonal variation in both institutions in the comparative analysis.

A university-hospital based study by (Adriana et al., 2003), in Bogotá, Colombia, developed and implemented an educational intervention to complement a new structured antibiotic order form. This intervention was performed after assessing the appropriateness of the observed antibiotic prescribing practices using a quasi-experimental study. An application of interrupted time series intervention analysis was conducted in three antibiotic groups (aminoglycosides, cephradine/cephalothin, and ceftazidime/cefotaxime) and their hospital weekly rate of incorrect prescriptions before and after the intervention. A fourth time series was defined on prophylactic antibiotic use in elective surgery.

Pre intervention models were used in the post intervention series to test for pre–post series level differences. An abrupt constant change was significant in the first, third, and fourth time series indicating a 47, 7.3, and 20% reduction of incorrect prescriptions after the intervention. The study concluded that a structured antibiotic order form, coupled with graphic and educational interventions can improve antibiotic use in the university hospital.

Medical experts and health administrators have for several years embarked on vaccination exercises to eradicate various forms of diseases. Girard (2000) used an ARIMA model with intervention analysis technique to analyze and assess the epidemiology situation of whooping cough in England and Wales for the period 1940 to 1990. The ARIMA modeling of this illness contained variables, such as the introduction of widespread vaccination in 1957 and the fall in the level of vaccination down to 30% in 1978. The results of the study confirmed the role of the intervention variables on the evolution of the morbidity due to whooping-cough, by quantifying their impact on the level of the morbidity, as well as the delay needed before they have an influence on the increase of recorded cases of whooping-cough.

2.3 Review of time series methods

2.3.1 Unit Root Tests

Time series modelling require the use of stationary data. On the contrary, most time series data are found to be non-stationary. Fuller (1976) and Dickey and Fuller (1979) advocated tests DickeyFuller (DF) test and Augmented Dickey-Fuller (ADF) test in which a null hypothesis is a nonstationary process with a unit root and an alternative hypothesis is a trend stationary process.

Several methods for testing unit root have been developed. Nelson and Plosser (1982) used the tests developed by Dickey and Fuller to test the economic indicators of the American economy. They found that almost all economic time series such as the Gross National Product have unit root. Next, Phillips and Perron (1988) weakened a strong assumption on the error term and extended the Dickey-Fuller test to a more general test Philips-Perron (PP) test). However, the PP-test did not alter the result of Nelson and Plosser (1982), even using the same data as Nelson and Plosser (1982).

Another important contribution on unit root test was made by Kwiatkowski et al., (1992). They developed a unit root test that reversed the null hypothesis and alternative hypothesis (KPSS test) and verified that only half of the economic time series had unit root using the same data set as Nelson and Plosser's (1982).

In addition, Christiano (1992) criticised Perron's exogenous treatment of a structural change and devised a method with which structural changes with a drift term and a trend can be detected endogenously and proposed a test whose null hypothesis is a unit root process without a structural change and whose opposing hypothesis is a stationary process with a structural change.

Again, another test whose null hypothesis is a unit root process without any change in a drift term and whose alternative hypothesis is trend stationary process with a structural break was proposed by Zivot and Andrews (1992). This proposed test can detect a time point of a structural change endogenously and its asymptotic distribution is constant regardless of the time points of structural changes.

Dickey et al. (1984) following the methodology suggested by Dickey and Fuller (1979) for the zero-frequency unit-root case, proposed the Dickey, Hasza and Fuller (DHF) test to test for seasonal unit root. The DHF test only allows for unit roots at all of the seasonal frequencies and has an alternative hypothesis, which is considered rather restrictive, namely that, all the roots have the same modulus. Trying to overcome these drawbacks Hylleberg et al. (1990) proposed a more general testing (HEGY's test) strategy that allows for unit roots at some (or even all) of the seasonal frequencies as well as the zero frequency. HEGY's methodology allows testing for unit roots at some seasonal frequencies without maintaining that unit roots are present at all seasonal frequencies.

Finally, Banerjee et al. (1992) proposed three kinds of unit root tests. Firstly, a recursive test that is extended on the basis of a structural stability test of Brown et al., (1975) which uses recursive residuals. Secondly, a rolling test that shifts a partial testing period successively among the whole sample period and thirdly a sequential test that conducts *t*-tests or Quandt likelihood ratio tests while shifting a time point of a structural change among the whole sample.

2.3.2 Traditional time series methods

Treating time series in a stochastic sense began in the mid-1920s (Gottman, 1981). Yule (1927) first developed an Autoregressive (AR) model when working on wolfer's sunspot data. Slutzky (1927) first developed a Moving Average (MA) model when studying a white-noise series. Box and Jenkins (1970) developed the Autoregressive Moving average (ARMA) model and gave a full account of the Integrated Autoregressive Moving average (ARIMA) model.

Furthermore, Mann and Wald (1943) proved a theorem to estimate the AR (p) parameters by the least squares method. Quenouille (1947) presented a simple test for AR (p) models and later extended to MA models. Also, Anderson (1971) developed a procedure to estimate the order of the AR model as well as the AR parameter.

In addition, a non-linear least squares technique procedure that led to a technique of approximated likelihood solution for ARMA (p, q) models was developed by Box and Jenkins (1976). Again, an exact likelihood method for estimating parameters of MA (q) models and for ARMA (p, q) models was developed by Newbold (1974). The Box-Pierce statistics was developed by Box and Pierce (1970) and modified by Ljung and Box (1978).

Next, an information criterion to assist in the selection of an ARIMA model was proposed by Akaike (1974). A model with the smallest Akaike Information Criterion (AIC) is the best model to have minimum forecast mean square errors. On the information criterion, Schwarz (1978) indicated that AIC was not consistent when probability approaches one, and proposed a Bayesian Information Criterion (BIC).

Also, Harvey and Phillips (1979) developed an exact likelihood procedure to estimate parameters of an ARIMA model in State-Space form. The State-Space models are also called Structural Time Series (STS) models. Many researchers have pointed out the advantages of the State-Space form over the ARIMA models (Durbin and Koopman, 2001). A time series might be characterised with trend, seasonal cycle and calendar variations, together with the effects of explanatory variables and interventions. These components can be processed separately and for different purposes for a State-Space model. On the contrary, the Box-Jenkins ARIMA model is a black-box model, which solely depends on the data without knowledge of the system structure that produces the data. The second advantage is the recursive nature of the State-Space model which obviously allows changes in the system overtime, while ARIMA models are homogenous throughout time, based on the stationary assumption.

Moreover, Granger and Joyeux (1980) and Hosking (1981) adopted an Autoregressive Fractionally Integrated Moving average (ARFIMA) model to study a long memory time series. The autocorrelation function in an ARFIMA (*p, d, q*) model decays at a hyperbolic rate for non-zero *d* which is slower than the usual geometric rate of a stationary ARMA (*p, q*) model. Another important contribution in the area of time series analysis was made by Engle (1982) when he introduced the Autoregressive Conditional Heteroscedasticity (ARCH) model, to model changing volatility. The non-linear term is the variance of the disturbance. An extension of the ARCH model to the Generalised Autoregressive Conditional Heteroscedasticity (GARCH) model was made by Bollerslev (1986). Again, Weiss (1984) proposed an ARMA-ARCH model, in which an ARMA model is used to model mean behavior and an ARCH model to model the residuals of the ARMA model. The quasi-maximum-likelihood method is used to estimate model parameters.

Furthermore, another remarkable contribution was made by Hillmer and Tiao (1979) in the area of multivariate time series. The authors developed an exact likelihood technique for Vector Autoregressive Moving average (VARMA) model. Harvey and Peters (1984) further proposed a state-space method to estimate parameters of VARMA (*p, q*) models.

In addition, a cointegrated multivariate time series and Error Correction Models (ECM) were proposed by Engle and Granger (1987). The cointegrated concept captures the phenomenon of univariate non-stationary time series moving together. The ECM procedure involves two steps. The first step involves modelling the long term relationship between endogenous and exogenous variables. The variables involved have to comply with two constraints; non-stationary and being stationary after first differencing. In the second step, the dynamic short term process is modelled and only stationary variables enter the regression equation.

Also, a Vector Autoregressive Fractionally Integrated Moving average (VARFIMA) was developed by Robinson and Yajima (2002) to handle multivariate time series cointegration problems.

2.4 Summary

This chapter dealt with reviewing of literature that is relevant to the study. Reviewing of the literature has exposed us to the diverse techniques that researchers have employed in modelling time series data. However, among the diverse techniques reviewed the vector autoregressive model was the one employed for this study.

CHAPTER THREE

RESEARCH METHODOLOGY

3.0 Introduction

The chapter is organized as follows: Section 3.1 presents the basic concepts and definitions of time series, and section 3.2 presents the components of time series. Section 3.3 presents the unit root test of variables to be used in the model. Section 3.4 presents the empirical strategy, section 3.5 discusses the model diagnostics while section 3.6 introduces the structural analysis.

3.1 Basic Concepts and Definitions of Time Series

3.1.1 Time Series Analysis

Time series is defined as a collection of observations or measurements on quantitative variables made sequentially or in a uniform set of time period, usually daily, weekly, monthly, quarterly, annually, and so on and so forth. Examples include total monthly crime for a jurisdiction for a period of ten years, daily stock prices of a firm for a period of one year, monthly electricity consumption for a household for a period of five years, etc. Time series analysis comprises methods or a process that break down a series into components and explainable portions that allows trends to be identified, estimates and forecasts to be made.

A sequence of random variables $\{Y_t: t = 0, \pm 1, \pm 2, \pm 3, \ldots\}$ is called a stochastic process and serves as a model for an observed time series.

Basically a time series is either an Autoregressive (AR) process, a Moving Average (MA) process or both i.e. Autoregressive Moving Average (ARMA).

We will let $\{Y_t\}$ denote any observed time series and $\{e_t\}$ represent an observed white noise series (i.e. a sequence of identically distributed, zero-mean, independent random variables).

In general a linear process, $\{Y_t\}$, is one that can be represented as a weighted linear combination of present and past white noise terms as

$$Y_t = e_t + \Psi_1 e_{t-1} + \Psi_2 e_{t-2} + \cdots \quad (3.1)$$

A Moving Average process results in the case when only a finite number of the Ψ- weights are non-zero. In such a case equation (3.1) can be re-written as:

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q} \quad 3.2$$

Such a series as in equation (3.2) is what is called a moving average process of order q. The terminology moving average arises from the fact that Y_t is obtained by applying the weights 1, $-\theta_1$, $-\theta_2$,….., $-\theta_q$ to the variables e_t, e_{t-1}, e_{t-2}, ... , e_{t-q} and then moving the weights and applying them to e_{t+1}, e_t, e_{t-1}, ... , e_{t+q-1} to obtain Y_{t+1} and so on.

On the other hand univariate autoregressive processes are as their name suggest- regression on themselves. Specifically a p-th order AR process $\{Y_t\}$ satisfies the equation,

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t \quad (3.3)$$

The explanation of equation (3.3) above is that the current value $\{Y_t\}$ is a linear combination of the p most recent past values of itself plus an “innovation” term e_t that incorporates everything new in the series at time t that is not explained by past values.

Basically time series analysis attempts to understand the underlying context of data points through the use of a model to forecast future values based on known past values. Such time series models include GARCH, TARCH, EGARCH, FIGARCH, CGARCH, ARIMA, VAR, COINTEGRATION, etc but the main focus of this study is based on the VAR model. Simply replacing scalars with matrices and scalar operations with matrix operations changes equation

(3.3) from a univariate autoregressive model to a vector autoregressive model which is the multivariate autoregressive model as seen in equation (3.13).

3.1.2 Lag

Lag is the time period between two observations. For example, lag 1 is between Y_t, and Y_{t-1}. Lag 2 is between Y_t and Y_{t-2}. Time series can also be lagged forward, Y_t and Y_{t+1}. The observation at the current time, Y_t, depends on the previous observation, Y_{t-1}.

3.1.3 Differencing

Differencing simply means subtracting the value of an earlier observation from the value of a later observation. Calculating differences among pairs of observations at some lag is just to make a non-stationary series stationary. The number of times a series is differenced before stationarity is obtained determines their order of integration.

3.1.4 Stationary and Non-stationary Series

A time series is said to be stationary if its statistical properties are invariant over time. This implies that the mean and the variance are the same for all times. Otherwise the series is non stationary. Using a non-stationary time series produces unreliable and spurious results and that leads to poor understanding and forecasting.

