FORECASTING UTILIZATION BY SUBSCRIBERS OF THE NATIONAL HEALTH INSURANCE SCHEME (NHIS)

A CASE STUDY OF THE NADOWLI DISTRICT MUTUAL HEALTH INSURANCE SCHEME (DMHIS)

By

Tindogo Rashid Mohammed B.Sc (Hons.)

A Thesis submitted to the Department of Mathematics, Kwame Nkrumah University of Science and Technology in partial fulfillment of the requirement for the degree of

MASTER OF SCIENCE

Faculty of Science, Institute of Distance Learning (IDL)

DECLARATION

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

TINDOGO RASHID MOHAMMED

(PG4068010) …………………….. ……………

(Student Name & ID) Signature Date

Certified by:

Dr. F.T. ODURO …………………….. ……………

(Supervisor's Name) Signature Date

Certified by:

Mr. K. F. DARKWA …………………….. ……………

(Head of Department, Mathematics) Signature Date

ABSTRACT

Since the implementation of Ghana's National Health Insurance Scheme (NHIS) in 2004, the monthly healthcare facility utilization by subscribers seems to be rising. This study presents an Auto Regressive Integrated Moving Average (ARIMA) model to ascertain the pattern of growth. It also presents a forecast for six leading months. Monthly time series utilization data for the Nadowli District Mutual Health Insurance Scheme (DMHIS) for the period March 2006 to June 2012 were analysed by time series methods. Appropriate Box-Jenkins model was fitted. Validity of the model was tested using standard statistical techniques. The forecasting power of the model was validated with six months in-sample observations. The identified model is a Seasonal Autoregressive Integrated Moving Average (SARIMA) presented as ARIMA(0,1,0)(1,0,0)^{12} i.e. $X_t = X_{t-1} + 0.492(X_{t-12} - X_{t-13}) + W_t$, with a twelve (12) months period.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AC	Auto-Correlations
ACF	Auto-Correlation Function
AIC	Akaike Information Criterion
AR	Autoregressive
ARIMA	Auto Regressive Integrated Moving Average
ARMA	Auto-Regressive Moving Average
BIC	Bayesian Information Criteria
DMHIS	District Mutual Health Insurance Scheme
GHS	Ghana Health Service
GOG	Government of Ghana
HMO	Health Maintenance Organisation
ILO	International Labour Organisation
IMF	International Monetary Fund
IPD	In-Patients Department
LI	Legislative Instrument
MA	Moving Average
MAPE	Mean Absolute Percentage Error
MDGs	Millennium Development Goals
MHIS	Mutual Health Insurance Scheme
MSE	Mean Squared Error
NHIA	National Health Insurance Authority
NHIC	National Health Insurance Council
NHIF	National Health Insurance Fund
NHIS	National Health Insurance Scheme

OPD	Out Patients Department
PAC	Partial Auto-correlation Coefficient
PACF	Partial Auto-Correlation Function
PNDC	Provisional National Defense Council
SAP	Structural Adjustment Programme
SARIMA	Seasonal Auto-Regressive Integrated Moving Average
SIC	Schwarz-Bayesian Information Criteria
SPSS	Statistical Product and Service Solutions (Statistical Package for the social sciences)
SSNIT	Social Security and National Insurance Trust
STEP	Strategies and Tools against social Exclusion and Poverty
VAT	Value Added Tax
WHO	World Health Organisation

ACKNOWLEDGEMENT

This thesis would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study.

First and foremost, my profound gratitude and deep regards to my supervisor, Dr. F.T. Oduro (Department of Mathematics-KNUST), whose help, stimulating suggestions and encouragement helped me to successfully complete this thesis.

I would also like to acknowledge with much appreciation the crucial role of my study companions and course mates Awal Mohammed (Techiman) and Jude Bumbokuri (World Vision - Nadowli), whose sincerity and encouragement I will never forget.

I am also obliged to thank the Claims Manageress of the Nadowli DMHIS, Miss Lily Yendaw, for making the data available for this work.

Lastly, I thank the Almighty, family and friends for their constant encouragement. They have been my inspiration as I hurdle all the obstacles in the completion this research work.

DEDICATION

I dedicate this thesis to my wife, Zubeidata Hakeem and child, Eaman-Ullah Rashid Tindogo whose invaluable time I stole to accomplish this task.

CHAPTER 1

INTRODUCTION

1.0 INTRODUCTION

This chapter constitutes the background of the study, statement of the problem, Aims and objectives of the study, significance of the study area, and limitations of the study area and the organization of the study.

1.1 BACKGROUND

Health care financing is of primary necessity, as health care itself. Access to health care is largely determined by health care financing and the World Health Organisation (WHO) has itemized health care financing among four essential needs every country has to take care of. (Owusu, 2009).

In Ghana, hospital fee system has been operational since the first colonial hospital was built in 1868 (Gyapong et al, 2007). Pre-Independence financing for modern health care services was mainly by out-of-pocket payments at the point of service use. Immediate post-independence saw the "free health care for all" policy. Challenges due to economic decline necessitated the introduction of user fees from the late 1970s to the 1980s, this culminated in the introduction of very small out-of-pocket payments at the point of service usage within the public sector healthcare services in 1972. According to Owusu (1999), this was not to recover cost but to discourage frivolous usage. As part of health sector reforms in 1980s, user fees known as the "cash and carry" was introduced as means of sustaining the health services, which adversely affected utilization. This was part of the Structural Adjustment Programme (SAP) undertaken by the PNDC with the International Monetary Fund (IMF) and the World Bank (WB). In 1985, Legislative Instrument (LI 1313) was passed; aim to recover at least 15 percent of recurrent health expenditure. The exemption policy was introduced to address the burden of user fees on the poor and certain vulnerable groups in the society. The Government of Ghana in an effort to offset the challenges commissioned various studies into alternative health financing, principally insurance based. Even though the government allocation to health has seen some level of improvement over the years, health care financing still remains a major challenge in the country.

Subsequent inequities in financial access to basic and essential clinical services resulted in the formation of community-based health insurance schemes introduced mostly by mission health facilities in conjunction with communities. This was mostly in the early 1990s. The first Mutual Health Insurance Scheme (MHIS) in Ghana is the Nkoranza Scheme, which was initiated by the Catholic Diocese of Sunyani in 1989. Other schemes, such as Damongo and Dangme West MHISs followed and became models for other communities to replicate. By the year 2001 a total of 47 of such schemes were operational nationwide (Owusu, 2010: pg 8). The emerging and promising awareness of the need for health insurance in the context of poverty reduction and social protection became even much clearer.

The NHIS bill passed into law in 2003 provided the basis for the establishment of MHIS at the district level. The LI which serves as a regulatory framework for the NHIS was passed in 2004. The National Health Insurance Scheme (NHIS) is a social intervention program intended to provide financial risk protection against out of pocket health care expenditure for all residents in Ghana. According to the National Health Insurance Authority (NHIA), 2010 Annual Report, the NHIS is currently operational in 145 districts across the country with a total cumulative membership of over 18 million, out of which over 8 million, representing 34% of Ghana’s current population are active card bearing members as at the close of 2010.

1.1.1 HISTORY OF HEALTH INSURANCE

According to Wikipedia (accessed 2011 November 15), accident insurance was first offered in the United States by the Franklin Health Assurance Company of Massachusetts. This firm, founded in 1850, offered insurance against injuries arising from railroad and steamboat accidents. Sixty organizations were offering accident insurance in the US by 1866, but the industry consolidated rapidly soon thereafter. In 1887, the African American workers in Muchakinock, Iowa, a company town, organized a mutual protection society. Members paid fifty cents a month or $1 per family for health insurance and burial expenses. In the 1890s, various health plans became more common. The first employer-sponsored group disability policy was issued in 1911.

Commercial insurance companies began offering accident and sickness insurance (disability insurance) as early as the mid-19th century. The first group medical plan was purchased from The Equitable Life Assurance Society of the United States by the General Tire & Rubber Company in 1934. Before the development of medical expense insurance, patients were expected to pay all other health care costs out of their own pockets, under what is known as the fee-for-service business model. During the middle to late 20th century, traditional disability insurance evolved into modern health insurance programs. Today, most comprehensive private health insurance programs cover the cost of routine, preventive, and emergency health care procedures, and also most prescription drugs, but this was not always the case.

During the 1920s, individual hospitals began offering services to individuals on a pre-paid basis. The first group pre-payment plan was created at the Baylor University

Hospital in Dallas, Texas. This concept became popular among hospitals during the Depression, when they were facing declining revenues. The Baylor plan was a forerunner of later Blue Cross plans. Physician associations began offering pre-paid surgical/medical benefits in the late 1930s Blue Shield plans. Blue Cross and Blue Shield plans were non-profit organizations sponsored by local hospitals (Blue Cross) or physician groups (Blue Shield). As originally structured, Blue Cross and Blue Shield plans provided benefits in the form of services rendered by participating hospitals and physicians ("service benefits") rather than reimbursements or payments to the policyholder.

Hospital and medical expense policies were introduced during the first half of the 20th century. During the 1920s, individual hospitals began offering services to individuals on a pre-paid basis, eventually leading to the development of Blue Cross organizations. The Ross-Loos Clinic, founded in Los Angeles in 1929, is generally considered to have been the first health maintenance organization (HMO). Henry J. Kaiser organized hospitals and clinics to provide pre-paid health benefits to his shipyard workers during World War II. This became the basis for Kaiser Permanente HMO. Most early HMOs were non-profit organizations. The development of HMOs was encouraged by the passage of the Health Maintenance Organization Act of 1973. The first employer-sponsored hospitalization plan was created by teachers in Dallas, Texas in 1929. Because the plan only covered members' expenses at a single hospital, it is also the forerunner of today's Health Maintenance Organizations (HMOs).

Employer-sponsored health insurance plans dramatically expanded as a result of wage controls during World War II. The labor market was tight because of the increased demand for goods and decreased supply of workers during the war. Federally imposed wage and price controls prohibited manufacturers and other employers raising wages high enough to attract sufficient workers. When the War Labour Board declared that fringe benefits, such as sick leave and health insurance, did not count as wages for the purpose of wage controls, employers responded with significantly increased benefits.

Employer-sponsored health insurance was considered taxable income until 1954.

In the United States, regulation of the insurance industry is highly Balkanized, with primary responsibility assumed by individual state insurance departments. Whereas insurance markets have become centralized nationally and internationally, state insurance commissioners operate individually, though at times in concert through a National Insurance Commissioners' organization. In recent years, some have called for a dual state and federal regulatory system for insurance similar to that which oversees state banks and national banks.

Elsewhere in the world, Germany has Europe's oldest universal health care system, with origins dating back to Otto von Bismarck's Social Legislation, which included the Health Insurance Bill of 1883, Accident Insurance Bill of 1884, and Old Age and Disability Insurance Bill of 1889. As mandatory health insurance, these bills originally applied only to low-income workers and certain government employees; their coverage, and that of subsequent legislation gradually expanded to cover virtually the entire population.

1.1.2 ACCESS TO HEALTH CARE

Access to health care or health services refers to the possibility that exist for people to make use of health care or health services. In order for everyone to enjoy access to health care or health services, steps must be taken to remove barriers, particularly economic, financial or cultural barriers, as well as those relating to the supply of health care when the latter is either non-existent or overburdened (and therefore inadequate to meet demand). Setting up a health insurance scheme facilitates access to health care and services by removing certain financial barriers, but does not always resolve problems of geographic or cultural accessibility (ILO / STEP, 2005: pg 1).

Geographical accessibility: Access to health care of acceptable quality by the inhabitants of a village may be limited by the distance between the village and healthcare providers, or by a lack of organized transport.

Cultural accessibility: Access to healthcare and the selection of treatment options may to some extent be influenced by social perceptions, attitudes towards illness and maternity, or family and community strategies for dealing with illness and maternity.

Financial accessibility: The ability of people to pay for the services that they need at the time of illness is by far the biggest barrier to access to health care or health services. According to the Ghana Health Service, GHS (2010 Annual Report), Health facilities now earn more than 80% of their IGF from insured clients. This makes NHIS the main source of funding for health care services in Ghana.

In Ghana, access to modern healthcare services is increasing steadily with the introduction of the Community Health Planning Services (CHPS) program. Almost every community has a CHPS zone or a health centre within reasonable distance.

1.1.3 NATIONAL HEALTH INSURANCE SCHEME IN GHANA

The National Health Insurance Scheme in Ghana was established by an Act of parliament, ACT 650, in 2003 and backed with a Legislative Instrument, LI 1809, 2004. The objective of the scheme is to ensure access to basic healthcare services to all residence of Ghana. It is a socialized mutual health insurance scheme governed by the principles of equity, solidarity and affordability. We will now take a look at ACT 650 and L.I.1809 to put the NHIS into context.

MANAGEMENT AND FUNDING

ACT 650 established the National Health Insurance Authority (NHIA) which is a regulator of the National Health Insurance Scheme. The governing body of the National Health Insurance Authority is the National Health Insurance Council (NHIC). The object of the NHIA is to secure the implementation of the National Health Insurance policy that ensures access to basic healthcare services to all residents.