3.1.5 Time Series Graph

Time series plot is simply a graph which displays observations on the y-axis against equally spaced time intervals on the x-axis. The time series plot specifically consists of:

Time scale (index, calendar, clock, or stamp column) on the x-axis; data scale on the y-axis; and lines displaying each time series as shown in the Figure 3.1 below for a given hypothetical data. The plots are usually used to: detect trends in the data over time; detect seasonality in the data; and compare trends across groups.

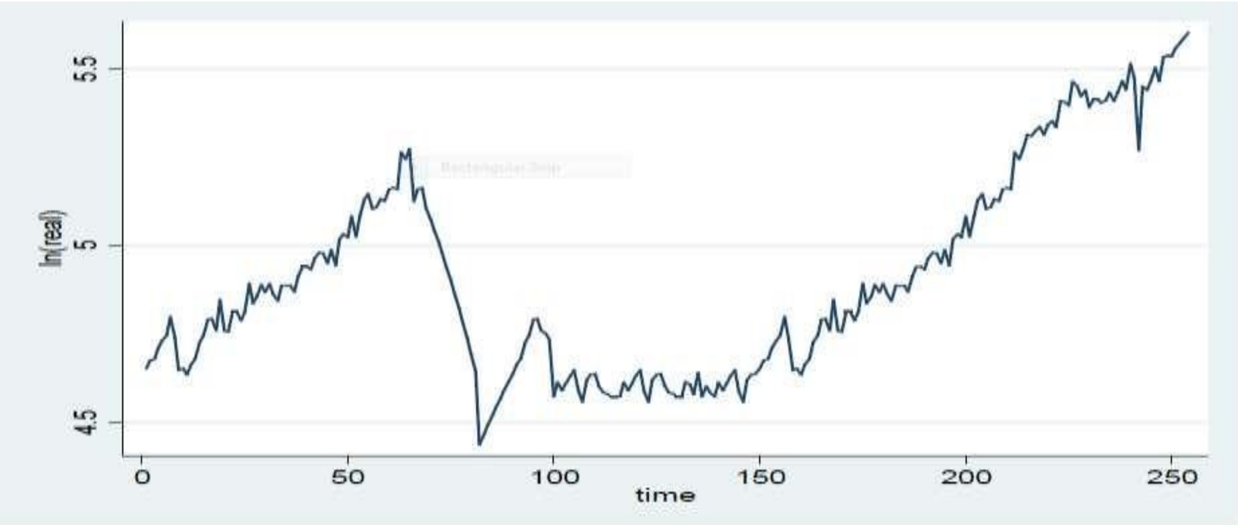

Figure 3.1: Time plot for a hypothetical data of 254 observations

3.2. Components of Time Series

A vital step in choosing appropriate modeling and forecasting procedure is to consider the type of data patterns exhibited from the time series graphs of the time plots. The sources of variation in terms of patterns in time series data are mostly classified into four main components. These components include seasonal variation; trend variation; cyclic changes; and "irregular" fluctuations.

3.2.1 Trend

The trend is simply the underlying long term behavior or pattern of the data or series. The Australian Bureau of Statistics (ABS, 2008) defined trend as the 'long term' movement in a time series without calendar related and irregular effects, and is a reflection of the underlying level. It may be as result of influences such as population growth, price inflation and general economic changes etc. The following graph depicts a series in which there is an obvious upward trend over time:

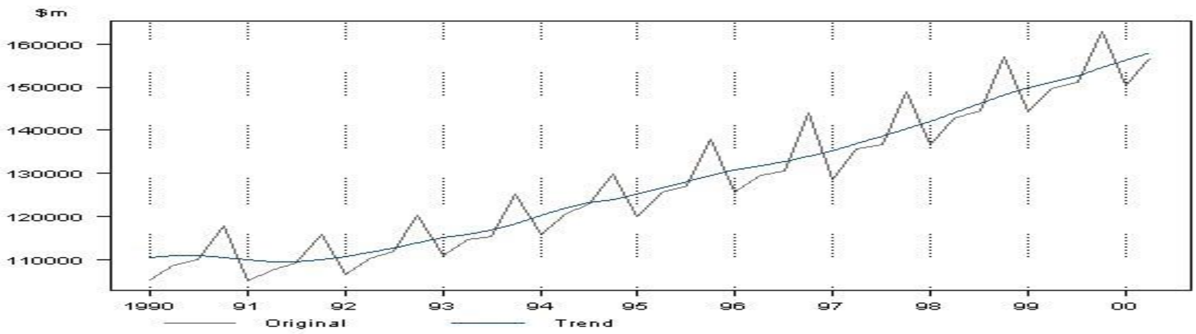

Figure 3.2: Upward trend graph of a hypothetical time series data

3.2.2 Seasonal variation

A seasonal effect is a systematic and calendar related effect. Some examples include the sharp escalation in most Retail series which occurs around December in response to the Christmas period, or an increase in water consumption in summer due to warmer weather. Other seasonal effects include trading day effects (the number of working or trading days in a given month differs from year to year which will impact upon the level of activity in that month) and moving holidays (the timing of holidays such as Easter varies, so the effects of the holiday will be experienced in different periods each year). Seasonal adjustment is the process of estimating and then removing from a time series influences that are systematic and calendar related. Observed data needs to be seasonally adjusted as seasonal effects can conceal both the true underlying movement in the series, as well as certain non-seasonal characteristics which may be of interest to analysts. Seasonality in a time series can be identified by regularly spaced peaks and troughs which have a consistent direction and approximately the same magnitude every year, relative to the trend as depicted in figure 3.3 .

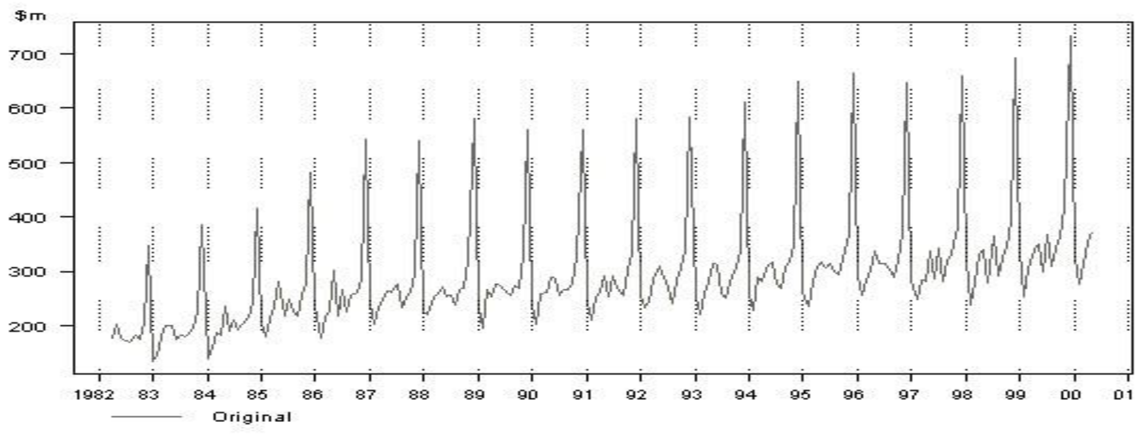

Figure 3.3: Graphical display of seasonal effect of a hypothetical data

Other techniques that can be used in time series analysis to detect seasonality include:

(i) A seasonal subseries plot which is a specialized technique for showing seasonality.

(ii) Multiple box plots can be used as an alternative to the seasonal subseries plot to detect seasonality.

(iii) The autocorrelation plot can help identify seasonality.

3.2.3 Cyclical variations

Cyclical variations are the short term fluctuations (rises and falls) that exist in the data that are not of a fixed period. They are usually due to unexpected or unpredictable events such as those associated with a business cycle's sharp rise in inflation or stock price, etc. The main difference between the seasonal and cyclical variation is the fact that the former is of a constant length and recurs at regular intervals, while the latter varies in length. More so, the length of a cycle is averagely longer than that of seasonality with the magnitude of a cycle usually being more variable than that of seasonal variation.

3.2.4 Irregular Variations

The irregular component (sometimes also known as the residual) is what remains after the seasonal and trend components of a time series have been estimated and removed. It results from short term fluctuations in the series which are neither systematic nor predictable. In a highly irregular series, these fluctuations can dominate movements, which will mask the trend and seasonality. Figure 3.4 is a graph which is of a highly irregular hypothetical time series.

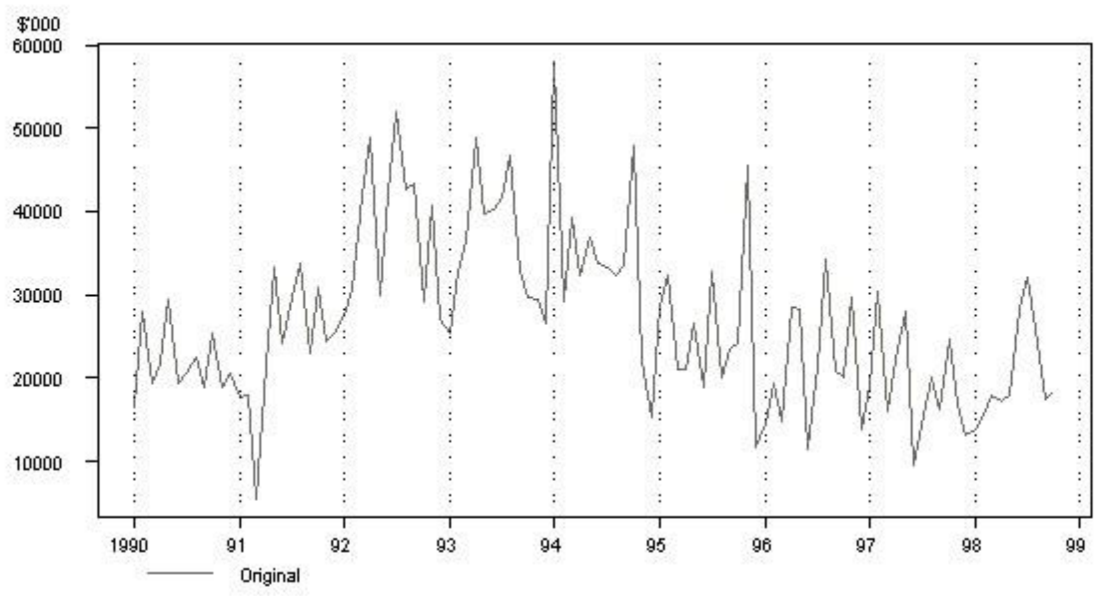

Figure 3.4: Typical irregular effect graph of a hypothetical time series data

3.3 Unit Roots

A unit root in a time series arises when either the AR (p) or the MA (q) polynomial of an ARMA (p,q) model has a root on or near the unit circle. But as stated already a VAR (p) process is the multivariate form of the univariate AR (p) process. And so specifically an AR (p) time series $\{y_t\}$ is said to be stationary if the roots of its associated polynomial

$$m^p - \emptyset_1 m^{p-1} - \emptyset_2 m^{p-2} - \cdots - \emptyset_p = 0 \tag{3.4}$$

are less than one in absolute value. But if the roots are equal to one or almost one in absolute value then the time series is not stationary and such an AR(p) series is said to have a unit root. By extension a VAR(p) process is said to have a unit root if its AR (p) process has a unit root. In such a case when a time series has unit roots we can make it stationary by differencing it. But on the contrary when an MA (q) polynomial has a root near one, then it means it has been over differenced.

3.3.1 Unit Root Test

A very important aspect of time series analysis is to ensure that the series is stationary. A stationary time series is one who's first and second moments are invariant of time. That is, the expected value of the time series does not depend on time and the auto-covariance function, $cov(y_t, y_{t+k})$ for any lag k, is only a function of k and not time, that is $\gamma_y(k) = cov(y_t, y_{t+k})$. Many methods have been proposed for testing stationarity of a time series data. These include both graphical and quantitative methods. The graphical approach includes observing the Autocorrelation function (ACF) plots. A strong and slowly decaying ACF will suggest deviation from stationarity. For the purpose of this study, in addition to the ACF, three quantitative techniques for testing unit root are employed. These are; the Augmented Dickey-Fuller (ADF) test, Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test and the Phillip-Perron (PP) unit root test.

3.3.2 Augmented Dickey-Fuller (ADF) Test

The ADF test proposed by Dickey and Fuller (1979) was an improvement on the Dickey-Fuller (DF) test. The test is based on the assumption that the series follows a random walk. Consider an autoregressive process of order one, AR (1), below

$$Y_t = \phi Y_{t-1} + \varepsilon_t \quad (3.5)$$

where ε_t denotes a serially uncorrelated white noise sequence with a mean of zero and constant variance. If ϕ=1, the equation (3.5) becomes a random walk model without drift, which is known as a non-stationary process.

The basic concept of the ADF test in this case is to simply regress Y_t on its lagged value Y_{t-1} and find out if the estimated ϕ is statistically equal to one or not. Equation (3.5) can be manipulated by subtracting Y_{t-1} from both sides to obtain

$$\Delta Y_t = \delta Y_{t-1} + \varepsilon_t \quad (3.6)$$

where $\delta = \phi - 1$, and $\Delta Y_t = Y_t - Y_{t-1}$.

In practice instead of estimation equation (3.5), we rather estimate equation (3.6) and test for the null hypothesis of $\delta = 0$ against the alternative $\delta \neq 0$. If $\delta = 0$, then $\phi = 1$, meaning that the series hav a unit root and hence not stationary. Under the null hypothesis $\delta = 0$, the *t*-value of the estimated coefficient of Y_{t-1} does not have an asymptotic normal distribution (Erdogdu, 2007).

The decision to reject the null hypothesis or not is based on the DF critical values of the τ-statistic. The DF test is based on the assumption that the error terms are uncorrelated. In such a situation the regression equation is presented as:

$$Y_t = \delta Y_{t-1} + \sum_{i=1}^{p} \gamma_i Y_{t-i} + \varepsilon_t \quad (3.7)$$

where $\sum_{i=1}^{p} \gamma_i Y_{t-i}$ is the sum of the lagged values of the dependent variable Y_t and p is the order of the autoregressive process. The parameter of interest in the ADF test is δ. For δ=0, the series contains unit root and hence non-stationary. The choice of the starting augmentation order depends on; data periodicity, significance of γ_i estimates, and white noise residuals. After preliminary estimation, non-significant parameter augmentation can be dropped in order to have more efficient estimates.

The test statistic for the ADF test is given by

$$F_{\tau} = \frac{\hat{\delta}}{SE\ (\hat{\delta})} \tag{3.8}$$

where $SE(\delta)$ is the standard error of the least square estimate of (δ).The null hypothesis that "the series is not stationary" is rejected if the test statistic is greater than the critical value in which case the associated p-value will be greater than the conventional 5% significance level.

3.3.3 Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test

Another complementary test for investigating the order of integration of a series Y_t is to test the null hypothesis that the data generating process is stationary $(H_0\text{: } Y_t \sim (\mathbf{0}))$ against the alternative that it is non-stationary $(H_1\text{: } Y_t \sim (\mathbf{1}))$ using the Kwiatkowski-Phillips-Schmidt-Shin procedure. Kwiatkowski et al., (1992) derived a test for this pair of hypotheses. The test assumes that if there is no linear trend term, the point of departure is a data generating process of the form

$$Y_t = X_t + \varepsilon_t \tag{3.9}$$

where is a random walk, $X_t = X_{t-1} + vt$, $vt \sim$ iid $(0,\sigma_v^2)$, and ε_t is a white noise sequence.

In this context, the foregoing pair of hypotheses is equivalent to the pair;

$$H_0 : \sigma_v^2 = 0 \text{ and}$$

$$H_1 : \sigma_v^2 = 0$$

If H_0 holds then Y_t is composed of a constant and a stationary process ε_t ; thus, Y_t is also stationary. Kwiatkowski et al, (1992) proposed the following test statistic:

$$KPSS = \frac{1}{T^2} \sum_{t=1}^{T} \frac{S_t^2}{\hat{\sigma}_\infty^2} \tag{3.10}$$

where T is the number of observations, $S_t = \sum_{j=1}^{t} \hat{\omega}_j$ with $\hat{\omega}_j = Y_t - \bar{Y}$ and $\hat{\sigma}_{\infty}^2$ is an estimator of the equation that is here below:

$$\sigma_{\infty}^2 = \lim_{T\to\infty} T^{-1} Var\left(\sum_{t=1}^{T} \varepsilon_t\right) \tag{3.11}$$

That is, $\hat{\sigma}_{\infty}^2$ is an estimator of the long-run variance of the process ε_t. If Y_t is a stationary process, then S_t is integrated of order one (**I (1)**) and the quantity in the denominator of the KPSS statistic is an estimator of its variance, which has a stochastic limit. The term in the denominator ensures that overall; the limiting distribution is free of unknown nuisance parameters. If, however, Y_t is integrated of order one (**I (1)**), the numerator will grow without bounds, causing the statistic to become large for large sample sizes. The null hypothesis of stationarity is rejected for large values of KPSS.

3.3.4 The Phillip-Perron Test

The Phillip Perron (PP) test was developed by Phillips (1987) and Phillips and Perron (1988). The PP tests are based on the following ADF regression, and the critical values are the same as those used for the ADF tests:

$$Y_t = \delta Y_{t-1} + \sum_{i=1}^{p} \gamma_i Y_{t-i} + \varepsilon_t \tag{3.12}$$

where ε_t is the error term.

The PP unit root test is utilized in this case in preference to ADF unit root tests for the following reasons. First the PP tests does not require an assumption of homoscedasticity of the error term (Phillips, 1987). Secondly, since lagged terms for the variable of interest are set to zero there is no loss of effective observations from the series (Perron, 1988), which is especially useful if the number of data points is limited.

The PP unit root test corrects the serial correlation and autoregressive heteroscedasticity of the error terms. This aims at providing unit root tests results that are robust to serial correlation and time dependent heteroscedasticity of errors.

In both the PP and ADF unit root tests the null hypothesis is that “the series is non-stationary” and this is either accepted or rejected by examination of the t-ratio of the lagged term Y_{t-1} compared with the tabulated values. If the t-ratio is greater than the critical value (in which case the p-value is less than the conventional 5% level of significance) the null hypothesis of a unit root (i.e. that the series is non-stationary) is rejected and the series is considered to be integrated of order zero, i.e. they are considered to be stationary at levels.

3.4 Empirical Strategy

The specific modeling technique used in this study is the Vector Autoregressive (VAR) modelling approach.

The use of the VAR model for time series was proposed by Sims (1980), motivated in part by questions related to the validity of the way in which a theory is used to provide a prior justification for the inclusion of a restricted subset of variables in the "structural" specification of each dependent variable.

The VAR approach has become somewhat standard in time series modeling because compared to the structural approach; it avoids the need to provide a dynamic theory specifying the relationships among the jointly determined variables; it can handle endogenous variables on both sides of the equation; it is capable of mixing of I(1) and I(0) variables in one system. In a VAR system, each variable is regressed on its own lags plus the lags of the other variables.

The appropriate lag length (p), which should be specified long enough for the residuals not to be serially correlated, can be determined using standard model selection criteria such as the Akaike Information Criterion (AIC), Schwarz information criterion (SIC), Hannan-Quinn information criterion (HQ) and the Final Prediction Error (FPE) information criteria.

3.4.1 The Vector Autoregressive model (VAR)

Assume Y_t is a k by 1 vector stochastic process. A p^{th} order VAR model of inter-related time series, written as VAR (p), is a process that evolves according to

$$\boldsymbol{Y_t} = \boldsymbol{\Phi_0} + \boldsymbol{\Phi_1 Y_{t-1}} + \boldsymbol{\Phi_2 Y_{t-2}} + \cdots + \boldsymbol{\Phi_p Y_{t-p}} + \boldsymbol{\epsilon_t} \tag{3.13}$$

where $\boldsymbol{\Phi_0}$ is a k by 1 vector of intercept parameters, $\boldsymbol{\Phi_j}$ is a k by k parameter matrices, with $j = 1,2, \ldots.., p$ and $\boldsymbol{\epsilon_t}$ is a vector white noise.

A vector white noise process has the same useful properties as a univariate white noise process; it is mean zero, has finite covariance and is uncorrelated with its past although the elements of a vector white noise process are not required to be contemporaneously uncorrelated.

3.4.2 Specification of the VAR model

A VAR model of order p, is specified as VAR (p) where p is also the lag length (i.e. the number of lags to be considered in the VAR system). Therefore a critical element in the specification of a VAR model is the determination of its lag length. The appropriate lag length (p) should be specified long enough for the residuals not to be serially correlated.

Various lag length selection criteria are defined by different authors like, Akaike's (1969) final prediction error (FPE), Akaike Information Criterion (AIC) suggested by Akaike (1974), Schwarz Criterion (SC) (1978) and Hannan-Quinn Information Criterion (HQ) (1979) just to mention a few.

Ivanov and Kilian (2005) recommends the AIC for the selection of lag length when the data is a monthly data. But the lag length for the VAR (p) model may be determined using any of the above mentioned model selection criteria.

The general approach to VAR model estimation is to fit the VAR (p) with the order p = 0,..., p_{max} and the value of p should minimize some model selection criteria. Model selection criteria for VAR (p) models have the form

$$IC(p) = ln|\bar{\boldsymbol{\Sigma}}(p)| + c_T.\varphi(n,p) \tag{3.14}$$

where $\bar{\Sigma}(p) = T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_t'$ is the residual covariance matrix without a degree of freedom correction from the VAR (p) model, c_T is a sequence indexed by the sample size T, and $\varphi(n, p)$ is a penalty function which penalizes large VAR (p) models.

The lag order is chosen by optimally balancing the term $ln|\bar{\Sigma}(p)|$ which is a non-increasing function of order n and $\varphi(n, p)$ which increases with n.

The three most common information criteria are the Akaike (AIC), Schwarz-Bayesian (BIC) and Hannan-Quinn (HQ):

$$AIC(p) = ln|\bar{\Sigma}(p)| + \frac{2}{T}pn^2 \tag{3.15}$$

$$BIC(p) = ln|\bar{\Sigma}(p)| + \frac{ln[T]}{T}pn^2 \tag{3.16}$$

$$HQ(p) = ln|\bar{\Sigma}(p)| + \frac{2\ln[lnT]}{T}pn^2 \tag{3.17}$$

3.5 Model Diagnostics

Once a VAR-model has been estimated, it is of pivotal interest to see whether the residuals obey the model's assumptions. That is, one should check for the absence of serial correlation and heteroscedasticity and also see if the error process is normally distributed. As a final check, one can conduct structural stability tests; i.e., CUSUM, CUSUM-of-squares, and/or fluctuation tests. For testing the lack of serial correlation in the residuals of a VAR (p)-model, the LM test proposed by Breusch [1978] and Godfrey [1978] are applied.

3.5.1 The Breush-Godfrey LM Test

The Breusch-Godfrey LM methodology for testing residual serial auto correlation assumes a VAR model for the error vector $\varepsilon_t = D_1\varepsilon_{t-1} + \cdots + D_h\varepsilon_{t-h} + v_t$ where v_t is white noise. ε_t is equal to v_t if there is no residual serial correlation. Therefore we test for the hypotheses

H_0: $D_1 = D_2 = \cdots = D_h = 0$ against

H_1: $D_j \neq 0$ for $j = 1, 2, \ldots, h$

For any K dimensional VAR, the Breusch-Godfrey LM test statistic is given by

$$LM_h = T\left(K - tr\left(\hat{\Sigma}_R^{-1}\hat{\Sigma}_e\right)\right) \tag{3.18}$$

where $\hat{\Sigma}_R$ and $\hat{\Sigma}_e$ assign the residual covariance matrix of the restricted and unrestricted model respectively, T is the sample size. The test statistic LM_h has a chi-squared distribution of $\chi^2(hk^2)$.

3.5.2 Test For Heteroscedasticity

The multivariate ARCH-LM test was used to test for heteroscedasticity. The multivariate ARCH-LM test is based on the regression:

$$vech(\hat{u}_t\hat{u}_t^T) = \beta_0 + B_1 vech(\hat{u}_{t-1}\hat{u}_{t-1}^T) + \cdots + B_q vech\left(\hat{u}_{t-q}\hat{u}_{t-q}^T\right) + v_t \quad (3.19)$$

where by v_t assigns a spherical error process and $vech$ is the column-stacking operator for symmetric matrices that stacks the columns from the main diagonal on downward. The dimension of β_0 is ½$[K(K+1)]$. The dimension for the coefficient matrices B_i with $i= 1, \dots, q$ is ½$[K(K+1)]$ by ½$[K(K+1)]$

The null hypothesis for this test: There is no heteroscedasticity.

ie H_0: $B_1 = B_2 = \cdots = B_q = 0$

The corresponding alternative hypothesis is: There is heteroscedasticity.

ie H_1: $B_1 \neq 0 \cap B_2 \neq 0 \cap \dots \cap B_q \neq 0$

The test statistic is:

$$VARCH_{LM}(q) = \frac{1}{2}TK(K+1)R_m^2 \quad (3.20)$$

where $R_m^2 = 1 - \frac{2}{K(K+1)} tr(\widehat{\Omega}\widehat{\Omega}_0^{-1})$

$\hat{\Omega}$ assigns the covariance matrix of the above defined regression model.