The direct implementation of the policy is carried out by Health Insurance Schemes. The following types of health insurance schemes may be established and operated in the country:

a. district mutual health insurance schemes (DMHIS),
b. private commercial health insurance schemes (PCHIS), and
c. private mutual health insurance schemes. (PMHIS)

The National Health Insurance Scheme (NHIS) basically is the complement of all District Mutual Health Insurance Schemes (DMHIS) in the country. It is a social not for profit scheme funded mainly by government through tax, payroll levy, contributions from the informal segment of society and donations. All these funds are paid into the National Health Insurance Fund (NHIF) which is managed by the NHIA. The NHIA gives subsidies and reinsurance to the DMHIS for the payment of claims and for the administration or day to day running of the schemes. The diagram below shows the cash flow of the NHIS

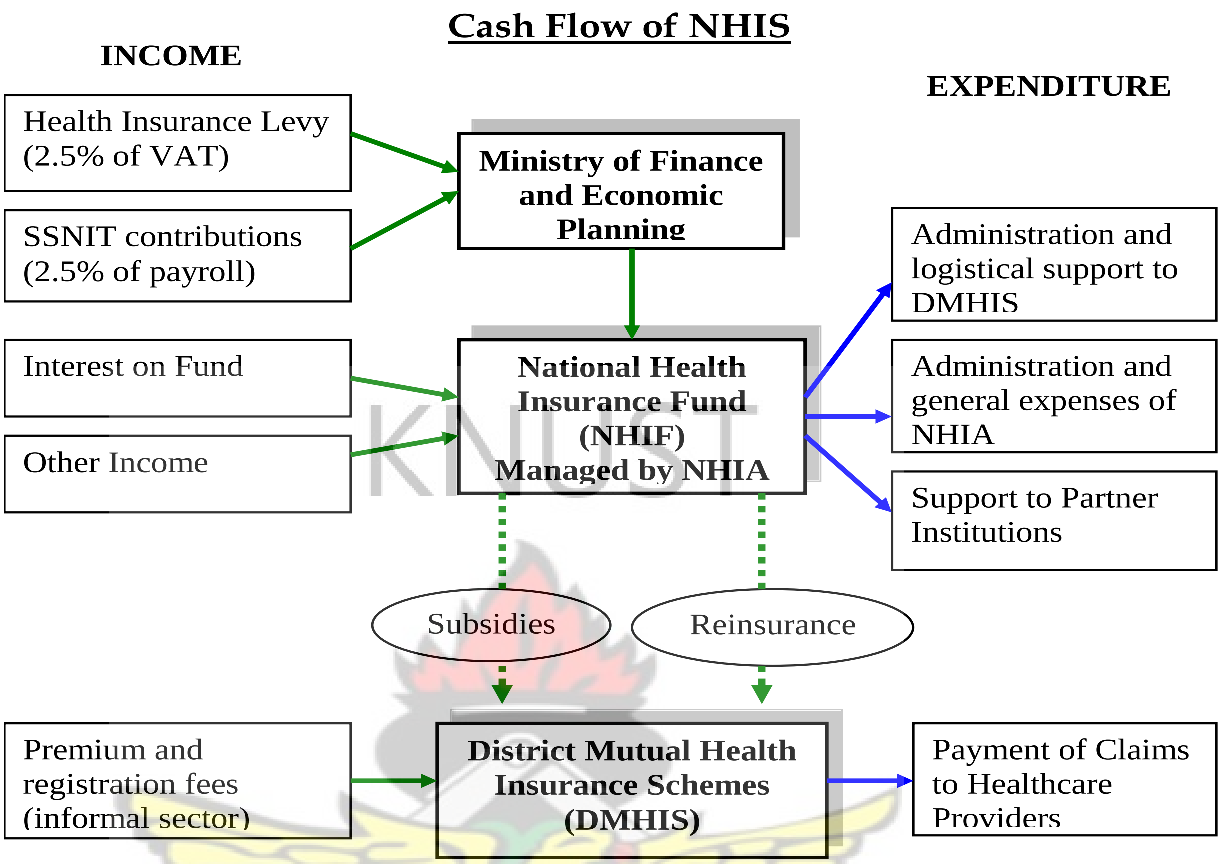


Figure 1.3 Cash flow of NHIS

Source: Benjamin A. Markin Yankah, NHIA 2011

The Private Commercial Health Insurance Scheme (PCHIS) is an insurance scheme operated by a private entity for profit. These are operated by approved insurance companies. Under this scheme, one can just walk into any of such companies and buy the insurance for yourself and your dependants just as you would for a car. Commercial health insurance companies do not receive subsidy from the National Health Insurance Fund and they are required to pay a security deposit before they start operations. The benefit packages for most private health insurance schemes are limited and they are also characterised by high premiums.

The third type of health insurance scheme in Ghana is known as the Private Mutual Health Insurance Scheme (PMHIS). This is an arrangement by any group of people (workers or any social group) coming together and making contributions to cater for their health needs, providing for services approved by the governing council of the scheme. Private Mutual Health Insurance Schemes are not entitled to subsidy from the National Health Insurance Fund. The motive for such schemes is not for profit but to protect their social group against catastrophic healthcare expenditure and cater for their most come health needs.

GOVERNING BODY OF THE AUTHORITY

The governing body of the Authority is a Council consisting of a chairperson and representatives of professional organizations and ministries implementing healthcare, social protection, law, insurance and finance as well as civil society

The chairperson and the other members of the Council are appointed by the President in accordance with article 70 of the constitution of Ghana.

MEMBERSHIP

The membership of the NHIS is open to all residents of Ghana. A person who registers with the scheme usually undergoes a waiting period of three months to qualify for free access to health care using the NHIS. The exception is pregnant women and children under five years of age. These categories of members are eligible for instant issuance of an NHIS ID card and immediate benefit. The aim is to accelerate the achievement of the Millennium Development Goals (MDGs), on the reduction of infant mortality and maternal deaths. The categories of membership of the NHIS are;

1. Informal Sector members
2. SSNIT Contributors
3. SSNIT Pensioners
4. Children under 18 years of age
5. Persons 70 years and above
6. Indigents and
7. Pregnant Women

Informal sector workers who do not contribute to the Social Security and National Insurance Trust (SSNIT) are the only category of members who pay premiums directly to the DMHIS at the point of registration. The premium ranges from GH¢ 7.20 to GH¢ 48.00. Persons 70 years and above, SSNIT Contributors, SSNIT Pensioners and Children under 18 years, pay only a registration fee at the point of registration. Pregnant Women and Indigents pay nothing at all at the point of registration. Those that are not required to pay premium are together referred to as the exempt group of members. Their premium is catered for by the National Health Insurance Authority through subsidies from the NHIF and other donations. The NHIF also provides reinsurance to the District Mutual Health Insurance Schemes. A person whose membership of the NHIS expires for more than three months is considered a defaulting member and for such persons, they have to undergo a waiting period of three months after renewal to rejoin the scheme. A member of the NHIS can be suspended or withdrawn from the scheme for inappropriate behavior which may be detrimental to the sustainability of the scheme.

BENEFIT PACKAGE

First and foremost, a member of the NHIS is entitled to a membership card. This enables him / her to be identified by accredited providers as beneficiaries of free healthcare services. Schedule II of the legislative Instrument, LI1809, Part 1 specifies the minimum healthcare benefit package under the National Health Insurance Scheme. The benefit package covers over 95% of all tropical diseases and health conditions (NHIA).

MINIMUM HEALTHCARE BENEFITS

1. Out-patient Services .

2. In Patient services

3. Oral Health services including

4. Eye Care services including

5. Maternity care including

6. Emergencies

All emergencies are also covered. These refer to crisis health situation that demand urgent intervention and include,

a. Medical emergencies;
b. Surgical emergencies including brain surgery due to accidents;
c. Paediatric emergencies;
d. Obstetric and Gynaecological emergencies including Caeserian Sections;
e. Road Traffic Accidents;
f. Industrial and workplace Accidents;
g. Dialysis for acute renal failure.

Part 2 of the same Schedule II, also spells out the list of services excluded from the benefit package. These healthcare services are not covered under the minimum benefits available under the National Health Insurance Scheme; however, Health Insurance Schemes may decide to offer any of them as additional benefits to their members.

EXCLUSION LIST

a. Rehabilitation other than physiotherapy;
b. Appliances and prostheses including optical aid, hearing aids, orthopedic aids, dentures;
c. Cosmetic surgeries and aesthetic treatments;
d. HIV retroviral drugs
e. Assisted Reproduction eg. Artificial insemination and gynaecological hormone replacement therapy;
f. Echocardiography;
g. Photography
h. Angiography;
i. Orthoptics;
j. Dialysis for chronic renal failure;
k. Heart and brain surgery other than those resulting from accidents;
l. Cancer treatment other than cervical and breast cancer;
m. Organ transplantation;
n. All drugs that are not listed on the NHIS Drug List;
o. Diagnosis and treatment abroad;
p. Medical examinations for purposes of visa applications, educational, institutional, driving licence;
q. VIP ward (Accommodation);
r. Morturary Services.

The following healthcare services are offered free of charge by the public health services and hence are also excluded from the minimum benefits package under the NHIS

a. Immunization;
b. Family planning;
c. In-patient and Out-patient treatment of mental illnesses;
d. Treatment of Tuberculosis, Onchocerciasis, Buruli Ulcer, Trachoma; and
e. Confirmatory HIV test on AIDS Patients.

1.1.4 NADOWLI DISTRICT MUTUAL HEALTH INSURANCE SCHEME

The Nadowli District Mutual Health Insurance Scheme started operations in the year 2005 with the registration of their members. Members started benefits in the year 2006 with a total of 135 claims made by various contracted providers of the scheme amounting to GH¢ 467.00. This was in March 2006 when their total membership of the scheme stood at 2,452. According to Mr. John Bosco Zury, Upper West Regional Manager of the NHIA in his third quarter 2011 assessment of the NHIS in the region, the total active membership of the Nadowli DMHIS stands at 40,318 accounting for 41.5% of the total population in the District. Coverage, he stated stood at over 90%, that is nine out of every ten members of the population in the district have registered with the scheme at one point or the other. He stated that Children under 18 years of age make up more than 50% of total membership of the scheme.

The scheme is govern by a board, which until 2009 were members of various stake holders in the district comprising chiefs, healthcare providers and community members. Now the scheme is governed by a four member caretaker committee comprising the District Coordinator Director, the District Finance Officer, a Representative from the NHIA regional office and the Scheme Manager who doubles as the secretary to the committee. Management of the scheme like all other DMHIS is in the hands of six(6) core staff of the scheme; Scheme Manager, Accountant, Public Relations Manager, Management Information Manager, Claims Manager and a Data Entry Clerk. Aside these, the scheme contract a lot of casual staff to run its operations. Currently, seven casual and temporary staffs are with the scheme. Following is the organogram of a typical District Mutual Health Insurance Scheme (DMHIS).

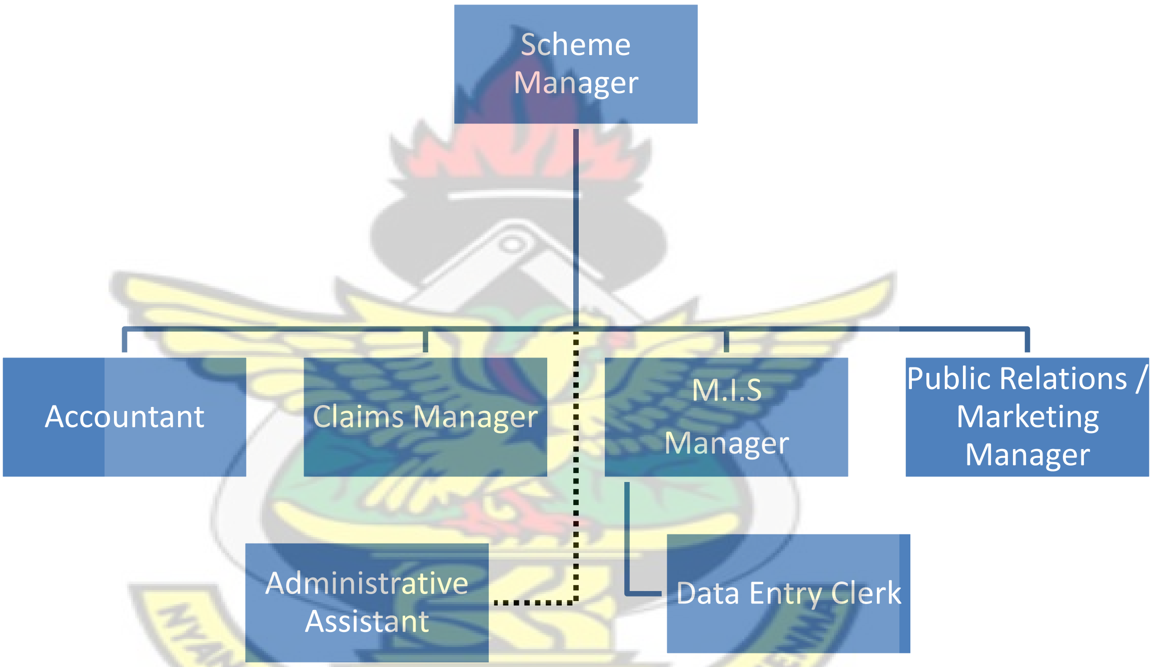


Figure 1.4 Organogram of a typical DMHIS

The salaries of the core staff are paid from the National Health Insurance Authority. The scheme raises money from processing fees and reactivation fees from which the other staff are paid from. Most of the funding of the scheme comes from the NHIA in the form of subsidies, reinsurance and administrative support.

The Nadowli DMHIS operates with the Tarriff list and the Medicines list developed by the NHIA. It contracts providers accredited by the NHIA to provide health services to their members based on the tariff and medicines list. The scheme then reimburses the providers monthly after the providers submit their claims to the scheme for vetting and verification. Over 95% of all claims submitted are Out Patients Department (OPD) cases with malaria accounting for a junk of the claims. In-Patients Department / Admissions (IPD) account consistently for less than 5% of total utilization in the districts. The IPD cases are mostly minor surgeries and maternity cases.

1.2 STUDY AREA (NADOWLI DISTRICT)

Nadowli district is centrally located in the Upper West region of Ghana. It lies between latitude 11’ 30’ and 10’ 20’ north and longitude 3’ 10’ and 2’10’ west. It is bordered to the south by Wa Municipality, west by Burkina Faso, north by Jirapa district and to the east by the Sissala East district. It covers a total land area of 2,742.50km^2 and extends from the Billi Bridge (4km from Wa) to the Dapuori Bridge (almost 12km from Jirapa) on the main Wa – Jirapa-Hamile road. From West to East, it extends from the Black Volta to Wahabu.

The District population is projected at 97,227 for 2010 from the 2000 population and housing census. Only twelve (12) out of the 158 settlements have populations above 2000. While about 45% of the population is aged between 0-14 years, the economically active population also constitutes 49% with the remaining 6% being the aged. The district has 2 major tribes, the Dagaaba and the Sissalas. The Dagaaba constitute 96% of the total population and the Sissalas represent 4%.. There are also three religious groups in the district including Christians (59%), Moslems (18%) and Traditional worshipers (23%).

The district depicts a typical rural economy dominated by the agriculture sector with the commerce and industrial sectors least developed. Agriculture alone accounts for about 85% of the labour force while commerce/service and industry account for 14% and 1% respectively.The Nadowli District has a total of 137 educational institutions comprising fifteen (15) Day Nurseries, Seventy (70) Primary, thirty-five (35) Junior Secondary Schools, five (5) Technical/ Vocational Institutes and three (3) Senior Secondary Schools. The District also has a total of 30 health facilities comprising 2 hospitals, 13 Health Centres and 15 Community clinics. Following is the Disease Prevalence in the District 2006 – 2008.

Table 1. 2 TOP CAUSES OF OPD ATTENDANCE 2006 - 2008

Condition	2006		Conditions	2007		Conditions	2008	
	No	%		No	%		No	%
Malaria	19700	62.5	Malaria	24555	47.2	Malaria	33849	48.2
ARI	3427	10.9	Other ARI(Acute	4579	8.8	ARI	6619	9.4
Skin Diseases	2707	8.6	Skin Diseases & Ulcers	2677	5.1	Skin Diseases	2938	4.2
Malaria in Pregnancy	1096	3.5	Diarrhoea Diseases	1443	2.8	Diarrhoea Diseases	2545	3.6
Diarrhoea Diseases.	1018	3.2	Pneumonia	1206	2.3	Acute Eye Infection	1268	1.8
Pneumoni a	982	3.1	Acute Eye infection	963	1.8	Home/Occupa tional Accidents	957	1.3
Acute Eye Infection	766	2.4	Home/Occupa tional	859	1.6	Rheumatism & joint pains	843	1.2
Hypertens ion	680	2.2	Malaria in Pregnancy	822	1.6	Acute Ear infection	809	1.2
Home & Occ. Accidents	640	2.0	Hypertension	788	1.5	Hypertension	790	1.1
Intestinal Worms	524	1.7	Acute Ear infection	619	1.2	Intestinal worms/parasit es	755	1.1
Total	43119	100		38,578	73.9		51,373	73.5

Source: Nadowli DHMT

1.3 PROBLEM STATEMENT

The health of a people is the wealth of any nation. It is universally accepted that investment in the healthcare of the people of any country is paramount in alleviating poverty and accelerating development. While many other economic, political and social advances must also be made, healthcare investment is the most critical. Poverty causes illness and illness results in more poverty, this vicious circle will spiral any community of people into deprivation.

Healthcare is a social problem and the solution to it lies outside individuals and their immediate environment. It is important for government's continued intervention through thoughtfully guided and collective actions. Health Insurance is one of such collective actions which advance the social welfare of any nation if properly managed and well resourced.

Over the last few years, the implementation of Ghana's Health Insurance has witness a tremendous increase in the utilization of health services by clients as a result of the high number of people who have registered to get treated free of charge for ailments and conditions that would have cost them millions of cedis under the cash and carry system. The utilization of healthcare services through the NHIS has become the order of the day because of its extended benefit package and low premiums. The open nature of the benefit has made it possible for subscribers to enjoy limitless attendance. It is therefore important to undertake an objective study on the recorded figures of the utilization, model the pattern and use the model to make future predictions for planning purposes by the management of the scheme.

1.4 OBJECTIVES OF THE STUDY

This research is undertaken to:

1. Model healthcare services utilization by subscribers of Nadowli DMHIS using ARIMA

2. Forecast future utilization based on the ARIMA model and interprete the results.

1.5 METHODOLOGY

Data on the monthly healthcare services utilization through the Nadowli DMHIS will be collated from January 2006 to June 2012 covering a period of six years six months. Time series analysis will then be done by using SPSS statistical package to select the best ARIMA model that will be used in predicting future utilization values for the scheme. Search on the internet will be used to obtain the related literature. Books from the main Library at KNUST and the Mathematics Department's library will be read in the course of the project.

1.6 JUSTIFICATION OF STUDY

The research is intended to collate the utilization figures of the Nadowli DMHIS from January 2006 to June 2012 and it is intended to serve the following purpose:

It will help management of the Nadowli DMHIS to objectively plan ahead of schedule in making decisions that concern their utilization; attendance of the registered clients at the various accredited healthcare centers for their healthcare needs and the expected claims to be paid. These will then serve as guideline for the evaluation of healthcare policies of the central Government.

1.7 LIMITATIONS / SCOPE

This study covers the operations of National health insurance at the Nadowli District Mutual Health Insurance Scheme.

A research of this nature cannot be carried out without some limitations. Because of Time and financial constraint, the study will take a critical look at only the utilization of members at the Nadowli District Mutual Health Insurance Scheme (DMHIS) as one of the schemes in the Upper West Region. As a result of this, the findings from this study may not apply to other Mutual Health Insurance Schemes as the Nadowli District Mutual Health Insurance Scheme may have its peculiar characteristics.

1.8 STRUCTURE / ORGANISATION OF THESIS

For the sake of convenience, this study has been arranged into five chapters. The first chapter constitutes the background of the study, statement of the problem, Aims and objectives of the study, significance of the study area, limitations of the study area and the organization of the study.

Chapter 2 covers literature review on the theoretical and empirical works of previous researches. Chapter 3 covers the research methodology which deals with the methods the research will use in achieving its target objective.

The presentations, interpretations and analysis of the data from the scheme are discussed in chapter 4.

Chapter 5, which is the conclusion part, contains summary of salient points, conclusion of the research work, suggestions and recommendations of the research work.

CHAPTER 2

LITERATURE REVIEW

2.0 INTRODUCTION

This chapter covers literature review on the theoretical and empirical works of previous researches.

According to Arsham (accessed 2011, December 13), the ability to model and perform decision modeling and analysis is an essential feature of many real-world applications ranging from emergency medical treatment in intensive care units to military command and control systems. Almost all managerial decisions involve forecasting. It is a continuous processing in every organization; the impact of the forecasts on actual performance is measured as time progresses; original forecasts are updated; and decisions are modified, and so on. There are various forms and methods of forecasting including Regression and Time Series Analysis. In this chapter I will review various scholarly works on Time Series Analysis; news papers, journals, books and other internet resources will be applied.

A Time Series is a set of observations collected in time order. Time Series forecasting is the use of a model to predict future values based on previously observed values. The development of Time Series started as far back to the Babylonian era.

Reviewing twenty five years of Time Series, De Gooijer and Hyndman (2006), wrote that early attempts to study time series, particularly in the 19th century, were generally characterized by the idea of a deterministic world. It was the major contribution of Yule (1927) which launched the notion of stochasticity in time series by postulating that every time series can be regarded as the realization of a stochastic process. Based on this simple idea, a number of time series methods have been developed since then. Workers such as Slutsky, Walker, Yaglom, and Yule first formulated the concept of autoregressive (AR) and moving average (MA) models. Wold's decomposition theorem led to the formulation and solution of the linear forecasting problem of Kolmogorow (1941). Since then, a considerable body of literature has appeared in the area of time series, dealing with parameter estimation, identification, model checking and forecasting.

A major revolution in the concept and study of time series however is the publication by Box and Jenkins in 1970; Time Series Analysis: Forecasting and Control.

2.1 APPLICATIONS OF TIME SERIES

Jones (1997) examined the relationship between life history characteristics and environmental predictability for mantled howler monkeys (Alouatta palliate Gray) at Hacienda La Pacifica, Costa Rica. A census with age structure was employed to estimate life history parameters including survivorship, fecundity and mortality. A time series analysis of rainfall at La Pacifica was conducted to test inferences from life history theory whereby variations in mortality across the lifespan are a function of environmental predictability. La Pacifica was found to be a relatively predictable environment, and, consistent with theory, howlers exhibit life history traits expected for their regime. These include low survivorship during more than one age class, iteroparity, a relatively small reproductive effort, a single young per litter, relatively few young across a lifetime, and a relatively long lifespan. The predictable environment of howlers at La Pacifica appears to favor adult over juvenile (including infant) survivorship, and howler life history is consistent with that for other large mammalian herbivores whose females may time reproductive investment to reduce the effects of environmental heterogeneity.

Ashuri and Lu (2010) compared the applicability and predictability of various approaches to forecasting the Construction Cost Index (CCI) which is a weighted aggregate index of the 20-city average prices of construction raw materials and labor published monthly by the professional magazine, Engineering News-Record (ENR). They largely focused around an examination of univariate time series approaches for in-sample and out of sample forecasting of CCI. What they found was that the seasonal Auto-Regressive Integrated Moving-Average (ARIMA) model is the most accurate time series approach for in-sample forecasting of CCI, while the Holt-Winters Exponential Smoothing (Holt-Winters ES) model is the most accurate time series approach for out-of-sample forecasting of CCI. Interestingly, they also found that several of the time series models provide more accurate out-of-sample forecasts than those produced by ENR's own CCI forecasters. Uncertainty about the future trends in construction costs is a significant risk factor in cost estimating and budgeting. They found that the development of a more informed prediction model using time series analysis can help reduce uncertainty about future construction cost.

Jensen (1999) noting that Time Series analysis of population abundance is not based on assumptions about the dynamics of populations, but that sometimes the results can be interpreted biologically or even difficult to interpret, compared a time series analysis of walleye field data with a time series analysis of simulated data from a population dynamics model. His aim was to better understand the results of a time series analysis of a walleye fish population, as it related to the walleye's life history. In the simulations, the nature of the time lags could be identified by changing model parameters. The simulations indicated that a partial auto-correlation coefficient (PAC) at lag 1 would result from density dependence, that a PAC at lag 5 would result from the time required for maturation, and that a negative sign at lag 5 would result from high larval survival.

Wongkoon et al (2007) conducted a study to predict Dengue Haemorrhagic Fever (DHF) incidence in Northern Thailand using Time Series Analysis Technique. This study aimed at developing a forecasting model on the number of DHF incidence in Northern Thailand using time series analysis. They developed Seasonal Autoregressive Integrated Moving Average (SARIMA) models on data collected between 2003 and 2006 and then validated the models using data collected between January-September 2007. The results showed that the regressive forecast curves were consistent with the pattern of actual values. The most suitable model was the SARIMA(2,0,1)(0,2,0)12 model with a Akaike Information Criterion (AIC) of 12.2931 and a Mean Absolute Percent Error (MAPE) of 8.91713. The SARIMA(2,0,1)(0,2,0)12 model fitting was adequate for the data with the Portmanteau statistic Q20 = 8.98644 (=27.5871, P>0.05). This indicated that there was no significant autocorrelation between residuals at different lag times in the SARIMA(2,0,1)(0,2,0)12 model.

Webster (2007) seek to find out if the term "evolutionary psychology" was supplanting "sociobiology" in the scientific literature. He specifically wanted to find out how influential E. O. Wilson's (1975) book, Sociobiology, is in establishing the discipline of the same name and similarly, how influential were the two Tooby-Cosmides chapters appearing in The Adapted Mind (Cosmides and Tooby, 1992; Tooby and Cosmides, 1992) are in establishing evolutionary psychology as a viable outgrowth of sociobiology. He carried out his work using quantitative analyses of publication trends. The Internet search engine Google Scholar was used to count the number of hits (i.e., the number of scholarly works, citations, etc.) for “sociobiology” and “evolutionary psychology” separately per year from 1960 to 2003. Interrupted time-series analyses revealed significant increases (intercept shifts) for sociobiology hits between 1974 and 1975, and for evolutionary psychology hits between 1991 and 1992. Evolutionary psychology hits also experienced a significant increase in change-over-time (a slope shift) between 1991 and 1992. Growth curve analyses revealed that the rate of growth for evolutionary psychology, which was accelerating over time, was significantly greater than that for sociobiology, which was decelerating.

Tulone and Madden (accessed 2011 December 15) presented a method for approximating the values of sensors in a wireless sensor network based on time series forecasting. More specifically, their approach relies on autoregressive models built at each sensor to predict local readings. Nodes transmit these local models to a sink node, which uses them to predict sensor values without directly communicating with sensors. When needed, nodes send information about outlier readings and model updates to the sink. They showed that this approach can dramatically reduce the amount of communication required to monitor the readings of all sensors in a network, and demonstrate that their approach provides provably-correct, user-controllable error bounds on the predicted values of each sensor.

Zhang and Kline (2007) wrote that forecasting of time series that have seasonal and other variations remains an important problem for forecasters. In this paper they presented a neural network (NN) approach to forecasting quarterly time series. With a large data set of 756 quarterly time series from the M3 forecasting competition, they conducted a comprehensive investigation of the effectiveness of several data preprocessing and modeling approaches. They considered two data preprocessing methods and 48 NN models with different possible combinations of lagged observations, seasonal dummy variables, trigonometric variables, and time index as inputs to the NN. Both parametric and nonparametric statistical analyses were performed to identify the best models under different circumstances and categorized similar models. Results indicate that simpler models, in general, outperform more complex models. In addition, data preprocessing especially with deseasonalization and detrending is very helpful in improving NN performance. Practical guidelines were also provided.

Brisson et al (access 2011, December 15) observed that, the growth rates of real output and real investment are two macroeconomic time series which are particularly difficult to forecast. Their paper considered the application of diffusion index forecasting models to solve this problem. They began by characterizing the performance of standard forecasts, via recently-introduced measures of predictability and the forecast content, noting the maximum horizon at which the forecasts have value. They then compared diffusion index forecasts with a variety of alternatives, including the forecasts made by the OECD and found gains in forecast accuracy at short horizons from the diffusion index models, but did not find evidence that the maximum horizon for forecasts can be extended in this way.

Mandal (2005) forecasted sugarcane production for three leading years. Yearly sugarcane production data for the period of 1950/51 to 2002/03 of India were analyzed by time-series methods. Autocorrelation and partial autocorrelation functions were calculated for the data. Appropriate Box-Jenkins autoregressive integrated moving average model was fitted. Validity of the model was tested using standard statistical techniques. The forecasting power of autoregressive integrated moving average model was then used to forecast sugarcane production for three leading years.

Dobre and Alexandru (2008) predicted the unemployment rate of Romania to be 4.06% for the month of January 2008. They modeled the evolution of unemployment rate using the Box-Jenkins methodology during the period 1998-2007. Monthly data was used. The empirical study relieved that the most adequate model for the unemployment rate is ARIMA (2,1,2). Using the model, the forecasts values of unemployment rate for January and February 2008 were determined.

Gilmour et al (2006) wrote that Intervention time series analysis (ITSA) is an important method for analysing the effect of sudden events on time series data. ITSA methods are quasi-experimental in nature and the validity of modelling with these methods depends upon assumptions about the timing of the intervention and the response of the process to it. In their publication, they described how to apply ITSA to analyse the impact of unplanned events on time series when the timing of the event is not accurately known and so the problems of ITSA methods are magnified by uncertainty in the point of onset of the unplanned intervention. They illustrated the methods using the example of the Australian Heroin Shortage of 2001, which provided an opportunity to study the health and social consequences of an abrupt change in heroin availability in an environment of widespread harm reduction measures. Application of these methods enables valuable insights about the consequences of unplanned and poorly identified interventions while minimising the risk of spurious results.

According to Taylor (2008), Predictions of call center arrivals are a key input to staff scheduling models. It is, therefore, surprising that simplistic forecasting methods dominate practice, and that the research literature on forecasting arrivals is so small. He evaluated univariate time series methods for forecasting intraday arrivals for lead times from one half-hour ahead to two weeks ahead. He analyzed five series of intraday arrivals for call centers operated by a retail bank in the UK. A notable feature of these series was the presence of both an intraweek and an intraday seasonal cycle. The methods considered include seasonal ARIMA modeling; periodic AR modeling; an extension of Holt-Winters exponential smoothing for the case of two seasonal cycles; robust exponential smoothing based on exponentially weighted least absolute deviations regression; and dynamic harmonic regression, which is a form of unobserved component state space modeling. The results indicate strong potential for the use of seasonal ARIMA modeling and the extension of Holt-Winters for predicting up to about two to three days ahead and that, for longer lead times, a simplistic historical average is difficult to beat. He found a similar ranking of methods for call center data from an Israeli bank.

Makridakis and Hibon, (accessed 2012 January 4) conducted a study of the Box-Jenkins methodology to ARIMA models and determined the reasons why in empirical test it is found that the post-sample forecasting accuracy of such models is worse than much simpler time series methods. They found out that the major problem is the way of making the series stationary in its mean (i.e., the method of differencing) that has been proposed by Box and Jenkins. They also demonstrated that if alternative approaches are utilized to remove and extrapolate the trend in the data, ARMA models outperform the corresponding methods involved in the great majority of cases. In addition it is shown that using ARMA models to seasonally adjusted data slightly improves post-sample accuracies while simplifying the use of ARMA models. The study also confirmed that transformations slightly improve post-sample forecasting accuracy, particularly for long forecasting horizons. Finally, it is demonstrated that AR(1) and AR(2), or their combination, produce as accurate post sample results as those found through the application of the Box-Jenkins methodology.

El-Mefleh and Shotar (accessed 2012 January 6) developed an autoregressive integrated moving-average (ARIMA) forecasting models for key Qatari macroeconomic variables. They developed, stimulated and then used the model for ex-post and ex-ante forecasts. They used the Percentage Forecast Inaccuracy (PFI) method to measure the precision of the models (forecast) and realized that the ARIMA models;

i. appear highly successful in forecasting government consumption and final consumption;
ii. are highly successful in forecasting four years out of six years of private consumption;
iii. are moderately successful in forecasting three years out of six years of GDP;
iv. are poor in forecasting (under forecast) capital formation, imports, and exports due to substantial increase in the price of oil;

v. increase their forecasts inaccuracy as the time span for forecast increased from one year to two years to three years.