The test statistic has a chi-squared distribution of $\chi^2(\frac{qK^2(K+1)^2}{4})$

3.6 Structural Analysis

The dynamic properties of a VAR (p) are often summarized using various types of structural analysis. The three main types of structural analysis summaries are the Granger causality test, the impulse response functions, and the forecast error variance decomposition.

3.6.1 Granger Causality

One of the main uses of VAR models is forecasting. The structure of the VAR model provides information about a variable's or a group of variables' forecasting ability for other variables. The following intuitive notion of a variable's forecasting ability is due to Granger (1969).

Granger Causality is the standard method to determine whether one variable is useful in predicting another and evidence of Granger causality is a good indicator that a VAR, rather than a univariate model, is needed. A scalar random variable $\{x_t\}$ is said to not Granger cause $\{y_t\}$ if

$$E[y_t|x_{t-1}, y_{t-1}, x_{t-2}, y_{t-2}, \dots] = E[y_t|y_{t-1}, y_{t-2}, \dots]$$

That is, $\{x_t\}$ does not Granger cause $\{y_t\}$ if the forecast of y_t is the same whether conditioned on past values of x_t or not. In such a case we say $\{x_t\}$ does not Granger cause $\{y_t\}$, or $\{x_t\}$ is exogenous to $\{y_t\}$.

Granger causality is not the same as causality in the philosophical sense. Granger causality does not claim that x is the reason for y in the sense like, for example, y moves because x moves. It just says that x is helpful in forecasting y, which might happen for other reasons than direct causality. A bivariate VAR (2) can generally be illustrated as:

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \phi_{11,1} & \phi_{12,1} \\ \phi_{21,1} & \phi_{22,1} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \phi_{11,2} & \phi_{12,2} \\ \phi_{21,2} & \phi_{22,2} \end{bmatrix} \begin{bmatrix} x_{t-2} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

In general if $\phi_{21,1} = \phi_{21,2} = 0$, then $\{x_t\}$ does not Granger cause $\{y_t\}$. If it happens to be the case that $\{x_t\}$ does not Granger cause $\{y_t\}$ and $\epsilon_{1,t}$ and $\epsilon_{2,t}$ have no contemporaneous correlation, then y_t is said to be weakly exogenous, and y_t can be modeled completely independently of x_t.

In a more general form, for any VAR (p) specification

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \cdots + \Phi_p y_{t-p} + \epsilon_t$$

$\{y_{j,}\}$ does not Granger cause $\{y_{i,}\}$ if $\phi_{ij,} = \phi_{ij,} = \dots = \phi_{ij,} = 0$

Granger causality can be tested in both directions and so depending on the objective of the study it can also be tested to see if y_t Granger causes x_t. This test can be conducted irrespective of whether x_t Granger causes y_t or not.

3.6.1.1 Granger Causality Test

Testing for Granger causality was conducted using Granger's OLS method. When using the OLS method to test for causality between two variables for example X and Y, to see whether past values of X are useful for predicting Y once Y's history has been modeled, the test is implemented by regressing Y on p past values of X, where p is the number of lags estimated by the various lag selection criteria. An F-test is then used to determine whether the coefficients of the past values of X are jointly zero. The null hypothesis for such a test is X does not Granger cause Y.

Now, considering a bivariate VAR (p) model written out in scalar form as:

$$x_t = \phi_1 + \sum_{i=1}^{p} \phi_{11}^{(i)} x_{i,t-i} + \sum_{i=1}^{p} \phi_{12}^{(i)} y_{i,,t-i} + \epsilon_{1t} \tag{3.22}$$

$$y_t = \phi_2 + \sum_{i=1}^{p} \phi_{21}^{(i)} x_{i,t-i} + \sum_{i=1}^{p} \phi_{22}^{(i)} y_{i,t-i} + \epsilon_{2t} \tag{3.23}$$

Then test for Granger causality from x to y is an F-test for the joint significance of

$\phi_{21}^{(i)},\ldots\ldots\ldots,\phi_{21}^{(\mathrm{p})}$ for the OLS regression in equation (3.23).

Similarly, the test for Granger causality from y to x is an F-test for the joint significance of

$\phi_{12}^{(i)},\ldots\ldots\ldots,\phi_{12}^{(\mathrm{p})}$ for the OLS regression in equation (3.22).

The F-statistic for testing the hypothesis

$$H_0: \beta_{k-p+1} = \beta_{k-2+p} = \cdots = \beta_k = 0 \text{ against}$$

$$H_1: \text{some } \beta_{k-i} \neq 0,\ i = 0, 1, \ldots, q-1$$

in any regression model $y = \beta_0 + \beta_{1x_1} + \cdots + \beta_k x_k + \mu$ (3.24)

is $F = \frac{(SSE_r - SSE_{ur})/q}{SSE_{ur}/(n-k-1)}$ (3.25)

where SSE_r is the residual sum of squares from the model under H_0 and SSE_u is the residual sum of squares for the model in equation (3.24).

Under the null hypothesis the test statistic in equation (3.25) is F-distributed with

$q = df_r - df_{ur}$ and $n - k - 1$ degrees of freedom, where df_r is the degrees of freedom of SSE_r and df_{ur} is the degrees of freedom of SSE_{ur}.

In the case of considering Granger causality in a bivariate VAR (p) model we drop $q = p$ variables in a model with $n = T$ observations and $k = 2p$ variables beyond the constant.

Hence,

$$F = \frac{(SSE_r - SSE_{ur})/p}{SSE_{ur}/(T-2p-1)} \sim F(p, T - 2p - 1) \dots\dots\dots\dots\dots\dots\dots. (3.26)$$

under H_0 : y does not Granger cause x in equation (3.22).

Again under H_0 : x does not Granger cause y in equation (3.23).

3.6.2 Impulse Response Functions

In the univariate case, the Auto Correlation Function (ACF) is sufficient to understand how shocks decay. However when analyzing a vector data, this is no longer the case. A shock to one series has an immediate effect on themselves and other variables in the system which, in turn, can feed back into the original variable. The impulse response function technique is used to study how one variable responds to a sudden change in other variables future time horizon. These sudden changes may be referred to as shocks or innovations or the unexpected changes in a variable (Harvey, 1994). This innovation is commonly referred to as impulse, to reflect primarily the notion of one time shock occurring at some point in the time line.

The impulse response function of any y_i which is an element of y, with respect to a shock in ϵ_j , an element of ϵ, for any j and i , is the change in $\boldsymbol{y_{it+s}}$, for any s ≥ 0, corresponding to a unit shock in $\boldsymbol{\epsilon_{j,t}}$

The impulse response function can be clearly illustrated through a vector moving average

(VMA). As long as $\boldsymbol{Y_t}$ is covariance stationary it must have a VMA representation of the form,

$$\boldsymbol{y_t} = \boldsymbol{\mu} + \boldsymbol{\epsilon_t} + \Xi_1\boldsymbol{\epsilon_{t-1}} + \Xi_2\boldsymbol{\epsilon_{t-2}} + \cdots$$

Using this VMA, the impulse response $\boldsymbol{y_i}$ with respect to a shock in $\boldsymbol{\epsilon_j}$ is simply $\{\mathbf{1}, \Xi_{1[ii]}, \Xi_{2[ii]}, \Xi_{2[ii]}, \ldots\}$ if $i = j$ and $\{\mathbf{0}, \Xi_{1[ij]}, \Xi_{2[ij]}, \Xi_{3[ij]}, \ldots\}$ otherwise.

Confidence bands can be constructed to determine whether an impulse response is large in a statistically meaningful sense. Since the parameters of the VAR are asymptotically normal (as long as it is stationary and the innovations are white noise), the impulse responses will also be asymptotically normal.

In this study the bootstrap method was used to compute confidence bands for the impulse response function.

The bootstrap method operates based on the principle that "if the residuals are realizations of the actual error process, one can use them directly to simulate this distribution rather than making an arbitrary assumption about the error distribution (e.g. i.i.d. normal)".

The procedure:

(i) Compute $\hat{\boldsymbol{\Phi}}$ from the initial data and estimate the residuals $\hat{\epsilon}_t$

(ii) Using $\hat{\boldsymbol{\epsilon}}$, compute a new series of residuals $\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}$ by sampling, with replacement, from the original residuals. The new series of residuals can be described as

$$\{\hat{\boldsymbol{\epsilon}}_{\boldsymbol{u1}}, \hat{\boldsymbol{\epsilon}}_{\boldsymbol{u2}}, \ldots, \hat{\boldsymbol{\epsilon}}_{\boldsymbol{uT}}\}$$

where $\boldsymbol{u_i}$ are i.i.d. discrete uniform random variables taking the values 1, 2, . . . , T . In essence, the new set of residuals is just the old set of residuals reordered with some duplication and omission.

(iii) Using $\hat{\boldsymbol{\Phi}}$ and $\{\hat{\boldsymbol{\epsilon}}_{\boldsymbol{u1}}, \hat{\boldsymbol{\epsilon}}_{\boldsymbol{u2}}, \ldots, \hat{\boldsymbol{\epsilon}}_{\boldsymbol{T}}\}$, simulate a time-series $\{\tilde{\boldsymbol{y}}_{\boldsymbol{t}}\}$ with as many observations as the original data. These can be computed directly using the VAR

$$\tilde{y}_t = \hat{\Phi}_0 + \hat{\Phi}_1 y_{t-1} + \cdots + \hat{\Phi}_{t-P} + \hat{\epsilon}_{u_t}$$

(iv) Using $\hat{y}_t$ compute estimates of $\hat{\boldsymbol{\Phi}}_{\boldsymbol{b}}$ from a VAR.

(v) Using $\breve{\Phi}_b$ compute the impulse responses $\{\Xi_j,\}$ where b = 1, 2, ….,B. These values are saved.

(vi) Return to step 2 and compute a total of B impulse responses. Typically B is between 100 and 1000.

(vii) For each impulse response for each horizon, sort the impulse responses. The 5th and 95th percentiles of this distribution are the confidence intervals.

3.6.3 Forecast Error Variance Decomposition

Variance decomposition refers to the breakdown of the forecast error variance into components due to shocks in the series. Basically, variance decomposition can tell a researcher the percentage of the fluctuation in a time series attributable to other variables at a selected time horizon. More precisely, the uncorrelatedness of the orthogonalized shocks υ_t's allow us to decompose the error variance of the s step-ahead forecast of y_{it} into components accounted for by these shocks, or innovations. Consider an orthogonalized VAR with m components in vector MA representation,

$$y_t = \sum_{l=0}^{\infty} \psi^*(l)\upsilon_{t-1} \tag{3.27}$$

The s step-ahead forecast for y_t is then

$$\mathrm{E}_t(y_{t+s}) = \sum_{l=s}^{\infty} \psi^*(l)\upsilon_{t+s-l} \tag{3.28}$$

Defining the s step-ahead forecast as

$$e_{t+s} = y_{t+s} - \mathrm{E}_t(y_{t+s}) \tag{3.29}$$

we get $e_{t+s} = \sum_{l=0}^{s-1} \psi^*(l)\upsilon_{t+s-l}$ (3.30)

and its i'th component is given by

$$e_{i,t+s} = \sum_{l=0}^{s-1} \sum_{j=1}^{m} \psi_{ij}^{*}(l) v_{j,t+s-l} = \sum_{j=1}^{m} \sum_{l=0}^{s-1} \psi_{ij}^{*}(l) v_{j,t+s-l} \quad (3.31)$$

Now, because the shocks are both serially and contemporaneously uncorrelated, we get for the error variance

$$V(e_{i,t+s}) = \sum_{j=1}^{m} \sum_{l=0}^{s-1} V(\psi_{ij}^{*}(l) v_{j,t+s-l}$$

$$= \sum_{j=1}^{m} \sum_{l=0}^{s-1} \psi_{ij}^{*}(l)^2 V(v_{j,t+s-l})$$

All shock components have unit variance, and this implies that

$$V(e_{i,t+s}) = \sum_{j=1}^{m} \left(\sum_{l=0}^{s-1} \psi_{ij}^{*}(l)^2 \right) \quad (3.32)$$

where $\sum_{l=0}^{s-1} \psi_{ij}^{*}(l)^2$ accounts for the error variance generated by innovations to y_j. Comparing this to the sum of innovation responses we get a relative measure how important variable j's innovations are in the explaining the variation in variable i at different step-ahead forecasts, i.e.,

$$R_{ij,s}^2 = 100 \frac{\sum_{l=0}^{s-1} \psi_{ij}^{*}(l)^2}{\sum_{k=1}^{m} \sum_{l=0}^{s-1} \psi_{ik}^{*}(l)^2} \quad (3.33)$$

3.7 Forecasting

The h-step ahead of a VAR (p) process, $\hat{\boldsymbol{y}}_{t+h|t}$ is given by the formula

$$\boldsymbol{E_t}[\boldsymbol{y_{t+h}}] = \sum_{j=0}^{h-1} \boldsymbol{\Phi_1^j \Phi_0} + \boldsymbol{\Phi_p^h y_t} \quad (3.34)$$

Forecasts from higher order VARs can be constructed by direct forward recursion beginning at $h = 1$, but it is often computed using the deviations from the VAR since it includes no intercept,

$\tilde{\boldsymbol{y}}_t = \boldsymbol{\Phi}_1 \tilde{\boldsymbol{y}}_{t-1} + \boldsymbol{\Phi}_2 \tilde{\boldsymbol{y}}_{t-2} + \cdots + \boldsymbol{\Phi}_P \tilde{\boldsymbol{y}}_{t-P} + \boldsymbol{\epsilon}_t$

Using the deviations form h-step ahead forecasts from a VAR (p) can be computed using the recurrence

$$\boldsymbol{E_t}[\tilde{\boldsymbol{y}}_{\boldsymbol{t+h}}] = \boldsymbol{\Phi_1 E_t}[\tilde{\boldsymbol{y}}_{\boldsymbol{t+h-1}}] + \boldsymbol{\Phi_2 E_t}[\tilde{\boldsymbol{y}}_{\boldsymbol{t+h-2}}] + \cdots + \boldsymbol{\Phi_P E_t}[\tilde{\boldsymbol{y}}_{\boldsymbol{t+h-P}}] \quad (3.35)$$

starting at $\boldsymbol{E_t}[\tilde{\boldsymbol{y}}_{\boldsymbol{t+1}}]$. Once the forecast of $E_t[\tilde{y}_{t+h}]$ has been computed the h-step ahead forecast of y_{t+h} ahead is constructed by adding the long run mean

$$E_t[y_{t+h}] = \mu + E_t[\tilde{y}_{t+h}] \quad (3.36)$$

CHAPTER 4

DATA ANALYSIS

4.0 Introduction

This chapter analyses, discusses and interprets the results obtained from the study. The data was analyzed with the R and EVIEWS softwares. The chapter has 3 main subdivisions which are descriptive analysis of the data, further analysis, and discussion of results.

4.1 Descriptive analysis

In the empirical analysis, two aggregate series of cases namely hypertension and heart disease cases are used. Some descriptive statistics including the mean, median, minimum and maximum values of cases are presented in the Table 4.1 below. The result from the table indicate that the minimum number of monthly cases within the study period is 3 for hypertension and 0 for cardiac cases. The maximum number of cases recorded for hypertension was 717 and that of heart diseases, 57. The average monthly number of cases was 132.92 for hypertension and 9.935 for heart disease cases. The sample data for the 5 years consist of 60 data points and the median number of cases for hypertension and heart disease cases are 65 and 6 respectively.

Table 4. 1 Descriptive Analysis of Hypertension and Heart Disease

Disease	Observation	Mean	Median	Min.	Max.
Hypertension	60	132.92	65	3	717
Heart	60	9.933	6	0	57

An exploration of the number of cases per month for hypertension is displayed in Table 4.2. The highest average number of cases for the study period between 1st January 2010 and 31st December 2014 corresponds to the month of January i.e. 270 cases on the average and this could be attributed to over indulgence in the consumption of fatty diets and alcoholic drinks during Christmas. A lot of adults are also over anxious around that period because they worry about what to provide for their families.

The least average number of cases corresponds to the month of October (46.8 number of cases). In terms of the maximum and minimum number of cases, the maximum number of hypertension cases occurred in the month of February (717 number of cases) and the least number of cases occurred in the month of October (3 cases).

Table 4.2: Monthly descriptive Statistics of Hypertension

Month	Mean	Min	Max	Median
January	270	29	710	217
February	215.2	4	717	86
March	106	29	279	70
April	60.2	28	96	48
May	65.6	28	104	65
June	62.4	10	213	29
July	136.8	4	551	33
August	188.4	5	601	65
September	120	7	293	43
October	46.8	3	171	20
November	159.4	10	517	77
December	163.2	21	523	84

Table 4.3 also reports summary statistics for heart disease cases as was done for hypertension. The highest average number of heart disease cases occurred in the month of March (16.8 cases),

and the least average number of cases occurred in the month of November (ie 4.4 average number of cases).

The maximum number of cases for heart diseases occurred in the month of March ie (57 cases) and the minimum was 0 in the month of June.

Table 4. 3: Monthly descriptive Statistics of Heart Disease cases

Month	Mean	Min	Max	Median
January	10.4	2	25	7
February	11.6	1	34	6
March	16.8	1	57	8
April	14.4	1	54	5
May	5.4	1	12	3
June	12	0	33	9
July	9.2	2	18	7
August	12.4	5	39	6
September	11.6	1	24	11
October	5.2	4	7	5
November	4.4	1	14	2
December	5.8	1	12	5

4.2 Further analysis

4.2.1 Unit root properties of individual series

The time series are first checked for stationarity before an attempt is made to fit a model. The variables have to be checked for unit roots and the order of integration of each series must be determined. The stationarity of the series was tested using time series plots followed by the

Augmented Dickey Fuller (ADF) test, Philip Perrons (PP) test, and the Kwiatkowski-Phillip-Schmidt-Shin (KPSS) test. For both the ADF test and the PP test, the null hypothesis is,

H_0 : "The series is not stationary"

and the alternate hypotheses are,

H_1 : "the series is stationary"

For the KPSS, the hypothesis are

H_0 : "the series is stationary"

H_1: "the series is non-stationary"

4.2.1.1 Unit Root Test Results for Hypertension

From Figure 4.1, it can be observed that even though some of the monthly number of cases for hypertension are very high, there is no trend. This seems to suggest that the series is stationary. A careful observation of the correlogram seems to support this observation since even though there is a significant spike at the first lag of the PACF, the ACF plots do not slowly decay. But these observations can only remain a suspicion until they have been confirmed by the various unit root tests. Tests for stationarity employed in this study included the ADF test, the PP test and the KPSS test.

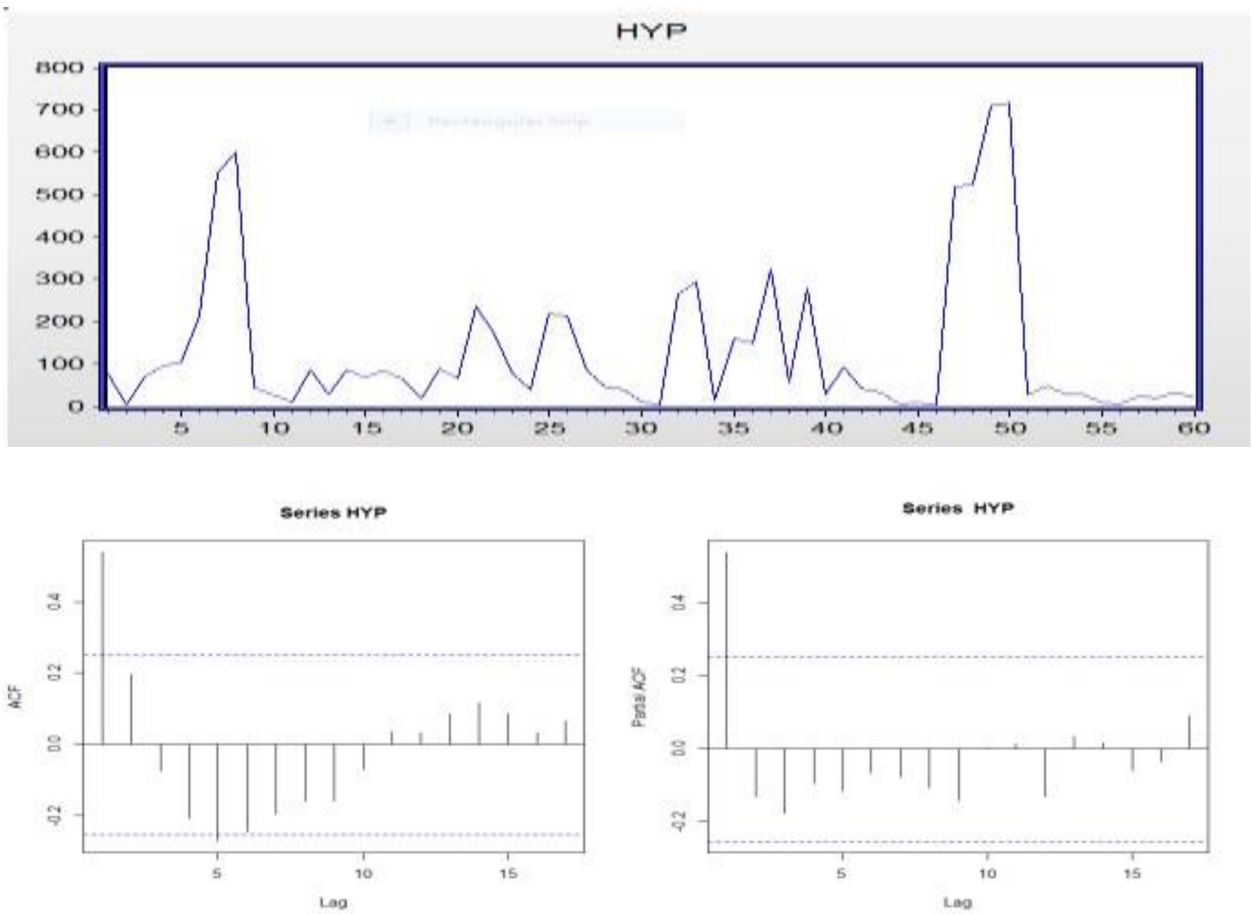

Figure 4.1: A time series display of the hypertension cases at levels

Quite clearly the series for hypertension is stationary whether or not there is an inclusion of a constant and/or trend or none of them. The evidence is seen from Tables 4.4, and 4.5. As a result there is no need to difference the series. In this case we say the series is integrated of order zero i.e. I(0).

Table 4.4: ADF test results for hypertension cases

HYPERTENSION	TEST STATISTIC	P-VALUE
CONSTANT	-4.1020	0.0020
CONSTANT+TREND	-4.0604	0.0118
NONE	-3.1675	0.0020

Source: Computations based on researchers own calculations from field data

From Table 4.4 all the p-values for the ADF tests are less than the conventional significance level of 0.05 which is an indication of stationarity per the test of hypotheses provided above.

Table 4.4: The PP stationarity test results for hypertension

HYPERTENSION	TEST STATISTIC	P-VALUE
CONSTANT	-4.1526	0.0017
CONSTANT+TREND	-4.1098	0.0103
NONE	-3.2019	0.0018

Source: Computations based on researchers own calculations from field data

Also from Table 4.5 all the p-values of the PP test are less than the conventional significance level of 0.05 which is an indication of stationarity per the hypotheses provided above.

Table 4.5: KPSS stationarity test results for hypertension

TEST	TEST STATISTIC	P-VALUE
KPSS	0.063	0.1

Source: Computations based on researchers own calculations from field data

The KPSS test results above confirms the results of the ADF and the PP test since for the KPSS test, a p-value greater than 0.05 is rather an indication of stationarity in the data.