Contreras et al (2003) wrote that Price forecasting is becoming increasingly relevant to producers and consumers in the new competitive electric power markets. Both for spot markets and long-term contracts, price forecasts are necessary to develop bidding strategies or negotiation skills in order to maximize benefit. Using the ARIMA methodology, they provided a method to predict next-day electricity prices. They presented a detailed explanation of the ARIMA techniques in Time Series analysis and proposed two ARIMA models to predict hourly prices in the electricity markets of Spain and California. The Spanish model needs 5 hours to predict future prices, while the California model requires 2 hours for prediction. These differences, they concluded, may reflect different bidding structures and ownership.

Georgakarakos et al (2006) used Time series analysis techniques (ARIMA models), artificial neural networks (ANNs) and Bayesian dynamic models to forecast annual loliginid and ommastrephid landings recorded from the most important fishing ports in the Northern Aegean Sea (1984–1999). The techniques were compared based on their efficiency in forecasting and their ability to utilize auxiliary environmental information. Applying a "stepwise modelling" technique, namely by adding stepwise predictors and comparing the quality of fit, certain inferences concerning the importance of the predictors were made. They observed that the ARIMA models predicted the test data very precisely (high R^2), especially if the target time series contained a strong autoregressive character, after they were first differenced to obtain stationarity ($R^2 > 0.96$). The influence of temperature on catches was mainly investigated by applying neural models, which predicted the monthly landings with high precision (R^2 = 0.89), even when incorporating in the model exclusively monthly sea surface temperature (SST) descriptors. Similarly, ANN models of annual landings containing monthly mean temperatures provided high precision (R^2 = 0.87) and valuable inference concerning the possible effect of the SST in certain months. Bayesian dynamic models also provided a high precision (R^2 = 0.96). They combined the information of both environmental and landing time series, namely the monthly mean temperatures and the monthly seasonality of the landings. The impact factors estimated from the model have the form of time series representing the temperature effect. The results reveal that both the monthly and the annual landings can be predicted and that the Bayesian model is the best performer overall, characterised by a higher number of stable forecasts, and forecasts with higher precision and accuracy, than the other methods. Evident from the application of the "stepwise modeling" technique, the researchers concluded that the incorporation of temperature descriptors can significantly improve the model performance.

Nogales and Conejo (2005) wrote that Forecasting electricity prices in present day competitive electricity markets is a must for both producers and consumers because both need price estimates to develop their respective market bidding strategies. This paper analyzed exhaustively a transfer function model to forecast day-ahead electricity prices based on both past electricity prices and demand. A model is built based on simulations conducted on data from PJB Interconnection for the year 2003. Exhaustive analysis using the data from the PJM Interconnection reveals an appropriate forecasting functioning of the technique pro-posed. The proposed model is compared with naïve and other techniques to assess its effectiveness. Their conclusion indicates that, the use of electricity demand series as an explicative variable improve predictions but not in a drastic manner and also from a detailed analysis of the numerical results available in the technical literature, they concluded that the quality of predictions using the proposed technique is generally superior to the quality of predictions using alternative procedure such as standard time-series models (ARIMA) or neural networks.

Smartdrill (2010) forecasting catalog sales of men's clothing using ARIMA time series analysis demonstrated how to build a seasonal ARIMA model using the autocorrelation and partial autocorrelation functions to identify the ARIMA orders. A number of candidate predictor variables were added to the model and evaluated based on their statistical significance. The final model, keeping only significant predictors, was compared to the model with no predictors. They concluded that the model with predictors did a better job of explaining the variance of the data.

CHAPTER 3

METHODOLOGY

3.0. INTRODUCTION

Forecasting has always been an important planning and control tool for most profit maximizing firms and agencies. Often, the financial well being or sustainability of an entire business operation may rely on the accuracy of the forecast since such information will likely be used to make interrelated budgetary and operative decisions in areas of personnel management, marketing and advertisement, purchasing, capital financing, investment among other things. The method of forecasting is as important as the forecast itself because any significant over or under forecast will burden the business with inventory load or create losses.

There are two main approaches to forecasting. Either the estimates of the future value of a variable is based on an analysis of factors which are believed to influence new observations of the variable (explanatory method) or the prediction is based on an inferred study of past general behavior of the variable over time (extrapolation method). For example, the belief that the yield of a farm will increase (decrease) because of the application of a particular fertilizer rather than the season of cultivation illustrates the difference between these two philosophies. It is possible that both approaches will lead to the creation of accurate and useful forecast but it must be remembered that even for a modest degree of desired accuracy, the former is often more difficult to implement and validate than the latter approach. In this chapter, our discussion focuses on the extrapolation or time series approach to forecasting.

3.1. TIME SERIES THEORY

Time Series data refers to observations on a variable that occurs in a time sequence. Mostly these observations are collected at equally spaced, discrete time intervals. When there is only one variable upon which observations are made then we call them a single time series or a univariate time series. One defining characteristic of time series is that this is a list of observations where the ordering matters. Ordering is very important because there is dependency and changing the order could change the meaning of the data. Unlike in standard linear regression, the data are not necessarily independent and not necessarily identically distributed. In analyzing time series it is important to consider the pattern of the data so that an appropriate model can be selected and used.

3.1.1. TIME SERIES COMPONENTS AND DECOMPOSITION

Every time series data consist of one or more distinguishable data patterns that can be decomposed. In general, in a given time series the following components can be recognized and separated:

i. Trend – Here on the average the data values of the time series tend to increase (decrease) over time. This component may also be referred to as a horizontal component when the values fluctuate around a constant value

ii. Seasonal – These are components / data that exhibit recurrent patterns that are related to calendar time. The time could be a season, month, quarter, a particular day of the week or time of a day.

iii. Cyclical – This refers to recurring up and down movement around trend levels that are not periodic / related to seasonality. Because there is no single explanation for cyclical fluctuations, they vary significantly in both length and magnitude.

iv. Irregular / Random / Error – These are erratic movements in a time series that follow no recognizable or regular patterns. These values remain unidentified after separating a time series into its components form i.e. trend, season and cyclical.

It is when these components are identified that an appropriate technique or method is selected for the analysis. This involves careful scrutiny of the recorded data plotted over time to help visualized the patterns while exploring the data.

3.2.MOVING AVERAGES AND EXPONENTIAL SMOOTHING METHODS

3.2.1. SIMPLE MOVING AVERAGES

The Moving Averages (MA) methods are the best –known methods of forecasting. This involves taking the sum of a set of numbers of specific periods and dividing this sum by the number of periods. The Simple moving averages approach is most effective and efficient for a data set that is stationary both in mean and in variance. In general, the moving average at time t, taken over N periods, is give by:

$$M_t^{[1]} = \frac{X_t + X_{t-1} + \cdots + X_{t-N-1}}{N},$$

where X_t is the observed response at time t. At each successive time period the most recent is included and the farthest observation is excluded for computing the average. Thus the name 'moving' averages. We have prior MA, weighted MA, centered MA etc as examples of simple moving averages

3.2.2. DOUBLE MOVING AVERAGES

Double moving averages is simply treating the moving averages $M_t^{[1]}$ over time as individual data points and obtaining a moving average of moving averages. Unlike the simple moving averages that is intended for data of a constant nature and no trend, the double moving averages is suitable for data with a linear or quadratic trend. The double moving average is an improvement of the simple moving average to correct for the biases and develop a better forecasting equation.

3.2.3. SIMPLE EXPONENTIAL SMOOTHING (SES)

Exponential smoothing technique is one of the most successful forecasting methods. Its added advantage lies in the fact that it can be modified efficiently and use effectively to cater for time series data with seasonal patterns. It is an averaging technique that uses unequal weights; however the weights applied to past observations decline in an exponential manner. The basic technique is known as the simple (single) exponential smoothing (SES).

Let the time series data be denoted by $X_1, X_2,\ldots,X_t$. Suppose we wish to forecast the next value of our time series X_{t+1} that is yet to be observed with forecast for X_t denoted by F_t. Then the forecast F_{t+1} is based on weighting the most recent observation X_t with a weight value α and weighting the most recent forecast F_t with a weight of (1-α) where α is a smoothing constant (weight) which lies between 0 and 1. Thus the forecast for the period t+1 is given by

$$F_{t+1} = F_t + \propto (X_t - F_t)$$

A small α provides a detectable and visible smoothing, while a large α, provides a fast response to the recent changes in the time series but provides a smaller amount of smoothing. An exponential smoothing over an already smoothened time series is referred to as Double Exponential Smoothing. When smoothing is applied three times, it is known as a Triple Exponential Smoothing.

3.2.4. DOUBLE EXPONENTIAL SMOOTHING (Holt)

Double exponential smoothing also known as the Holt's linear exponential smoothing applies the process above to account for trends. The forecast is found by having two more equations to the SES model to deal with; for level and for trend. The smoothing weights α and β are each selected from the range (0.1, 0.2,…, 0.9), and then the combination of α and β that corresponds to the lowest Mean Square Error (MSE) is selected.

3.2.5. TRIPPLE EXPONENTIAL SMOOTHING (Winters)

Triple Exponential Smoothing or Winter's exponential smoothing is recommended when seasonality exist in the time series data. It is based on three smoothing equations, one for level, one for trend and the last for seasonality. It is basically similar to the Holt's method with one more equation to deal with seasonality. There are two different Winter's methods depending on whether seasonality is modeled in an additive or multiplicative way.

3.3.CORRELOGRAMS

3.3.1. AUTOCORRELATION FUNCTIONS (ACF)

Autocorrelation refers to the way the observations in a time series are related to each other. It is the serial correlation of equally spaced time series between its members one or more lags apart. It is also known as the lagged correlation, and persistence. Thus the autocorrelation for a given series $\mathbf{X_t}$ at lag **p** is the correlation between the present $\mathbf{X_t}$ and its value **p** periods before**,** $\mathbf{X_{t\text{-}p}}$, and is given by

$$\Gamma_{p=\text{correlation}(X_t,X_{t-p})=\frac{\sum_{t=1}^{n-p}(X_t-\bar{X})(X_{t-p}-\bar{X})}{\sum_{t=1}^{n}(X_t-\bar{X})^2}}$$

Unlike the statistical data which are random samples allowing us to perform statistical analysis, the time series are strongly autocorrelated, making it possible to predict and forecast. Three tools for assessing the autocorrelation of a time series are the time series plot, the lagged scatter plot, and at least the first and second order autocorrelation values. Autocorrelation ranges from -1 to +1 and, Box and Jenkins has suggested that maximum number of useful autocorrelations are roughly n/4, where n is the number of periods upon which information on $\mathbf{X_t}$ is available.

3.3.2. PARTIAL AUTOCORRELATIONS FUNCTIONS (PACF)

Partial autocorrelations are used to measure the degree of association between X_t and X_{t-p} when the X –effects at other time lags 1, 2, 3,…., p-1 are removed. In other words the partial autocorrelation is similar to autocorrelation, except that when calculating it, the (auto) correlation with all the elements within the lag is partialled out (Box and Jenkins, 1976). If a lag of one is specified; where there are no intermediate elements within the lag, then the partial autocorrelation is equivalent to autocorrelation.

3.4.AUTOREGRESSIVE MODELS (AR)

One of the simplest time series models is the autoregressive (AR) models. This is a model in which we use a linear model to predict the value of a variable using its value at a previous time. The number of time we move back to predict the present value is known as the order of the model or lag value of the series. For example if $\mathbf{X_t}$ denote the value of a series at time **t**, then $\mathbf{X_{t-1}}$ denotes the value of the series one time before **t**. $\mathbf{X_{t-1}}$ is described as the lag 1 value of $\mathbf{X_t}$ and the model called an AR(1); autoregressive model of order one. Usually, the plot of a series against its lag values indicates if an AR model will be a useful model.

Theoretically, the AR model is written as:

$$\mathbf{X_t = \delta + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + w_t}$$

Where, $\mathbf{w_t}$ is the random error term which are identical and independent normal distributions with mean zero(0) and variance, $\boldsymbol{\sigma_w^2}$. $\boldsymbol{\varphi_1, \varphi_2, \varphi_3 \ldots \varphi_p}$ are the autoregressive parameters, $\boldsymbol{\delta}$ is a constant term. For example an AR(1) is written as

$$X_t = \delta + \varphi_1 X_{t-1} + w_t$$

3.5.MOVING AVERAGE MODELS (MA)

A moving average term in a time series model is a past error multiplied by a coefficient. Theoretically, the qth order moving average model is denoted by MA(q) and written as:

$$X_t = \mu + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q} ,$$

Where, w_t,..., w_{t-q} are the white noise error terms and they are identically, independently distributed, each with a normal distribution with mean 0 and a constant variance. μ is the mean of the series and θ_1,…..θ_q are the parameters of the model. For instance a first order moving average model will be specified as MA(1):

$$X_t = \mu + w_t + \theta_1 w_{t-1}$$

The second order moving average model, denoted by MA(2) is:

$$X_t = \mu + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2}$$

3.5.1. INVERTIBILITY OF MA MODELS

An MA model is said to be invertible if it is algebraically equivalent to a converging infinite order AR model. By converging, it means that the AR coefficients decrease to zero (0) as the series move back in time.

3.6.STATIONARITY OF A TIME SERIES PROCESS

A time series is said to be stationary (random) if it has a constant mean, constant variance and constant autocorrelation function (ACF) through time. This means that, if different subsets of a time series sample are considered, the different subsets will typically have means, variances and autocorrelation functions that do not differ significantly. There are several ways to ascertain stationarity of a time series process. The most common is an examination of the graph or time plot of the data; there should be no consistent trend (upward or downward) or seasonality in the time series. Non stationarity in mean is corrected through appropriate differencing of the data and non-stationarity in variance is achieved by some modes of transformation, say, taking the natural log of the series. Other means include examination of autocorrelation and partial autocorrelation functions (ACF and PACF) and the Dickey Fuller test.

3.7.DIFFERENCING

Differencing is a process of computing the difference between every two successive values in a time series. Often differencing is used to account for non-stationarity that occurs in the form of trend or seasonality or both. Differencing that accounts for trend is referred to as regular (non-seasonal) differencing and when it accounts for seasonality it is known as seasonal differencing. Thus if $\mathbf{X_t}$ (t=1,2,3,….,n) denotes the original series, the first order non seasonal difference (regular) is $\mathbf{Y_t=X_t - X_{t-1}}$.

The difference $\mathbf{X_t - X_{t-1}}$ can be expressed as (1-B)$\mathbf{X_t}$. , where **B** is the backshift operator.

Using B before either a value of the series X_t or an error term W_t means to move that element back one time e.g. $\mathbf{B}X_t=X_{t-1}$

A power of B means to repeatedly apply the backshift in order to move back a number of time periods that equal the "power".Thus $B^2X_t=X_{t-2}$. It must however be noted that the backshift operator is not operated on coefficients because they are fixed/constant quantities that do not move in time. That is $\mathbf{B}Ø_1=Ø_1$.