4.2.1.2 Unit root test results for Heart disease cases

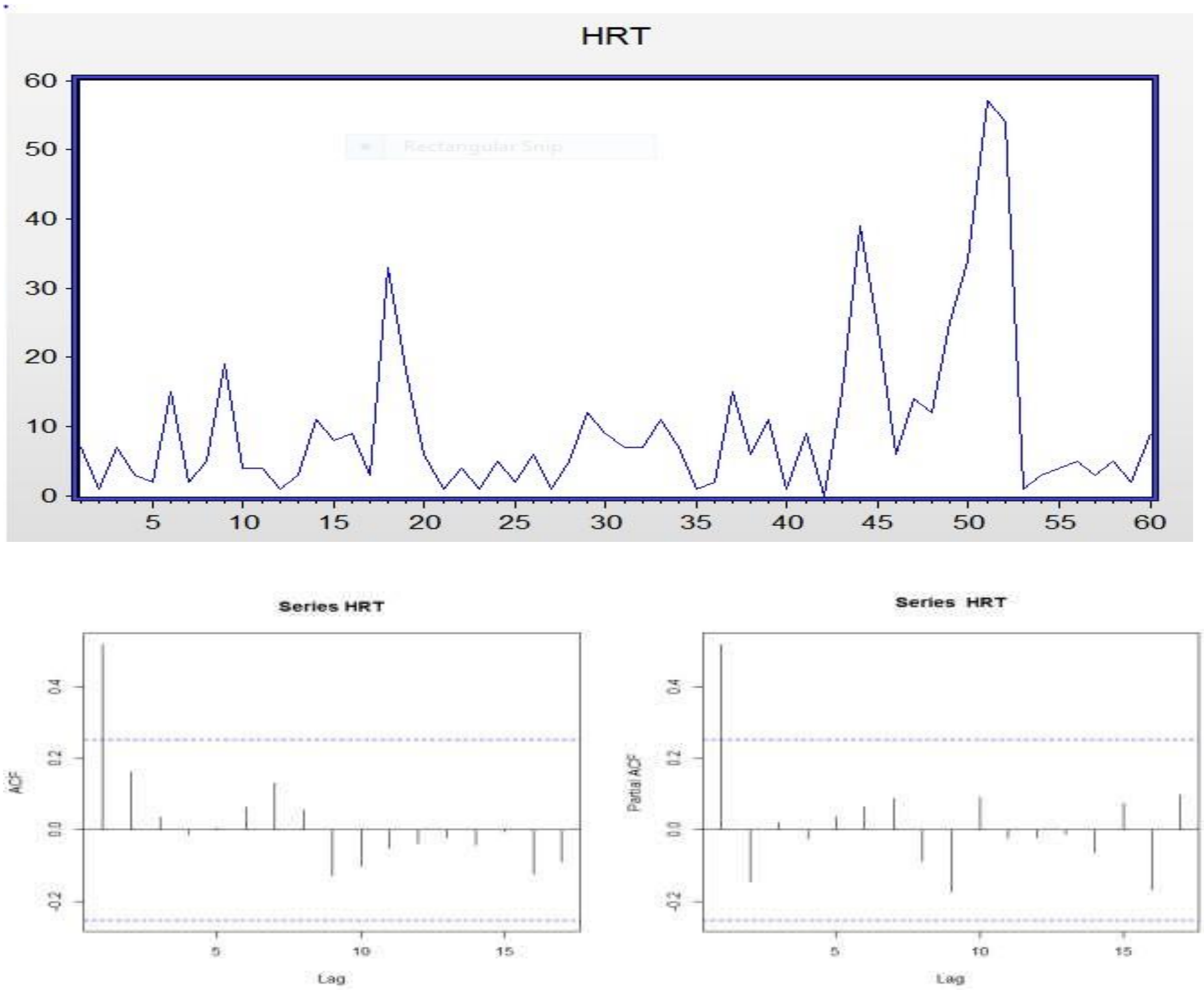

Figure 4.2: A time series display of heart disease cases at levels

From Figure 4.2, the PACF has a significant spike at lag one and again the lags of the ACF does not decay gradually. For both ADF test and PP tests, a p-value less than the conventional critical value of 5% signifies stationarity. However for the KPSS test, a p-value greater than the conventional 5% is rather an indication of stationarity. Hence the PP test and the KPSS test confirms the result of the ADF test.

Table 4.6: ADF unit root test results for heart disease cases

HEART DISEASE	TEST STATISTIC	P - VALUES
CONSTANT	-4.2458	0.0013
CONSTANT + TREND	-4.4031	0.0045
NONE	-3.0821	0.0026

Table 4.7: PP unit root test results for heart disease cases

HEART DISEASE	TEST STATISTIC	P - VALUES
CONSTANT	-4.2153	0.0014
CONSTANT + TREND	-4.2973	0.0061
NONE	-2.9628	0.0037

Table 4.8: KPSS unit root test results for heart disease cases

TEST	TEST STATISTIC	P - VALUE
KPSS	0.3962	0.0788

Results from Tables 4.7, 4.8, and 4.9, indicate that the heart cases just like the hypertension cases are stationary at levels.

4.3 Estimation of the VAR order

The order of a VAR model is the number of lags included in the model. The appropriate lag length (p), should be estimated long enough for the residuals not to be serially correlated. The lag length plays a crucial role in diagnostic tests as well as in the estimation of VAR models for impulse response analysis and variance decomposition (Bhasin, 2004).

Four of the selection criteria i.e. the Final Prediction Error (FPE), Akaike Information Criterion (AIC), the Schwarz Information Criterion(SC), and the Hannan-Quinn Information Criterion(HQ) support the inclusion of 2 lags as seen from Table 4.9. Hence, lag 2 is considered the optimal choice for the VAR order.

Table 4.9: Results for Model Specification

Included observations: 48

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-504.9961	NA	5122210.	21.12484	21.20281	21.15430
1	-484.7641	37.93500	2605350.	20.44851	20.68241	20.53690
2	-477.0053	13.90128	2230305.*	20.29189*	20.68172*	20.43920*
3	-473.9574	5.206705	2326778.	20.33156	20.87733	20.53781
4	-470.7547	5.204377	2416924.	20.36478	21.06648	20.62995
5	-468.5069	3.465370	2619797.	20.43779	21.29542	20.76189
6	-467.8973	0.889098	3051114.	20.57905	21.59262	20.96208
7	-463.7580	5.691514	3080954.	20.57325	21.74275	21.01521
8	-462.6477	1.434126	3548493.	20.69365	22.01909	21.19454
9	-461.2269	1.716794	4060543.	20.80112	22.28249	21.36093
10	-451.4470	11.00239*	3305429.	20.56029	22.19759	21.17903
11	-444.0895	7.664094	3004380.	20.42039	22.21363	21.09806
12	-442.6806	1.350151	3538460.	20.52836	22.47753	21.26495

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)

The Akaike Information Criteria (AIC) and the Schwarz Bayesian Information Criteria (BIC) are the two main penalty function statistics which penalizes fitted models based on the principle of parsimony.

Apart from the fact that the majority of the criteria support the inclusion of two lags, the AIC which was the main guideline also supports that. Therefore the estimated VAR is VAR(2).

We proceeded to estimate the long run relationship for heart disease cases and hypertension cases.

The long run relationship results for heart disease cases estimated by the VAR (2) model is as presented Table 4.11.

Table 4.10: The long run estimation results for heart disease cases

variable	Lag	Estimate	std error	t-value	p-value
Heart	1	0.4084	0.1258	3.2460	0.0020 **
Hypertension	1	0.0070	0.0080	0.8710	0.3874
Heart	2	-0.0917	0.1184	-0.7750	0.4419
Hypertension	2	0.0261	0.0086	3.0510	0.0035 **
Constant		2.4621	1.8934	1.3000	0.1991

The long run relationship results for hypertension cases estimated by the VAR (2) model is as presented Table 4.12

Table 4.11: The long run estimation results for hypertension cases

Variable	Lag	estimate	std error	t-value	p-value
Heart	1	-1.6614	2.1361	-0.778	0.4402
Hypertension	1	0.6070	0.1361	4.459	0.0000***
Heart	2	1.2000	2.0092	0.597	0.5529
Hypertension	2	-0.0968	0.1454	-0.666	0.5082
Constant		71.4235	32.1427	2.222	0.0306 *

Hence the estimated VAR (2) model is specified as:

$$HRT_t = 0.4084\,HRT_{t-1} + 0.0070\,HYPt_{t-1} - 0.0917HRT_{t-2} + 0.0261\,HYP_{t-2} + 2.4621 \qquad (4.1)$$

$$HYP_t = -\,1.6614\,HRT_{t-1} + 0.6070\,HYP_{t-1} + 1.200HRT_{t-2} - 0.0968HYP_{t-2} + 71.4235 \qquad (4.2)$$

It can be observed from equation (4.1) that heart disease is significantly affected positively by 40% when there is one unit change in its lagged values. It can also be inferred from the equation that heart diseases are affected positively by almost 3% when there is two unit change in the lagged values of hypertension. From equation (4.2) it can be observed that hypertension is affected by almost 61% when there is a unit change in its lagged values and also affected by 71% of the constant term.

4.4 Model Diagnostics

A test of misspecification must be conducted to find out whether the model is an appropriate representation.

4.4.1 Test of Residual Serial Correlation

Table 4.13 presents the results for the Breusch-Godfrey LM test for the residual serial correlation of VAR (2) model.

The hypotheses for this test is:

H_0 : “There is no serial correlation in the residual of this model”

H_1 5%. : “The residuals of the model are serially correlated”

The result seen from Table 4.13 gives a clear indication that there are no residual auto correlations since the associated p-value is greater than the conventional significance level of

Table 4.12: Test of residual autocorrelations

LAG	CHI – SQUARED VALUE	LM TEST (P-VALUE)
2	6.8816	0.5495

4.4.2 The ARCH (Multivariate Test)

The model was also checked for heteroscedastic arch effects. The hypothesis for this test

H_0: "There is no serial heteroscedasticity in the residauls"

H_1: "There is heteroscedasticity in the residuals"

Table 4.13: Test for heteroscedaticity

LAG	CHI SQUARED VALUE	P – VALUE
2	**22.6914**	**0.2027**

As seen from Table 4.14 the alternate hypothesis was rejected and so there are no ARCH effects.

4.4.3 Stability Test

A test for the stability of the model parameters was performed using the OLS-CUSUM test. From Figure 4.7, it was observed that the cumulative residuals of the model fall within the 95% confidence band. It can therefore be concluded that the parameters of the model are structurally stable.

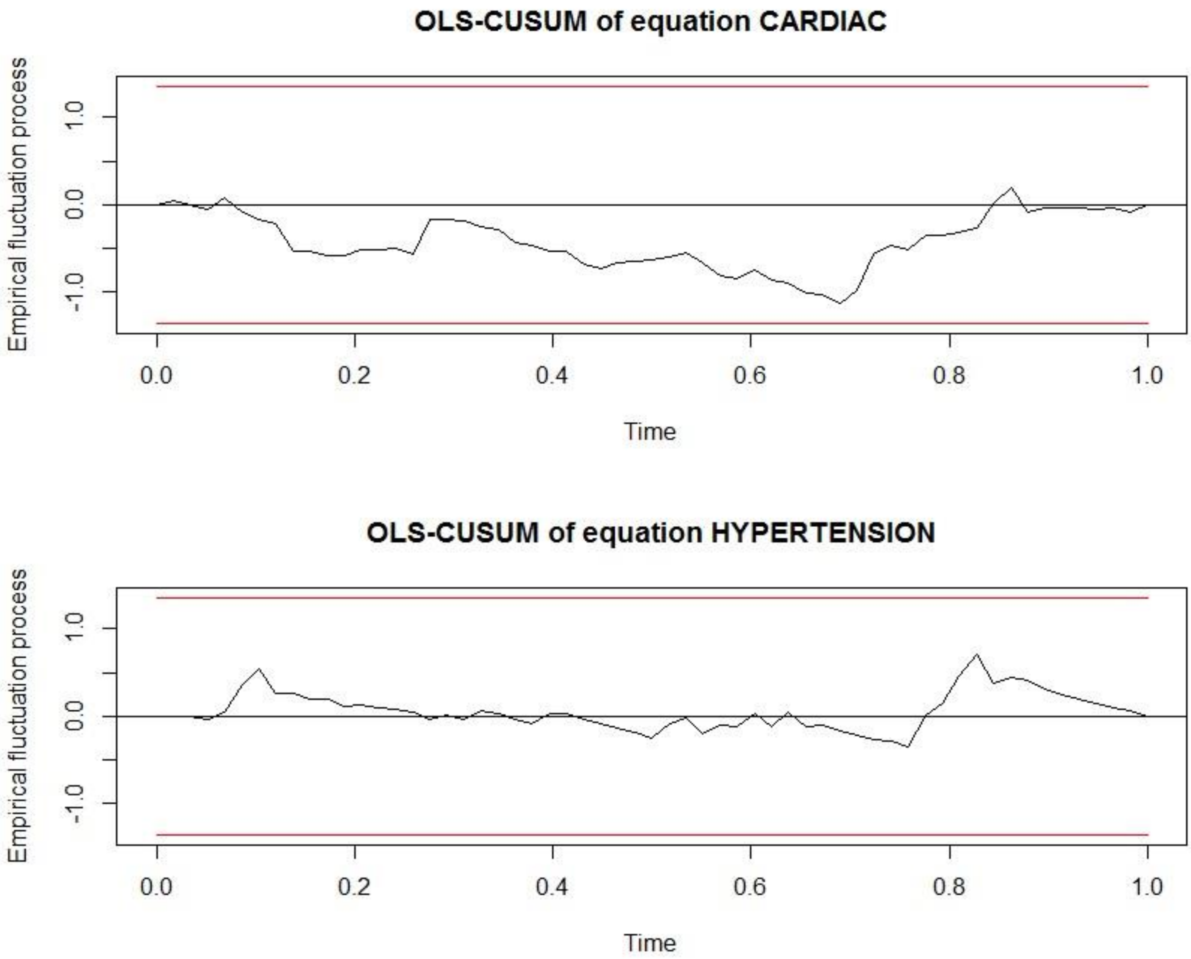

Figure 4.3: The OLS-CUSUM test for stability

4.5 Forecast for hypertension and heart disease cases

Since the model passed almost all the relevant diagnostic tests, it was used to make forecasts for heart disease and hypertension. The results are as shown in the tables below:

Table 4.14: Monthly forecasts of heart disease cases for the year 2015

MONTH	FORECAST	LCL	UCL
January	7	-11.0340	24.95995
February	5	-14.0400	25.05432
March	7	-15.4183	28.67774
April	8	-15.5193	32.43230
May	9	-14.8445	34.19446
June	10	-14.4445	34.70181
July	10	-14.3964	34.75353
August	10	-14.4355	34.71623
September	10	-14.4422	34.71005
October	10	-14.4190	34.73432
November	10	-14.3935	34.76046
December	10	-14.3792	34.77487

Table 4.15: Monthly forecast of hypertension cases for the year 2015

MONTH	FORECAST	LCL	UCL
January	68	-237.1020	373.9469
February	110	-248.6830	468.9887
March	131	-236.7848	498.5151
April	136	-232.2367	503.8059
May	135	-233.0077	503.1622
June	134	-233.7635	502.4418
July	135	-233.4546	502.7906
August	135	-232.7432	503.5444
September	136	-232.2143	504.0878
October	136	-231.9958	504.3081
November	136	-231.9617	504.3422
December	136	-231.9768	504.3272

4.6 Structural Analysis

The dynamic properties of a VAR (p) are often summarized using various types of structural analysis. The three main types of structural analysis summaries are

1. Granger causality test
2. Impulse response functions
3. Forecast error variance decompositions

4.6.1 Granger Causality Test

Granger causality test is considered a useful technique for determining whether one time series is good for forecasting the other. The concept of Granger-causality test is explored when the coefficient of the lagged of the other variable is not zero.

Table 4.16 presents the result from the Granger causality tests.

Table 4.16: Granger causality test

NULL HYPOTHESIS	F-STATISTIC	P-VALUE
Hypertension does not Granger-cause Heart Disease	-28.938	0.0005
Heart Disease does not Granger-cause Hypertension.	-20.3293	0.7209

The result shows that hypertension Granger-causes Heart diseases at the conventional significance level of 5%. This means that the past information of hypertension can be used to forecast future values of heart disease. The opposite however is not true as far as this study is concerned. From Table 4.16, clearly the test failed to reject the hypothesis that “heart disease does not Granger cause Hypertension”. This means causality is unidimensional since it is only one of the two variables that Granger causes the other.

4.6.2 Impulse Response Function

Impulse responses trace out the responsiveness of the dependent variables in the VAR model to shocks by each of the variables. It helps in identifying the reaction of the variable when a positive shock of one standard deviation is given to the error terms and to know in what manner the variables react to each other. So for each variable from each equation separately a unit shock is applied to the error terms and the effects upon the VAR system over time are noted. A standard decomposition is used in order to identify the shocks on level of the endogenous variables in the VAR.

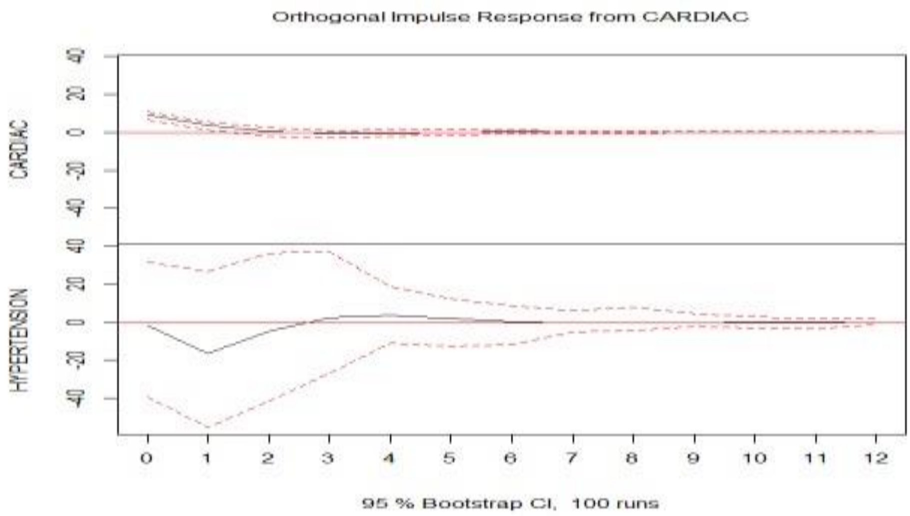

Figure 4.4: Graph of Impulse Response from Hypertension Disease.

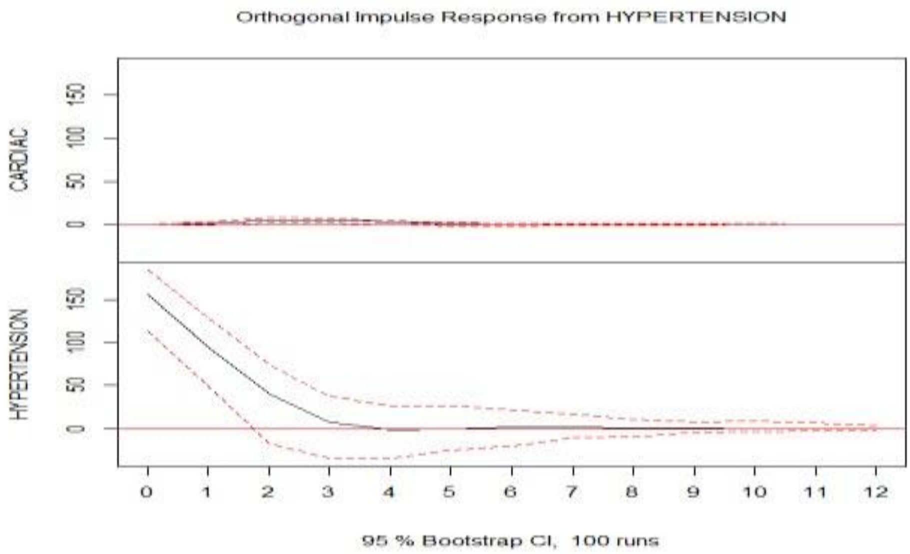

Figure 4.5: Graph of Impulse Response from Heart Disease.

The impulse response is represented in graphical form. The x-axis gives the time horizon or the duration of shock whilst the y-axis gives the direction and intensity of the impulse or percent variation in the dependent variable away from its base line level. The red dashed lines indicate the lower and upper bound of the confidence interval whilst the black lines within the red dashed lines traces out the percent variation. Figure 4.4 shows the response from Heart diseases to the standard deviation innovations of hypertension and response to their own shocks (Heart Disease shocks).

4.6.2.1 Analysis of the Response from Heart Disease cases

From Figure 4.4 it can be inferred that at the application of one shock of standard deviation to the cases of heart diseases the response to its own shock (by heart cases) is a gentle decrement in number of cases from period one to period three and then it maintains uniformity through out to the end.

Also at one standard deviation shock to hypertension cases the response from cardiac cases is an immediate decrement followed by a gentle increment from period one to period four after which it remains constant throughout to the end.

4.6.2.2 Analysis of Response from hypertension cases

From Figure 4.5 when one standard deviation is applied to cardiac cases, the response from hypertension is an insignificant increment from period one to period two after which it decreases gently to uniformity throughout for the rest of the period.

Hypertension cases reacts to a one unit shock to themselves by a sharp decrease in cases from the first period until the fourth period after which it remains constant on the x-axis.

4.6.3 Forecast Error Variance Decomposition

Variance decomposition offer a slightly different method for examining the VAR system dynamics. The decomposition used to understand the proportion of the fluctuation in a series explained by its own shocks as well as shocks from other variables. The results of the decomposition of the endogenous variables of the model are presented in Tables 4.17 and 4.18. The results provide percentages of the forecast error in each variable attributable to the innovations of the other variable over a time period.

Table 4.17: FEVD for Heart Cases

MONTH	SHOCKS FROM CASES OF HEART DISEASE (%)	SHOCKS FROM CASES OF HYPERTENSION (%)
1	100.00	0.00
2	98.81	1.19
3	77.88	22.12
4	66.09	33.91
5	66.30	36.70
6	63.02	36.98
7	63.02	36.98
8	63.03	36.97
9	63.02	36.97
10	63.02	36.98

The variance decomposition analysis result for Heart cases shows that at the initial stages variation in heart diseases is explained by its own lags until from the third month where 77.88%

of the variability in heart disease fluctuations is explained by its own lags. This further reduces to between 66.10 and 66.00 for the rest of the months.

Table 4.18: FEVD for Hypertension

MONTH	SHOCKS FROM CASES OF HYPERTENSION (%)	SHOCKS FROM CASES OF HEART DISEASE (%)
1	99.99	0.01
2	99.19	0.81
3	99.16	0.84
4	99.14	0.86
5	99.11	0.90
6	99.10	0.90
7	99.10	0.90
8	99.10	0.90
9	99.10	0.90
10	99.10	0.90

In the case of hypertension as shown in Table 4.18, more than 99% of its variability is explained by its own shocks from the first month to the tenth month. Heart once again does not have any significant variability to affect hypertension. Therefore variability in hypertension cases is explained significantly by their own throughout the ten months.

CHAPTER FIVE

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

5.0 Overview

This chapter consists of the summaries of findings and conclusion of the study. Recommendations and suggestions are made accordingly for further research.

5.1 Summary of findings

The maximum number of monthly cases recorded for hypertension for the study period was 717 and that of heart diseases, 57. The average monthly number of cases was 132.92 for hypertension and 9.935 for heart disease cases. The highest average number of monthly heart disease cases was associated with the months of March (16.8 cases), and the least average number of monthly heart disease was associated with the months of November (i.e. 4.4 average number of cases). The highest average number of monthly hypertension cases for the period selected for the study, was associated with the months of January. This could be due to over indulgence in the consumption of alcoholic beverages and fatty diets during Christmas celebrations in the months of December. The least number of monthly average hypertension cases was associated with the months of October.

Granger causality tests applied indicated that hypertension could Granger cause heart disease but the opposite was not true i.e. heart diseases could not Granger cause hypertension. Therefore the causality between hypertension and heart diseases was unidirectional.

The main aim of this study was to investigate the relationship between hypertension cases and cases of heart disease in the municipality of Ho between 1st January, 2010 and 31st December, 2014. It was observed that hypertension cases are positively related to heart disease cases.

Which means as hypertension cases increases, those of heart diseases also increase and the converse is also true i.e. as hypertension cases decreases those of heart diseases decrease.

It was also observed that, heart disease are affected by immediate past movements of their own cases as well as past two movements of hypertension cases. These lagged variables are therefore useful in predicting future values of heart disease cases. Accordingly, heart disease cases are affected positively by almost 41% when there is a unit change in their lagged values and almost 3% when there is two unit change in the lagged values of hypertension.

From our analysis it can be seen that hypertension cases can only be significantly explained by their first lag and it is seen at this lag that a unit change in lagged values of hypertension explains almost 61% hypertension cases. It was evidently clear that lags associated with heart disease cannot be used to explain hypertension at the conventional 5% level of significance and this confirms the Granger causality test results where it was seen that past values of heart diseases are not helpful in the prediction of hypertension.

The appropriate number of lags for the VAR modelling was identified to be two i.e. VAR (2).

The impulse response function employed revealed that heart disease shocks have no effect on hypertension but shocks from hypertension had a significant positive effect on heart diseases.

The variance decomposition analysis on Table 4.17 confirms the findings of the results from the impulse response analysis. The variability of both heart disease cases and hypertension cases are highly explained by their own shocks.

5.2 Conclusions

Our forecast for the year 2015 predicts a slight increase for heart disease cases in the Ho municipality.

The fact that the trend or increment is slight does not mean stakeholders in the area should relax on their oars. But rather, the Ministry of Health (MoH) should liaise with health workers in the area to provide intensive education to the indigenes on risk factors of heart diseases, stressing especially on hypertension and the need to report to medical facilities to check their blood pressure.

5.3 Limitations of study

The project is limited in terms of the length of data, as data for the years below 2010 was not readily available as at the time of request. Other setbacks include time constraints and difficulty in obtaining relevant materials.

5.4 Recommendations

A lot of sensitization programs must be organized by health workers and stake holders in the catchment of Ho so as to create awareness of hypertension and its complications. This can prompt people to go for regular checkups so that, should they be positive, they can start treatment early before it results into complications such as heart disease and stroke. Therefore such programs must throw more lights on how high blood pressure can lead to other dangerous conditions such as stroke and heart disease.

The general public should also be encouraged to eat balanced diets and fruits, and not to overindulge in the consumption of alcohol, fatty and salty diets.

I also recommend that any further studies on this topic using the VAR modeling approach should also include stroke to further confirm the result of this study.

REFERENCES

Addo J, Smeeth L, Leon DA. Hypertension in sub-saharan Africa: a systematic review. Hypertension.2007 Dec;50(6):1012-8.

Amoah AG. Hypertension in Ghana: a cross sectional community prevalence study in greater. Accra. Ethn Dis. 2003; 13: 310-315. Accra.

Addo, J. (2006). *The changing patterns of hypertension in Ghana: A study of four rural communities in the Ga district.* Available from:http//mxi.ishib.org/ED/journal/16-4/ethn-16-04-894.pdf. [accessed in January 2015]

Adriana P., Rodolfo J. D., Benigno R., Amparo Y. C., Victor D., Juan L.M., Maria C.C., (2003). An interrupted time series analysis of parenteral antibiotic use in Colombia. Journal of Clinical Epidemiology, Volume 56, Issue 10, Pages 1013-1020. htpp://sciencedirect.com/science/journal/08954356 [accessed in January 2015]

Anderson, T. W., (1971). The Statistical Analysis of Time Series. New York, Wiley.

Akaike, H., (1974). A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control, 19(6):716-723.

Australian Bureau of Statistics [ABS]. (2008). *Retail trade trends*. Retrieved on December 2012 from http://en.wikipedia.org/wiki/Australian_ Bureau_ of_Statistics

Burket BA. Blood pressure survey in two communities in the Volta region, Ghana, West Africa. *Ethn Dis.* 2006;16:292-294.

Bovet P, Gervasoni JP, Ross AG, Mkamba M, Mtasiwa DM, Lengeler C, Burnier M, Paccaud F:Assessing the prevalence of hypertension in populations: are we doing it right? J Hypertens 2003, 21:509-517.

Beaglehole R, Bonita R, Horton R, Adams C (2011). Priority actions for the non-communicable disease crisis. The Lancet. 2011;377(9775):1438-47.

Box, G. E. P., and Jenkins, G. M., (1976). Time Series Analysis: Forecasting and Control. Holden-Day, San-Francisco.

Banerjee, A., Lumsdaine R., Stock J. H., (1992). Recursive and Sequential Tests of the Unit-root and Trend-break hypothesis; Theory and International Evidence. *Journal of Business and Economic Statistics*, 10: 271-287.

Brown, R. L., Durbin, L., and Evans, J. M., (1975). Techniques for Testing the Consistency of Regression Relationships over time. *Journal of the Royal Statistical Society*, Series B 37: 149-192.

Box, G. E. P., and Pierce, D. A., (1970). Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Models. *Journal of the American Statistical Association*, 65: 1509-1526.

Box, G. E. P., and Jenkins, G. M., (1970). Time Series Analysis: Forecasting and Control. Holden-Day, San-Francisco.

Bollerslev, T., (1986). Generalised Autoregressive Conditional Heteroscedasticity. Journal of Econometrics, 31: 307-327.

Cappuccio FP, Micah FB, Emmett L, Kerry SM, Antwi S, Martin-Peprah R, (2004). Prevalence, detection, management, and control of hypertension in Ashanti, West Africa. Hypertension. 2004;43:1017-1022.

Christiano, L. J., (1992). Searching for a Break in GNP. Journal of Business and Economic Statistics, 10: 237-250.

Cunha J. P & Marks J, W. 2011. High blood pressure (hypertension). Available at:
http://www.medicinenet.com/high_blood_pressure/article.htm. [Accessed on January 2015]

Dickey, D. A., and Fuller, W. A., (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit-root. *Journal of the American Statistical Association*, 74: 427-431.

Durbin, J., and Koopman, S. J.,(2001). Time Series Analysis by State Space Methods. Oxford University Press, Oxford.

Dickey, D.A. and W.A. Fuller (1979). "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," Journal of the American Statistical Association, 74, 427–431.

de-Graft Aikins A, Unwin N, Agyemang C, Allotey P, (2010). Tackling Africa's chronic disease burden: From the local to the global. Global Health. 2010 Apr 19;6:5.

Engle, R. F., (1982). Autoregressive Conditional Heteroscedasticity with estimates of the variance of United Kingdom Inflation. *Econometrica*, 50: 987-1007.

Ezzati M, Lopez AD, Rodgers A, Vander Hoorn S, Murray CJL. Selected major risk factors and global and regional burden of disease. The Lancet. 2002;360(9343):1347-60.

Engle, R. F., and Granger, C. W. J., (1987). Cointegration and Error Correction: Representation, Estimation and Testing. *Econometrica*, 55: 251-276.

Erdogdu, E., (2007). Electricity Demand Analysis Using Cointegration ARIMA Modelling. A Case Study of Turkey. *Energy Policy*, 35: 1129-1146.

Gottman, J. M., (1981). Time Series Analysis: A Comprehensive Introduction for Social Scientists. Cambridge, UK: Cambridge University Press.

Girard D.Z., (2000). Intervention times series analysis of pertussis vaccination in England and Wales. Health Policy (2000) 54: 13-25.

Goka, F.(2007). Trend analysis of diseases reported at outpatient departments: A case study of the Greater Accra Region. An MSc. dissertation presented to the Department of Mathematics and Statistics, University of Cape Coast, Cape Coast, Ghana.

Gaziano TA, Galea G, Reddy KS. Scaling up interventions for chronic disease prevention: the evidence. The Lancet. 2007;370(9603):1939-46.

Greater Accra Regional Health Directorate: Annual Report 2007. Accra: Ghana Health Service 2008.

Ghana Health Service. The health sector in Ghana: facts and figures 2009. Accra: GHS; 2009.

Harvey, A. C., and Phillips, G. D. A., (1979). Maximum Likelihood Estimation of Regression Models with Autoregressive Moving Average Disturbances. *Biometrika*, 66: 49-58.

Hillmer, S. C., and Tiao, G. C., (1979). Likelihood Function of Stationary Multiple Autoregressive Moving Average Models. *Journal of the American Statistical Association*, 74: 652-660.

Harvey, A. C., and Peters, S., (1984). Estimation Procedures for Structural Time Series Models. LSE Econometrics Programme Discussion Paper A 44.

Hosking, J. R. M., (1981). Fractional Differencing. *Biometrika*, 68: 165-176.

Hendriks M, Brewster L, Wit F, Bolarinwa O, (2011). Cardiovascular disease prevention in rural Nigeria in the context of a community based health insurance scheme: Quality Improvement Cardiovascular care Kwara-I (QUICK-I). BMC Public Health. 2011;11(1):186.

Hylleberg, S., Engle, R. F., and Granger C.W. J., and Yoo, B. S., (1990). Seasonal Integration and Cointegration. Journal of Econometrics, 44: 215-238.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y., (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit-root; How Sure are we that Economic Time Series have a unit-root? *Journal of Econometrics*, 54: 159-178.

Kearney PM, Whelton M, Reynolds K, Muntner P, Whelton PK, He J. Global burden of hypertension: analysis of worldwide data. Lancet. 2005 Jan 15-21;365(9455):217-23.

Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, (2010). A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990?2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet. 2012;380(9859):2224-60.

Lopez A, Mathers C, Ezzati M, (2006). Global burden of disease and risk factors. Washington, DC: Oxford University Press and World Bank, 2006.

McAlister FA, Straus SE: Measurement of blood pressure: an evidence based review. BMJ 2001, 322:908-911.

Mann, H. B., and Wald, A., (1943). On Stochastic Limit and Order Relationships. *Annals of Mathematical Statistics*, 14: 217-226.

Nelson, C. R., and Plosser, C. I., (1982). Trends and Random Walks in Macroeconomic Time Series. *Journal of Monetary Economics*, 10: 139-162.

Newbold, P., (1974). The Exact Likelihood Function for a Mixed Autoregressive Moving Average Process. *Biometrika*, 61: 423-426.

Owusu I. K. 2007. Causes of Heart Failure as Seen in Kumasi, Ghana. Available at :http://www.ispub.com. Accessed 15 January 2011.

Opie LH, Seedat YK. Hypertension in sub-Saharan African populations. Circulation. 2005 Dec 6;112(23):3562-8.

Phillips, P. C., B., and Perron, P., (1988). Testing for a Unit-root in Time Series Regression. Biometrika, 75: 335-346.

Phillips, P.C.B., (1987): “Time Series Regression with a Unit Root,” *Econometrica*, 55, 277-301.

Quenouille, M. H., (1947). Notes on the Calculation of the Autocorrelation of Linear Autoregressive Schemes. *Biometrika*, 34: 365-367.

Robinson, P. M., and Yajima, Y., (2002). Determination of Cointegration Rank in Fractional Systems. *Journal of Econometrics*, 106: 217-241.

Slutzky, E. E., (1927). The Summation of Random Causes as the Source of Cyclic Processes. *The Problem of Economic Conditions, edition by the Conjecture Institute Moscow*, 3(1): 34-64.

Schwarz, G. E., (1978). Estimating the Dimensions of a Model. *Annals of Statistics*, 6(2): 461-464.

Seedat YK. Hypertension in developing nations in sub-Saharan Africa. Journal of human hypertension. 2000;10-11(14):739-4.

Schwandt H, Wealth Shocks and Health Outcomes: Evidence From Stock Market Fluctuations, CEPS Discussion Paper No 1281, July 2014

Suleman N. and Sarpong S., Statistical modeling of hypertension cases in Navrongo, Ghana, West Africa.

Twagirumukiza M, Van Bortel LM. Management of hypertension at the community level in sub-Saharan Africa (SSA): towards a rational use of available resources. J Hum Hypertens. 2011 Jan;25(1):47-56.

Unwin N, Setel P, Rashid S, Mugusi F, Mbanya JC, Kitange H, (2001). Noncommunicable diseases in sub-Saharan Africa: where do they feature in the health research agenda? *Bull World Health Organ.* 2001;79:947-953.

van de Vijver SJ, Oti SO, Agyemang C, Gomez GB, Kyobutungi C. Prevalence, awareness, treatment and control of hypertension among slum dwellers in Nairobi, Kenya. J Hypertens. 2013 May;31(5):1018-24.

Weiss, A. A., (1984). ARMA Models with ARCH Errors. *Journal of Time Series Analysis*, 5: 129-143.

Whelton PK, Beevers DG, Sonkodi S. Strategies for improvement of awareness, treatment and control of hypertension: Results of a panel discussion. J Hum Hypertens. 2004 Aug;18(8):563-5.

Yule, G.U., (1927). On a Method of Investigating Periodicities in Distributed Series with Special Reference to Wolfer's Sunspot Numbers. *Philosophical Transaction of Royal Society of London, Series A*, 226: 267-298.

Yusuf S, Reddy S, Ounpuu S, Anand S. Global burden of cardiovascular diseases: part I: general considerations, the epidemiologic transition, risk factors, and impact of urbanization. Circulation. 2001;104:2746-2753.

Zivot, E., and Andrews, D. W., (1992). Further Evidence on the Great Crash, the Oil-Price Shock and the Unit-root hypothesis. *Journal of Business and Economic Statistics*, 10: 251-270.

Zareian Z. 2004. Hypertensive disorders of pregnancy. Available at: http://www.sciencedirect.com [accessed on January 2015]