The alternative notation for a difference is: $\nabla = 1 - \mathrm{B}$

Thus $\nabla \mathrm{X_t} = (1 - \mathrm{B})\mathrm{X_t} = \mathrm{X_t} - \mathrm{X_{t-1}}$

For data that exhibits seasonality; where the differences from the previous year may be on the average, about the same each month of a year, a subscript is used to define a difference of lag equal to the subscript. That is

$$\nabla_s \mathrm{X_t} = \mathrm{X_t} - \mathrm{X_{t-s}}$$

where **s** is the seasonal lag (s=12 for monthly data, s=4 for quarterly measurements). A superscript is used to repeat the differencing a specified number of times. As an example

$$\nabla^2 \mathrm{X_t} = (1 - \mathrm{B})^2 \mathrm{X_t} = (1 - 2\mathrm{B} + \mathrm{B}^2)\mathrm{X_t} = \mathrm{X_t} - 2\mathrm{X_{t-1}} + \mathrm{X_{t-2}}$$

3.8.MIXED MODELS (ARMA)

As discussed above, a model with only autoregressive terms is referred to as an AR(p) model and one with only moving average terms is referred to as an MA(q) model. However some models contain both terms and these models are termed ARMA(p,q) models, for autoregressive moving average model. This is the case when there is no differencing. When differencing is performed the mixed model is referred to as an autoregressive integrated moving average (ARIMA) model and denoted as ARIMA(p,d,q), where **p** is the order of autoregression, **d** is the level of differencing and **q** is the order of the moving average.

3.9.AUTOREGRESSIVE INTEGRATED MOVING MODELS (ARIMA)

Generally, an ARIMA model is characterized by the notation ARIMA(p,d,q) where p,d and q denote orders of auto-regression, integration(differencing) and moving average respectively. In ARIMA, time series is a linear function of past actual values and random shocks. For instance, if $\mathbf{X_t}$ is a time series process, the first order AR(1) and MA(1) are given as follows:

ARIMA(1,0,0) or AR(1) : $X_t = \delta + \varphi_1 X_{t-1} + w_t$

ARIMA(0,0,1) or MA(1) : $X_t = \mu + w_t + \theta_1 w_{t-1}$

Alternatively, the model may be a mixture of these processes and of higher orders and can be stated as

$$\mathbf{X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q} + \varepsilon_t}$$

Where $\boldsymbol{\varepsilon_t}$ ‘s are independently and normally distributed with zero mean and constant variance σ^2 for t=1,2,…,n. In practice, the values of **p** and **q** lie between 0 and 3.

3.9.1. NON SEASONAL ARIMA MODELS

ARIMA (Box-Jenkins) models are models that possibly may include autoregressive (AR) terms, moving average (MA) terms and differencing (integration) operations. It may be specified as:

AR(p) or ARIMA(p,0,0) – a model with only ‘p’ AR terms e.g. when p=1; AR(1) or ARIMA(1,0,0)

MA(q) or ARIMA(0,0,q) – a model with only ‘q’ MA terms e.g. if q=2; MA(2) or ARIMA(0,0,2)

ARMA(p,q) – a model with ‘p’ AR terms and ‘q’ MA terms with no differencing involved e.g. ARMA(1,1) for one AR term and one MA term.

When differencing is required in the model it is specified as ARIMA(p,d,q), where the ‘d’ refers to the order of differencing. For instance ARIMA(1,1,1) is a model with one

AR term and one MA term being applied to the variable $Z_t = X_t - X_{t-1}$ i.e. a first difference. A first difference might be used to account for a linear trend in a data set. The differencing order refers to the successive first differences. For instance, if the order of differencing is d=2, it implies the analysed variable is $Z_t=(X_t-X_{t-1})-(X_{t-1}-X_{t-2})$ i.e. a first difference of first differences.

To identify a possible model, a time series plot of the observed data series, the autocorrelation function (ACF) and partial autocorrelation function (PACF) are examined to guess the orders of the various terms in an ARIMA. The things to look out for are possible trend, seasonality, outliers, constant variance and non constant variance.

No particular model is spotted by examining the time series plot of the observed data, but this gives an idea as to the need for various possible actions. For example, if there is a random walk i.e. an obvious linear trend (upward or downward) a first difference may be needed. A differencing order of two may be needed for a quadratic trend. Rarely is a differencing order beyond two (required), if that be the case then techniques like smoothing may be considered instead as over differencing may introduce unnecessary levels of dependency. A natural logarithm or a square root transformation is recommended if there is a curved upward trend accompanied by increasing variances in the data as the series progresses into time.

The autocorrelation functions (ACFs) and partial autocorrelation functions (PACFs) combined gives an overall nature of the model. This requires a lot of experience and experimentation (guesses) but the following general guidelines can be applied in identifying the various terms in an ARIMA model.

3.9.2. IDENTIFYING ORDER OF DIFFERENCING, AR AND MA TERMS

- If a series has positive autocorrelations out to a high number of lags, then it may need a higher order of differencing.
- If the lag 1 autocorrelation is zero or negative, the series does not need a higher level of differencing. The same is the case if the autocorrelations are all small and with no pattern.
- A stationary series requires no differencing and a model with no order of differencing often include a constant term which represents the mean of the series.
- If a first difference of the series yields only non significant autocorrelations then the series is called a random walk model i.e. $X_t=\delta+X_{t-1}+W_t$ (the data are dependent and are not identically distributed; increasing mean and variance through time)
- If the PACF of the differenced series displays a sharp cutoff and or the lag 1 autocorrelation is positive, then consider adding one or more AR terms. The lag beyond which the PACF cuts off is indicative of the number of AR terms.
- If the ACF of the differenced series displays a sharp cutoff and or the lag 1 autocorrelation is negative, then the addition of an MA term should be considered. The lag beyond which the ACF cuts off is the indicated number of MA terms.
- It is possible for an AR term and an MA term to cancel each other's effect and as such if a mixed model, ARMA, seem to fit the data, try a model with one fewer AR term and one fewer MA term and a combination of such manipulations.

3.9.3. ESTIMATING AND DIAGNOSING A POSSIBLE MODEL

Once a guess of the possible models is made, computer software such are SPSS, SAS, R or Minitab is used to estimate the coefficients. Most software uses maximum likelihood estimation methods to make the estimates. The following are ensured in the estimation:

- Significance of coefficients
- ACF of the residuals: for a good model, all autocorrelations for the residuals series should be non-significant. If this is not the case, the chosen model should be revised.
- Box-Pierce (Ljung) test: this is applied to the residuals from the model fit to determine whether residuals are random. Randomness of residuals indicates that the model provides an adequate fit to the data series. The value of this statistic must be non-significant.
- Plot of residuals verses fit / time series plot of residuals: a standardized residuals plot mostly indicates that there is no trend in the residuals, no outliers and in general no changing variance across time

If any of the conditions stated above is violated, the chosen model is revised by reexamining the ACF and PACF to move in a different direction. If more than one model looks good, the criteria following is used to decide which one is the best:

- Models with fewest parameters are favoured (parsimony)
- Pick models with generally lowest standard errors for prediction of the future
- Statistics such are MSE, BIC or SIC, and AICs are compared and the model with the lowest of these statistics is more desirable

3.10. SEASONAL ARIMA MODELS

Seasonality in a time series is a regular pattern of changes that repeats over specific time periods. For example every year the pattern of rains in a particular region changes with the months of the year. If S defines the number of time periods until the pattern repeats again, 'S' can be define as S=12 (months per year) or S=4 (quarters per year). It may also be days of the week, weeks of the month and so on. In a seasonal ARIMA model, seasonal AR and MA terms predict X_t using data values and errors at times with lags that are multiples of S (length of season). For example with monthly data (S=12), a seasonal first order AR would use X_{t-12} to predict X_t, and a second order seasonal AR model would use X_{t-12}, and X_{t-24} to predict X_t. Similar a first order seasonal MA model would use the error W_{t-12} as a predictor just as a seasonal MA(2) would use W_{t-12} and W_{t-24} for prediction.

Seasonality usually causes the series to be non stationary because of the seasonal changes in mean. This makes differencing necessary for seasonal data to achieve stationarity. Seasonal differencing is defined as a difference between a value and a value with lag that is a multiple of the seasonal period, S. For instance monthly data (S=12) will have a seasonal difference as $(1 - B^{12})X_t = X_t - X_{t-12}$. The differences from the previous year may be about the same for each month of the year to yield a stationary series. Seasonal differencing removes seasonal trend and can also get rid of seasonal random walk type of non-stationarity.

It must also be noted that when the data series has trend, non seasonal differencing may be applied to "detrend" the data. For this purpose, usually a first non-seasonal difference is enough to attain stationarity. That is $(1 - B)X_t = X_t - X_{t-1}$. When both seasonality and trend are present it may be necessary to apply both a first order non-seasonal and a seasonal difference. In which case the ACF and PACF of the following equation needs to be examined:

$$(1-B^{12})(1-B)X_t = (X_t - X_{t-1}) - (X_{t-12} - X_{t-13})$$

Removal of trend does not imply removal of dependency; the mean part, μ_t, which may include a periodic component, may have been removed. What is actually done is breaking the dependency down to recent things that have happened and long-range things that have happened.

Short run non seasonal behavior may still matter in seasonal data and contribute to a seasonal model. The seasonal ARIMA model incorporates both non-seasonal and seasonal factors in a multiplicative model. The model is written in the following notation:

$$\text{ARIMA}(p, d, q) * (P, D, Q)S$$

Where p, d, q, are the non-seasonal orders AR, differencing and MA respectively. P,D.Q are the seasonal AR order, seasonal differencing order and seasonal MA order respectively. S=time span of repeating seasonal pattern. This model can be written more formally as:

$$\Phi(B^S)\varphi(B)(X_t - \mu) = \Theta(B^S)\theta(B)w_t \qquad \text{(without differencing)} \ldots\ldots\ldots \text{(a)}$$

The non-seasonal components are:

AR: $\varphi(B) = 1 - \varphi_1 B - \cdots - \varphi_p B^p$

MA: $\theta(B) = 1 + \theta_1(B) + \cdots + \theta_q B^q$

The seasonal components are:

Seasonal AR: $\Phi(B^S) = 1 - \Phi_1 B^S - \cdots - \Phi_P B^{PS}$

Seasonal MA: $\Theta(B^S) = 1 + \Theta_1 B^S + \cdots + \Theta_Q B^{QS}$

3.10.1. IDENTIFYING THE SEASONAL COMPONENTS OF THE MODEL

A time series plot of the data will give strong indication of seasonality. Examine the pattern across the time units (months, quarters etc) to confirm if there is indeed seasonality. In addition to the identifying rules for non-seasonal behavior, the following guidelines should be considered in identifying seasonal ARIMA (SARIMA) model parameters.

Differencing may be required to make the series stationary. The general guidelines are:

- If the series has a strong and consistent seasonal pattern, then the use of an order of seasonal differencing is required; but never an order of seasonal differencing more than one or two order of total differencing (seasonal and non seasonal)
- Apply both seasonal and non-seasonal difference as two successive operations when both trend and seasonality is present in the data.
- If there is a linear trend and no obvious seasonality, only a first difference may be required. If the trend is curved, a transformation should be considered before differencing.
- If there is seasonality and no trend, take a difference at lag S (seasonal span or period). Example, 12^{th} difference for monthly data with seasonality.

Also, an examination of the ACF and PACF of the difference data will give indication of the need for seasonal MA and or seasonal AR;

- Spikes in the ACF at low lags indicate non-seasonal MA terms
- Spikes in the PACF at low lags indicate possible non-seasonal AR terms
- If the autocorrelation at the seasonal period is positive, a seasonal AR term in the model is suggested.

- If the autocorrelation at the seasonal period is negative, then a seasonal MA term in the model is suggested.

Mixing seasonal AR and seasonal MA terms in a model is however not recommended and the use of more than one kind of either term should be avoided.

3.11. ARIMA METHODOLOGY / ART OF ARIMA MODEL BUILDING

Autoregressive integrated moving average (ARIMA) modeling, also known as the Box-Jenkins modeling approach, requires significantly large data set and the development of ARIMA model for any variable involves several steps; model identification, model estimation, model verification and then forecasting. A detailed description of the procedure is indicated in figure 3.1 below.

3.11.1. MODEL IDENTIFICATION

ARIMA model is estimated only after transforming the variable under forecasting into a stationary series. As stated earlier, a stationary series is one whose values vary over time only around a constant mean, constant variance and constant autocorrelation. A look at the graph of the data and structure of the autocorrelation and partial correlation coefficients may provide clues for the presence of stationarity. Another way of checking for stationarity is to fit a first order autoregressive AR(1) model for the raw data and test whether the coefficients ‘ $\boldsymbol{\varphi_1}$ ‘ is less than one. If the series is found to be non-stationary, stationarity can be achieved mostly by differencing the series. The number of times the series is difference to achieve stationarity is reflected in the “**d**” parameter. Stationarity in variance is achieved by some form of transformation, say, taking the natural log of the series. This is applicable for both seasonal and non-seasonal stationarity. Thus if ‘X_t’ denotes the original series, the non-seasonal difference of first order is:

$$Y_t = X_t - X_{t-1}$$

and can be followed by a seasonal difference if needed as:

$$Z_t = Y_t - Y_{t-s} = (X_t - X_{t-1}) - (X_{t-s} - X_{t-s-1})$$

In order to determine the necessary level of differencing, the plot of the data and autocorrelograms are examined. Significant changes in level (strong upward or downward changes) usually require first order non-seasonal (lag 1) differencing; strong changes in slop usually require second order non-seasonal differencing. Seasonal patterns require respective seasonal differencing. If the estimated autocorrelation coefficients decline slowly at longer lags, first order differencing is usually needed. It must however be noted that some time series may require little or no differencing, and that over differenced series produces less stable coefficients estimates.

The next step in the identification process is to find the initial values for the orders of seasonal and non-seasonal parameters, **p,q,** and **P**,**Q**. This is where the number of '**p**'s and '**q**'s parameters for non-seasonal and seasonal (if needed) are decided to yield an effective but still parsimonious model of the process. A parsimonious model means that it has the fewest parameters and greatest number of degrees of freedom among all models that fit the data. The parameters are obtained by looking for significant autocorrelation and partial auto correlation coefficients. In practice, the numbers of the parameters very rarely need to be greater than 2. For instance if second order autocorrelation coefficient is significant, then an AR(2) or MA(2) or ARMA(2) model could be tried to start with. This is not a hard and fast rule, as sample autocorrelation coefficients are poor estimates of population autocorrelation coefficients. But they give an indication of the possible models and serve as initial values while the final models are achieved after going through the stages repeatedly; alternating and increasing (decreasing) the AR and MA terms.

3.11.2. MODEL ESTIMATION AND VERIFICATION

The identification stage yields one or more tentative models that provide statistically adequate representation of the available data. The precise estimates of the parameters of the models are obtained by least squares according to Box and Jenkins. Standard computer packages like SPSS, SAS, R etc are available for finding the estimates of the relevant parameters using iterative procedures.

The model verification is concerned with checking the residuals of the model to see if they contain any systematic pattern which still can be removed to improve on the chosen ARIMA. This is done through examining the autocorrelations and partial autocorrelations of the residuals of various orders.

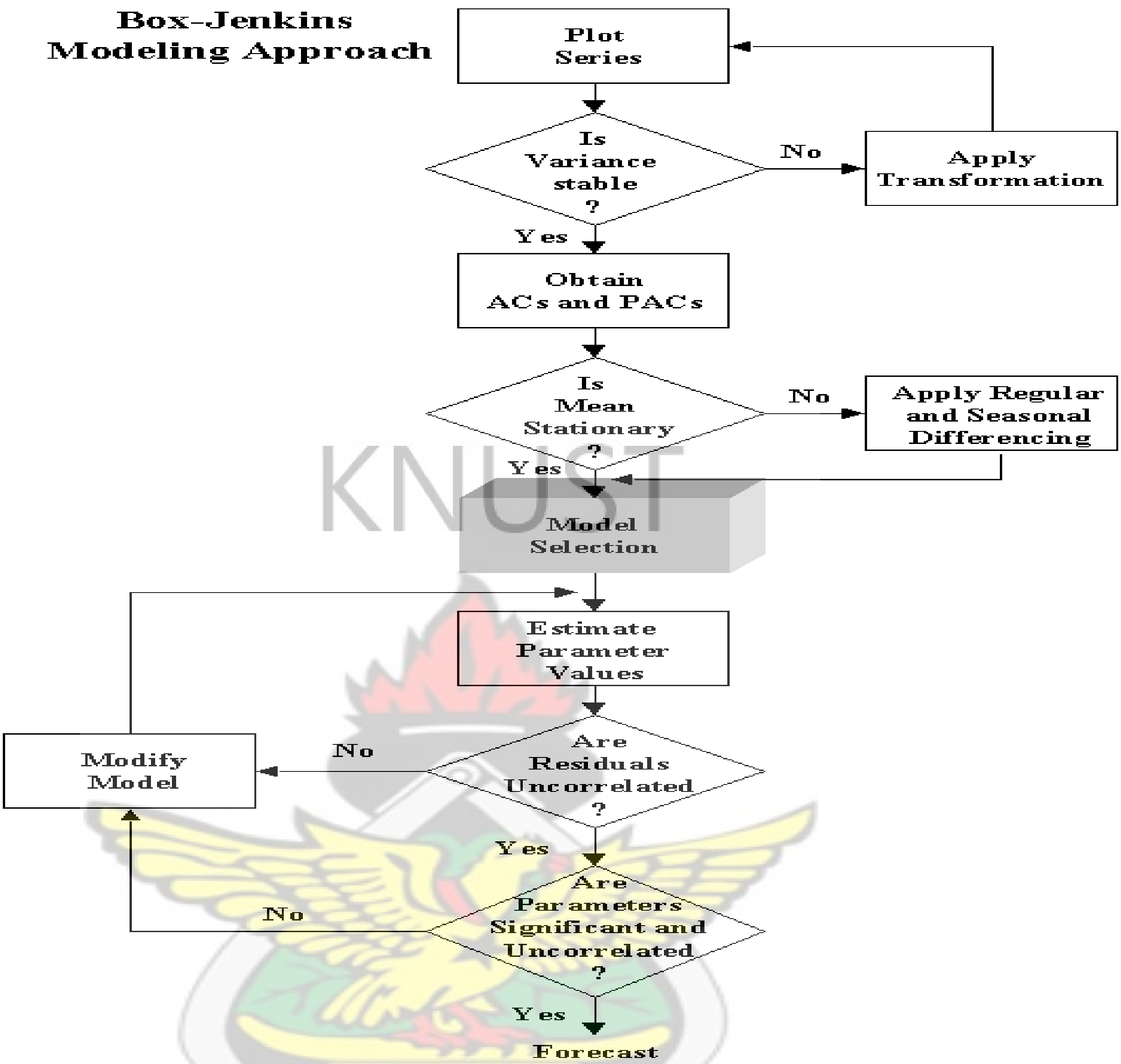


Figure 3. 2: Box-Jenkins modeling approach

Source: adapted from Professor Hossein Arsham

3.11.3. DIAGNOSTICS

Various combinations of AR and MA will yield various models individually and collectively. The best model is obtained with the following diagnostics

i. Low Akaike Information Criteria (AIC) / Bayesian Information Criteria (BIC) /Schwarz-Bayesian Information Criteria (SBC). Lowest values of these statistics are favoured in deciding which model is best and more reliable.

ii. Non significances of auto correlations of residuals via Portmonteau test (Q-test based on Chisquared statistics) – Box –Pierce or Ljung-Box tests. The Ljung-Box statistic (modified Box-Pierce statistic) is a function of the accumulated sample autocorrelation, r_j, up to any specified time lag m. As a function of m, it is defined as: $(m) = n(n+2)\sum_{j=1}^{m}\frac{{r_j}^2}{n-j}$, where n is the number of usable points after any differencing operations. Q(m) should be zero (0) for any lag m, a significant Q(m) for residuals indicates a possible problem with the model.

3.11.4. FORECASTING

The main objective of ARIMA model is forecasting. There are two types of forecast; in-sample forecast and post sample forecast. The former is use to generate confidence in the model and the later is to produce predictions for the feature to enable planners make strategic decisions. The sample period forecasts are obtained by putting the actual values of the explanatory variables in the estimated equation. Important measures of sample period forecast such as mean absolute percentage error (MAPE) is used to judge the forecasting ability of the fitted model.

3.12. CONCLUSION

Like any forecasting technique, the ARIMA methodology does not guarantee perfect forecasts but can be successfully used for forecasting long time series data. The ARIMA technique is suitable for any time series with any pattern of change and it is a good technique for predicting the magnitude of any variable. It does not require the forecaster to choose a priori the value of any parameter. Its disadvantage is its requirement of a significantly large amount of data. It tells “what” will happen based on what has already happened and not “why” it happens. It is often referred to as the ‘black box’ model.

CHAPTER 4

DATA ANALYSIS AND RESULTS

4.0. INTRODUCTION

This chapter employs the Box-Jenkins method (ARIMA) to analyze the data of health utilization by subscribers of the Nadowli DMHIS and to use this data to model and forecast utilization pattern for the next twelve (12) months ending in December, 2012. This methodology requires a significantly large data set and involves primarily three steps in developing a model for any variable: identification, estimation and verification.

4.1. DATA PRESENTATION

In this study, we used the monthly utilization data of subscribers of the Nadowli District Mutual Health Insurance Scheme (DMHIS) for the period January, 2006 – June 2012. This is purely secondary data from the scheme's claims processing department. The data from January 2006 to December was used to create the ARIMA model and the data from January 2012 to June 2012 was used to validate the model. The full data set is shown in Appendix 1.

4.2. PRELIMINARY ANALYSIS

The following figure indicates a time series plot of the monthly utilization by subscribers of the Nadowli DMHIS for the period. It is observed that, there is an increasing upward trend and a form of pattern that seem to repeat every year and peaking in October each year. The peaks are not repeated with the same intensity; peaking increases with increasing time. These indicate that the series is not stationary. A stationary series is the one whose values vary over time only around a constant mean and constant variance. Stationarity is a necessary condition for an ARIMA model to be developed.

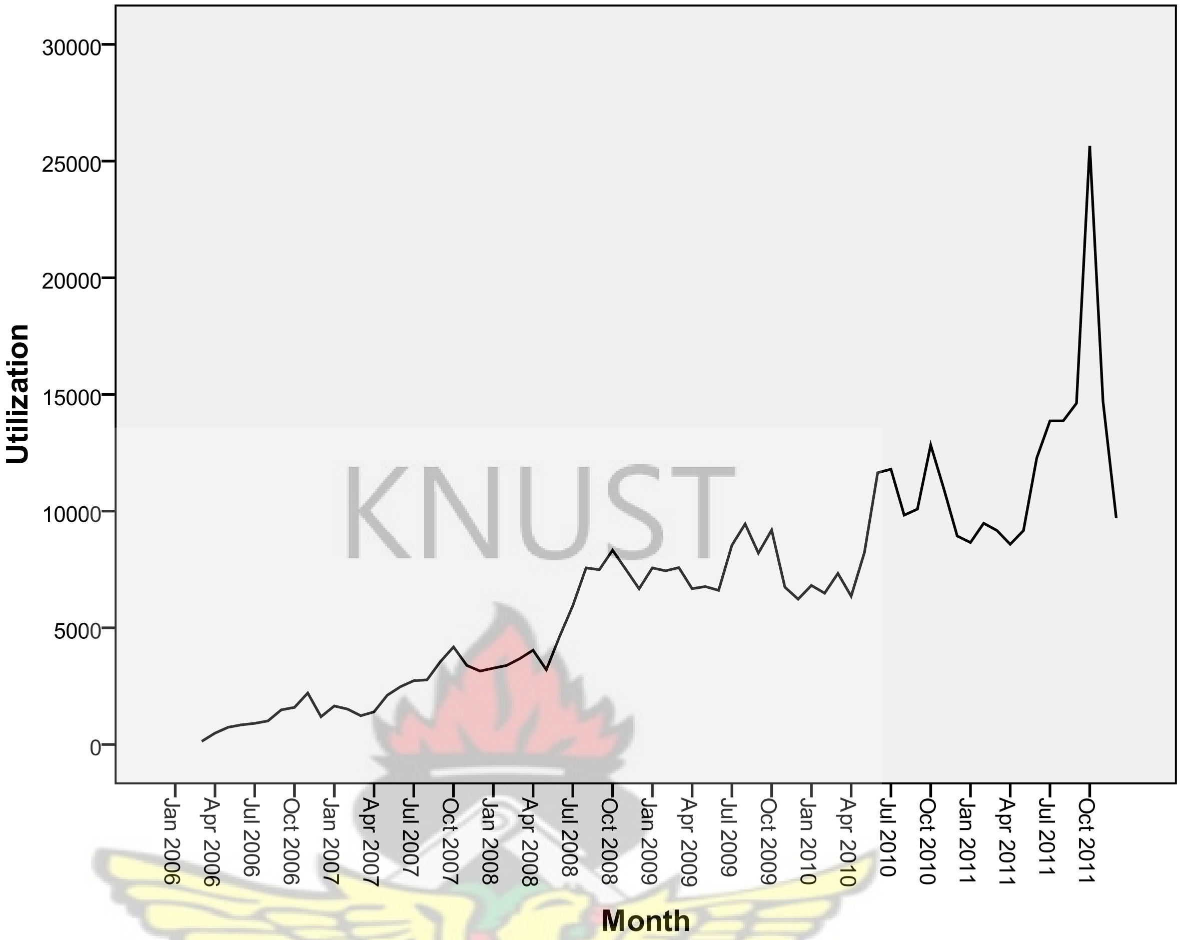


Figure 4. 9: A time series plot of monthly utilization by NHIS subscribers in Nadowli District

Non stationarity is confirmed by the correlograms. The PACF indicates non-stationarity because atleast, one of the vertical bars is higher than the horizontal lines that indicate the cut-off points for statistical significance. Furthermore, the non stationarity is of order one (1) as only the first lagged bar is significantly higher than the cut-off line. This indicates that a first difference transformation will probably get rid of the non-stationarity.

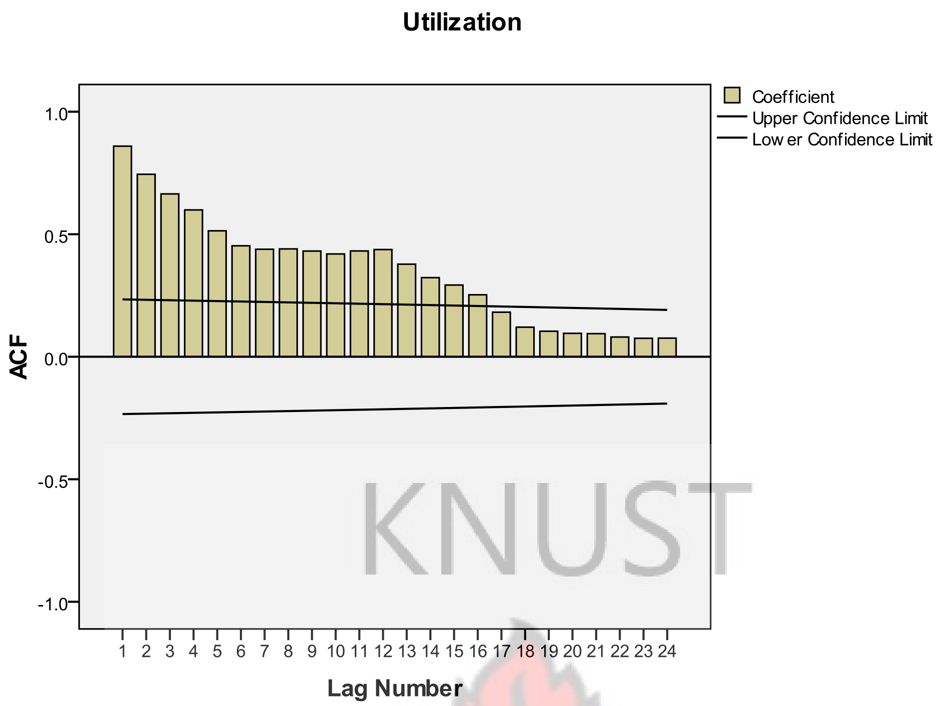


Figure 4. 10: ACF Healthcare Utilization Data

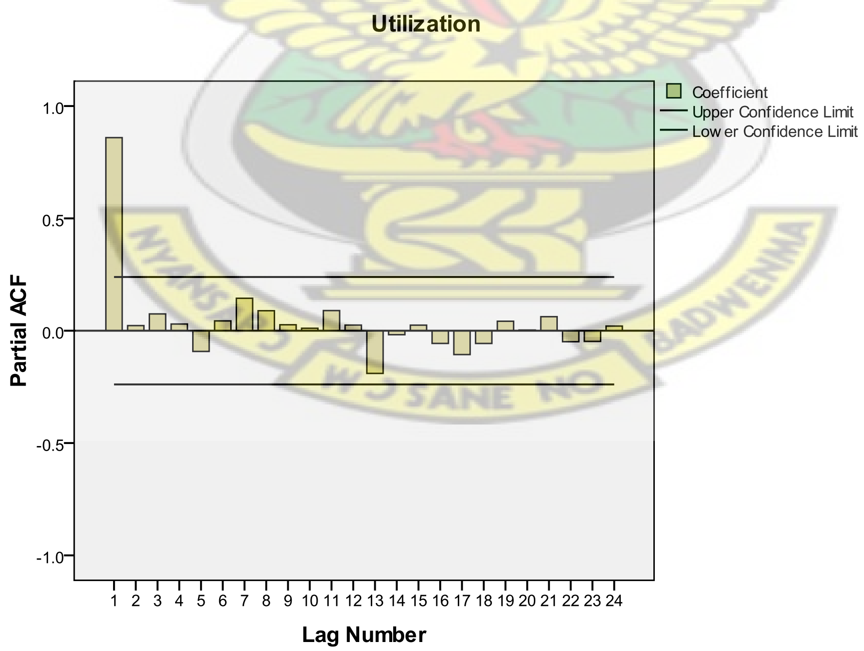


Figure 4. 11: PACF of Healthcare Utilization Data

Also, in order to stabilise the variance, a natural log transformation is required; Natural log transformation of the data may yield a more even peaking as the time series progresses to stationarize the variance.

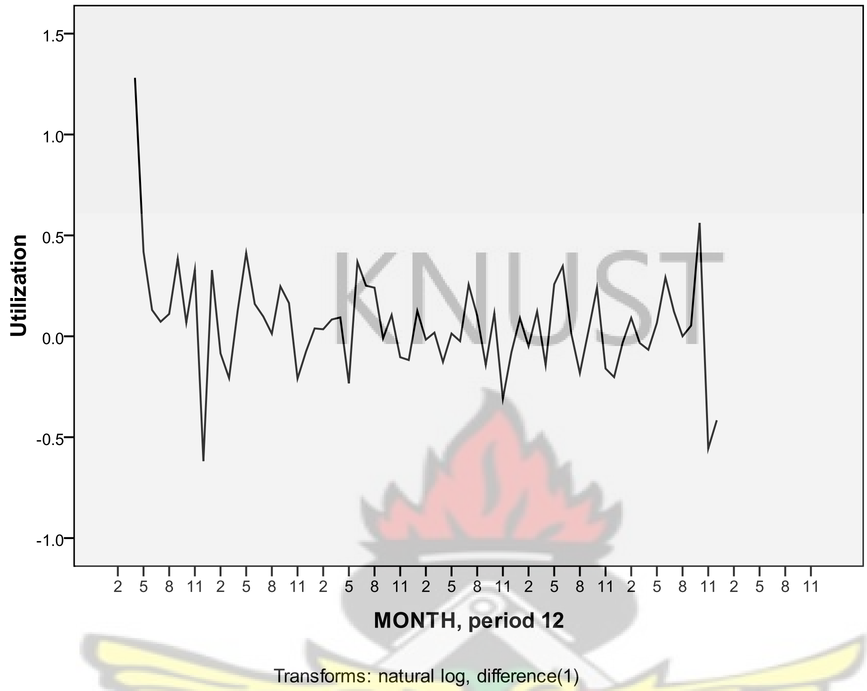


Figure 4. 12: Time series plot of first difference and Log transform of utilization data

The transformation as indicated by the plot in figure 4.4 above depicts a more stationary time series. This is confirmed by the correlograms. All the coefficients are within the confidence bounds, an indication of a stationary series. Also, the ACF and PACF here show sizable coefficients at lag 12, this suggest seasonality in the series because it is generally agreed that for seasonal series, ACF and PACF will show sizable coefficients at multiples of the seasonal lag (in addition to their overall patterns reflecting the non seasonal components of the series).

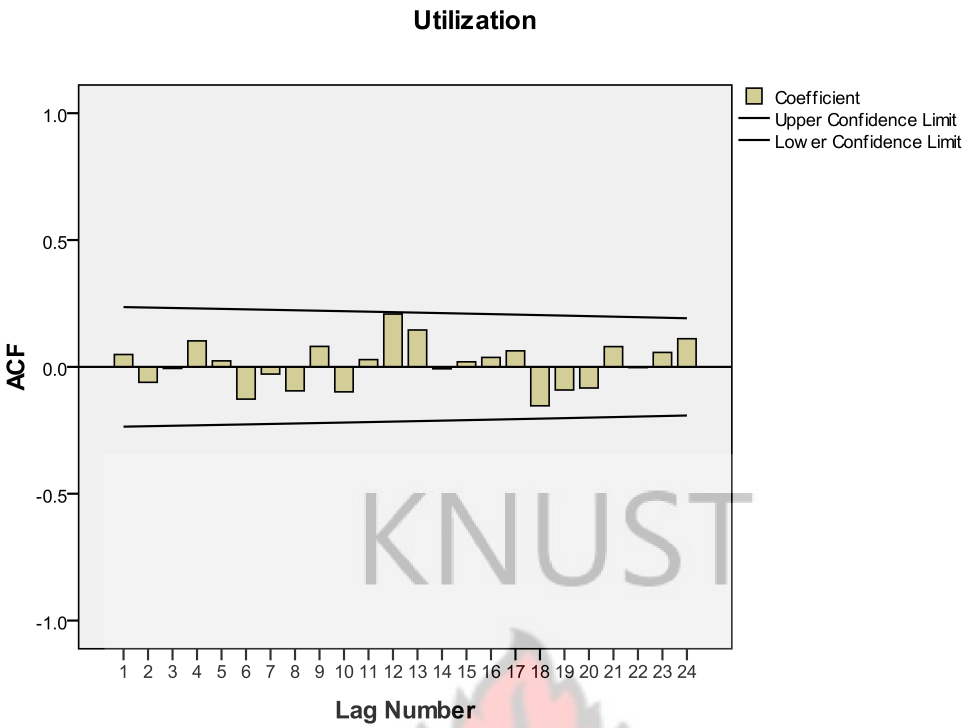


Figure 4. 13: ACF of stationary series of utilization data

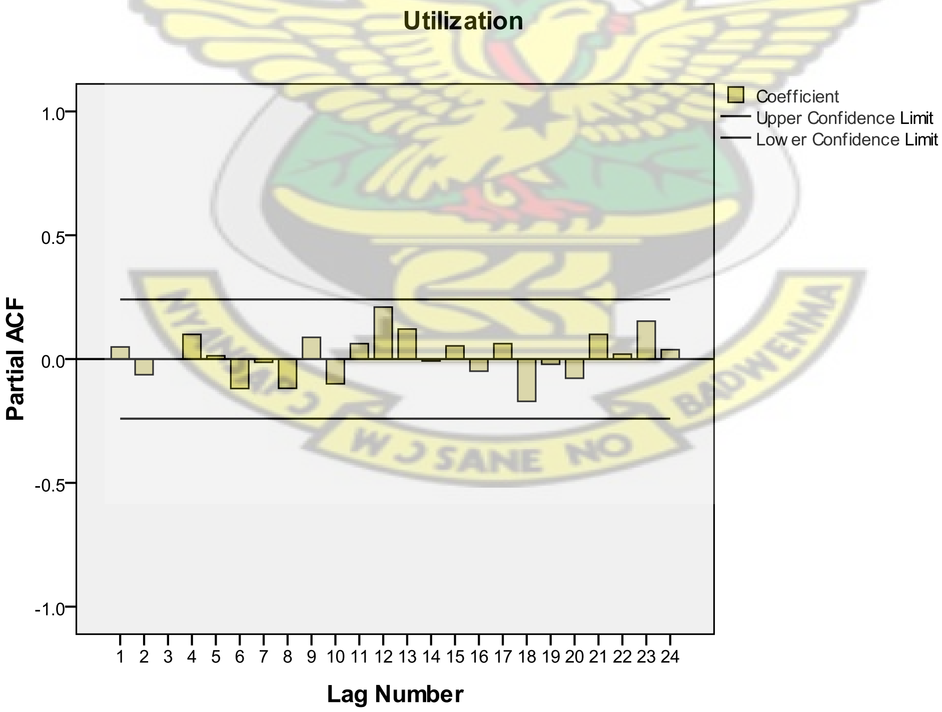


Figure 4. 14: PACF of stationary utilization series

Now that a stationary series have been obtained, a model can be formulated and tested.

4.3.MODEL IDENTIFICATION AND SELECTION

After obtaining a stationary series by differencing once and taking a log transformation, the PACF of the differenced series displays a positive autocorrelation at lag-1. This suggest that there is atleast one Autoregressive term, AR(1). The correlograms also suggest a seasonal model. The following tentative models are selected after several combinations of AR terms:

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

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

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

Table4. 1: Model statistics for tentative ARIMA models

	Model type	**ARIMA(1,1,0)(1,0,0)**	**ARIMA(0,1,0)(2,0,0)**	**ARIMA(0,1,0)(1,0,0)**
Model fit statistics	Normalized BIC	15.226	15.246	15.217
	RMSE	1903.670	1923.149	1954.029
Ljung Box Q(18)	Statistics	14.601	13.365	14.875
	DF	16	16	17
	Sig.	0.554	0.646	0.604

Statistics for the various selected models are displayed in Table 4.1 above. It is observed that, all the Ljung-Box statistics of the models have p-values greater than α-value (0.05) indicating that the models are all valid models. However the best model is ARIMA(0,1,0)(1,0,0). This is because it has the least BIC value, the highest degree of freedom and the least number of parameters (parsimony). Following is a detail presentation of the model ARIMA(0,1,0)(1,0,0).

4.4.ESTIMATION

The model parameters were estimated using SPSS package 17.0. The following tables show the reports. Table 4.2 shows a Ljung-Box statistics of 14.875 with a p-value of 0.604 greater than the α-value (0.05) indicating that the model is a valid model.

Table4. 2: Description of the best ARIMA model

	Model Type
Model ID Utilization Model_1	ARIMA(0,1,0)(1,0,0)

Table4. 3 : Model statistics

Model	Number of Predictors	Model Fit statistics			Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	R-squared	Normalized BIC	Statistics	DF	Sig.	
Utilization-Model_1	0	-.024	.811	15.217	14.875	17	.604	0

Hypothesis:

H_O: Model is inadequate for prediction purposes

H_1: Model is adequate for prediction purposes

From the above, with a Ljung-Box Q Statistic (14.875) and a P-value (0.604), the null hypothesis Ho is rejected and we conclude that the model is statistically significant and adequate for prediction purposes. The co-efficient of determination R^2 = 81.1% means that 81.1% of the total variation in the utilization values can be explained by the chosen mathematical model.

Table4. 4 : ARIMA Model Parameters

					Estimate	SE	t	Sig.
Utilization-Model_1	Utilization	Natural Log	Difference		1			
			AR, Seasonal	Lag 1	.492	.113	4.371	.000

4.5.VERIFICATION

The model verification is concerned with checking the residuals of the model to see if they contain any systematic pattern which still can be removed to improve on the chosen ARIMA model. The various correlations up to 24 lags were computed as displayed in Table 4.5 and the residual plots of ACF and PACF shown in figure: 4.2 below.

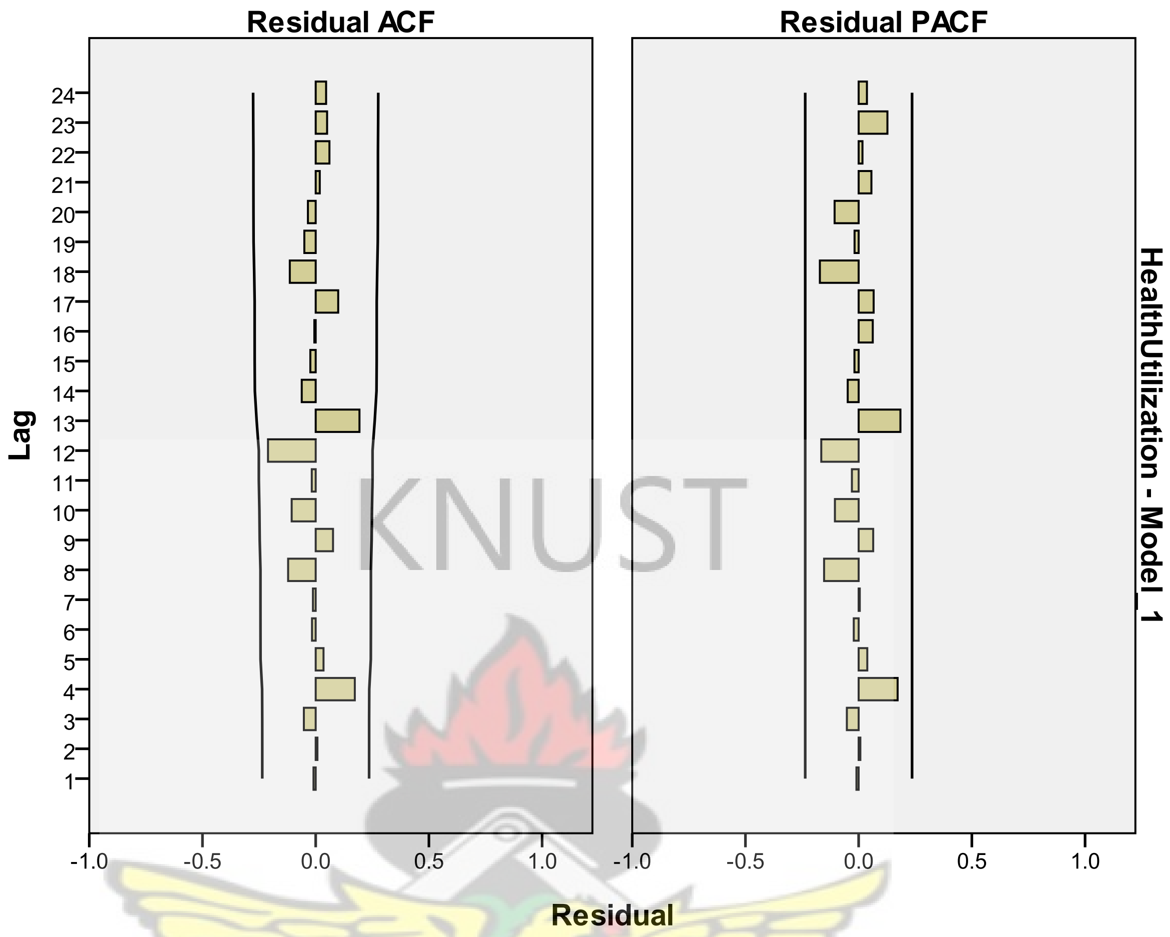


Figure 4. 15: ACF and PACF of residuals of fitted ARIMA model

Table4. 5 : Autocorrelations and partial autocorrelations of residuals

Lag	ACF	SE		Lag	PACF	SE
1	-.010	.120		1	-.010	.120
2	.008	.120		2	.008	.120
3	-.052	.120		3	-.052	.120
4	.173	.121		4	.172	.120
5	.035	.124		5	.038	.120
6	-.016	.124		6	-.022	.120
7	-.012	.124		7	.005	.120
8	-.122	.124		8	-.153	.120
9	.077	.126		9	.064	.120
10	-.106	.127		10	-.105	.120
11	-.017	.128		11	-.030	.120
12	-.211	.128		12	-.165	.120
13	.193	.133		13	.185	.120
14	-.062	.137		14	-.048	.120
15	-.024	.137		15	-.019	.120
16	-.006	.138		16	.062	.120
17	.099	.138		17	.066	.120
18	-.114	.139		18	-.171	.120
19	-.050	.140		19	-.018	.120
20	-.034	.140		20	-.106	.120
21	.018	.140		21	.056	.120
22	.061	.140		22	.017	.120
23	.050	.141		23	.127	.120
24	.046	.141		24	.037	.120

An examination of the residual ACF and PACF indicates that none of these correlations is significantly different from zero at a reasonable level. This shows that the selected model is an appropriate ARIMA model. This also indicates “good fit” of the model. Thus the fitted ARIMA model for health utilization by subscribers of the NHIS in Nadowli is : ARIMA(0,1,0)(1,0,0).

More formally, this can be written in terms of the model parameters i.e.

Seasonal AR components: $\emptyset(B^S) = 1 - \emptyset B^S$

Non-Seasonal first difference: $(1 - B)X_t = X_t - X_{t-1}$

The final model is: ARIMA$(0,1,0)(1,0,0)^{12}$ and can be written as

$\emptyset(\mathbf{B}^s)(1 - \mathbf{B})X_t = (1 - \emptyset\mathbf{B}^s)(X_t - X_{t-1})$

Where, S is the Seasonal period=12, Ø is the parameter estimate (Seasonal Autoregressive estimate) = 0.492, X_t is the study variable (Utilization), and **B** is the backshift operator, that is $\mathbf{B}Y_t=Y_{t-1}$, $\mathbf{B}^{12}Y_t=Y_{t-12}$ etc

Let $\emptyset(B^S)(1 - B)X_t = W_t$

This implies that, $(1 - \emptyset B^S)(X_t - X_{t-1}) = W_t$

$$X_t - X_{t-1} - \emptyset B^S X_t + \emptyset B^S X_{t-1} = W_t$$

$$X_t - X_{t-1} - \emptyset X_{t-s} + \emptyset X_{t-1-s} = W_t$$

Making $\mathbf{X_t}$ the subject,

$$X_t = X_{t-1} + \emptyset X_{t-S} - \emptyset X_{t-1-S} + W_t$$

$$X_t = X_{t-1} + \emptyset(X_{t-S} - X_{t-1-S}) + W_t$$

But S=12, Ø=0.492 and Wt is the error term.

Thus the ARIMA$(0,1,0)(1,0,0)^{12}$ is $X_t = X_{t-1} + 0.492(X_{t-12} - X_{t-13}) + W_t$

The model has AR terms at lags 1, 12 and 13.

4.6.FORECASTING

ARIMA models are developed primarily to forecast the corresponding variable. Sample period forecast is used to develop confidence in the model and a post-sample period forecast is used to generate genuine forecasts for use in planning and other purposes. Table 4.6 and Figures 4.5, 4.6 display the observed values, fitted values, and forecast as well as the confidence level of health care utilization by subscribers in the Nadowli District.

Table4. 6: Observed utilization verses forecasted utilization

Month	Utilization	Predicted value from Utilization Model_1	LCL from Utilization model_1	UCL from Utilization model_1	Deviation (Error)	NRESIDUAL
Jan-06						
Feb-06						
Mar-06	135					
Apr-06	486	141	77	238	345	1.280934
May-06	739	506	276	856	233	0.419089
Jun-06	842	769	420	1301	73	0.130482
Jul-06	905	877	478	1482	28	0.072155
Aug-06	1011	942	514	1593	69	0.11076
Sep-06	1485	1052	574	1780	433	0.384475
Oct-06	1589	1546	843	2614	43	0.06769
Nov-06	2207	1654	903	2798	553	0.328529
Dec-06	1189	2297	1254	3886	-1108	-0.61852
Jan-07	1651	1238	675	2093	413	0.328269
Feb-07	1517	1719	938	2907	-202	-0.08465
Mar-07	1233	1579	862	2671	-346	-0.20728
Apr-07	1397	2387	1415	3789	-990	-0.50535
May-07	2111	1770	1049	2809	341	0.206641
Jun-07	2478	2321	1376	3683	157	0.096092
Jul-07	2733	2647	1569	4201	86	0.062448
Aug-07	2768	2975	1764	4722	-207	-0.04177
Sep-07	3545	3448	2044	5472	97	0.05825
Oct-07	4180	3778	2240	5997	402	0.131469

Nov-07	3389	5065	3003	8040	-1676	-0.37141
Dec-07	3145	2577	1528	4090	568	0.229595
Jan-08	3270	3811	2259	6048	-541	-0.12253
Feb-08	3386	3234	1917	5132	152	0.076506
Mar-08	3680	3152	1869	5003	528	0.185248
Apr-08	4041	4034	2392	6403	7	0.032139
May-08	3201	5104	3026	8101	-1903	-0.43615
Jun-08	4628	3571	2117	5668	1057	0.289798
Jul-08	5948	5007	2968	7947	941	0.202739
Aug-08	7568	6170	3658	9794	1398	0.234613
Sep-08	7490	8812	5224	13986	-1322	-0.13209
Oct-08	8325	8374	4964	13291	-49	0.024625
Nov-08	7505	7741	4589	12286	-236	-4.83E-04
Dec-08	6672	7458	4421	11837	-786	-0.08089
Jan-09	7567	7011	4156	11129	556	0.106701
Feb-09	7442	7936	4704	12596	-494	-0.03381
Mar-09	7580	7993	4738	12686	-413	-0.02259
Apr-09	6675	8183	4851	12987	-1508	-0.17319
May-09	6767	6136	3637	9739	631	0.12834
Jun-09	6605	8364	4958	13275	-1759	-0.20561
Jul-09	8536	7704	4567	12228	832	0.133007
Aug-09	9450	9907	5873	15725	-457	-0.01679
Sep-09	8200	9693	5746	15384	-1493	-0.13678
Oct-09	9188	8905	5279	14134	283	0.061762
Nov-09	6743	9001	5336	14286	-2258	-0.25838
Dec-09	6228	6561	3889	10413	-333	-0.02157
Jan-10	6816	6831	4049	10842	-15	0.028285
Feb-10	6484	6969	4132	11062	-485	-0.04174
Mar-10	7331	6745	3999	10706	586	0.113734
Apr-10	6354	7099	4209	11268	-745	-0.08047
May-10	8225	6595	3909	10467	1630	0.251359
Jun-10	11646	8379	4967	13299	3267	0.359706

Jul-10	11796	13621	8075	21619	-1825	-0.11338
Aug-10	9827	12785	7579	20292	-2958	-0.23267
Sep-10	10087	9448	5601	14996	639	0.09592
Oct-10	12832	10998	6519	17455	1834	0.184722
Nov-10	10938	11361	6735	18032	-423	-0.00748
Dec-10	8934	10844	6428	17212	-1910	-0.16329
Jan-11	8654	9628	5708	15282	-974	-0.07623
Feb-11	9481	8705	5160	13817	776	0.115837
Mar-11	9170	10383	6155	16480	-1213	-0.09376
Apr-11	8579	8811	5223	13985	-232	0.00375
May-11	9167	10042	5953	15938	-875	-0.06069
Jun-11	12274	11214	6648	17799	1060	0.120761
Jul-11	13866	12733	7549	20211	1133	0.11566
Aug-11	13868	13066	7746	20739	802	0.089998
Sep-11	14622	14482	8585	22985	140	0.040095
Oct-11	25647	16969	10060	26934	8678	0.443476
Nov-11	14701	24442	14490	38795	-9741	-0.47794
Dec-11	9697	13719	8133	21775	-4022	-0.31653
Jan-12	9538	9841	5834	15620	-303	
Feb-12	10021	10612	4976	20034	-591	
Mar-12	11872	10762	4186	23047	1110	
Apr-12	9277	10737	3550	25450	-1460	
May-12	10424	11436	3265	29535	-1012	
Jun-12	14157	13610	3394	37875	547	
Jul-12		14898	3271	44298		
Aug-12		15360	2990	48471		
Sep-12		16253	2821	54136		
Oct-12		22091	3433	77313		
Nov-12		17319	2420	63438		
Dec-12		14549	1834	55588		

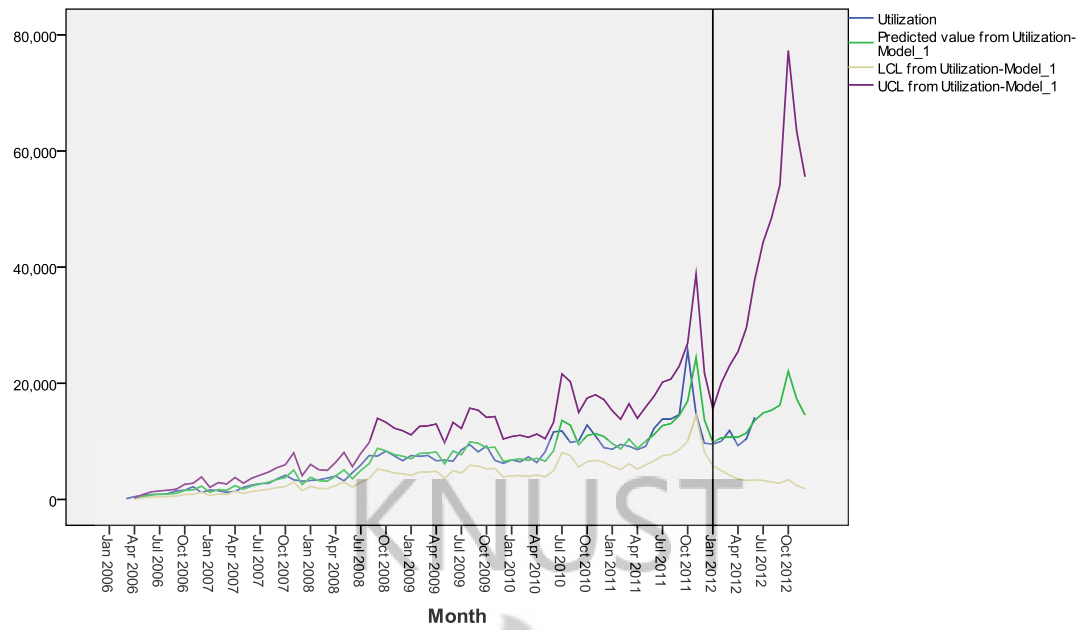


Figure 4. 8: A display of sample and post sample forecast with ARIMA(0,1,0)(1,0,0)

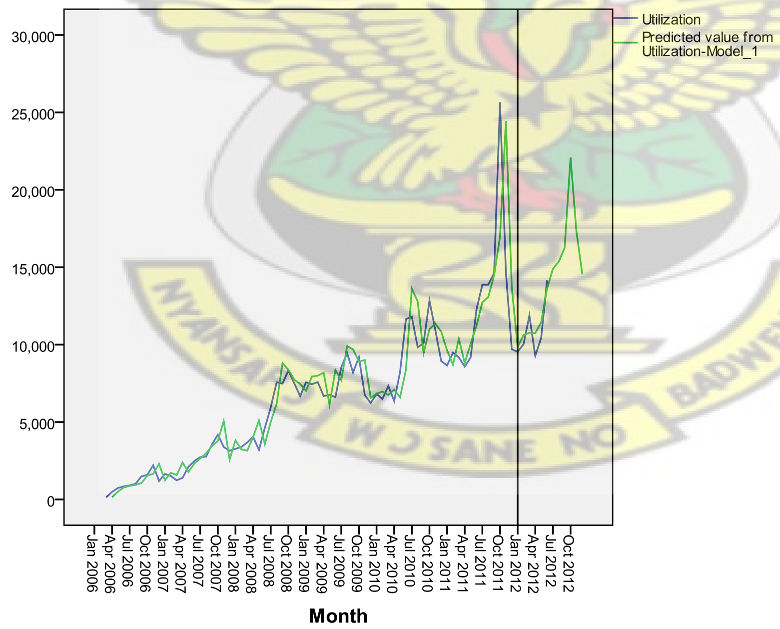


Figure 4.9: A display of observed values against fitted values for sample and post sample forecast with ARIMA(0,1,0)(1,0,0)

4.7 FORECASTING ERROR ANALYSIS

We cannot expect a time-series forecast to be perfect; it will surely and always have prediction errors. The table below summarises some of the error statistics for validating the model base on post sample observations.

Table 4.7 : Error analysis of forecasted utilization

Month (n=6)	Observed utilization (X)	Predicted utilization(Y)	Error $e = X - Y$	Absolute Error $= \lvert e \rvert$	e^2	Fractional Error $= \frac{\lvert e \rvert}{X}$
Jan-12	9,538	9,841	-303	303	91,809	0.0318
Feb-12	10,021	10,612	-591	591	349,281	0.0590
Mar-12	11,872	10,762	1110	1110	1,232,100	0.0935
Apr-12	9,277	10,737	-1460	1460	2,131,600	0.1574
May-12	10,424	11,436	-1012	1012	1,024,144	0.0971
Jun-12	14,157	13,610	547	547	299,209	0.0386
SUM			**-1709**	**5023**	**5,128,143**	**0.4773**

Table4. 8: Statistics for forecasted errors

Cummulative Forecast Error	Mean Error	Mean Squared Error	Root Mean Squared Error	Mean Absolute Deviation	Mean Absolute Percentage Error
CFE	**ME**	**MSE**	**RMSE**	**MAD**	**MAPE**
-1709	284.83	854690.50	924.49	837.17	7.955

With a MAPE of 7.955, the post sample forecast inaccuracy is very low. Thus the ARIMA(0,1,0)(1,0,0) is very good for predicting future values with about 92% accuracy.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.0 INTRODUCTION

Here, a summary of the findings based on specific objectives of the study is stated and recommendations are made based on these findings. The core objective of the study is to Study and characterize the trend or pattern of healthcare services utilization by subscribers of the NHIS in the Nadowli District, build an ARIMA model for prediction and use it to forecast the monthly utilizations for the year 2012 for the Nadowli DMHIS.

5.1 SUMMARY OF FINDINGS

The following results were obtained from the research:

- The best ARIMA model for characterizing monthly healthcare utilization in the district is ARIMA(0,1,0)(1,0,0)^{12} i.e. $X_t = X_{t-1} + 0.492(X_{t-12} - X_{t-13}) + W_t$

  This is a seasonal model with 12 months period as can be seen from the above model statement, it has a linear trend with a difference of about one (1). This suggest that there is room for improvement and that utilization will continue to grow in the short to medium term as more people subscribe to the scheme due to the growing awareness.
- Overall, the pattern of utilization by subscribers increases with time
- Utilization is seasonal, that is there is a pattern that repeats each year. This suggests that utilization is related to environmental factors / weather patterns.
- The lowest utilization values each year are recorded around April and May and the peak utilization month each year is October.

5.2 CONCLUSION

The main research objectives were to model healthcare services utilization by subscribers of the NHIS in the Nadowli District. It was also to use the Box-Jenkins methodology to model the pattern and make forecast for the future based on the time series model developed.

The Box-Jenkins (ARIMA) methodology for analyzing and modeling time series is characterized by three steps:

1) Model identification, 2) Parameter estimation, and 3) Model validation

The validated model is then used to forecast future time series values. The unstationary utilization time series data from January, 2006 to December, 2011 was differenced once and a natural log transformation performed to achieve stationarity. ACFs and PACFs were used to guess several tentative models and the best ARIMA model identified as $\text{ARIMA}(0,1,0)(1,0,0)^{12}$ for the utilization series. The validity of the estimated model was confirmed using the Ljung-Box Q Statistic. The adequacy and goodness of fit of the model was guaranteed from the observations of residual plots of the model. The residual ACF and PACF indicated that none of the correlations is significantly different from zero at a reasonable level; a confirmation of the assumption of normality. This showed that the selected model is fitted well. The forecast for January to December, 2012 were generated with this model and compared with actual post-sample observations for the first six months of 2012. The variation between the forecast and post-sample observations were very minimal attesting to the adequacy of the model. The final model is: $\text{ARIMA}(0,1,0)(1,0,0)^{12}$ and can be written as

$$\emptyset(\mathbf{B}^{s})(1-\mathbf{B})\mathrm{X}_{t} = (1-\emptyset\mathbf{B}^{s})(\mathrm{X}_{t}-\mathrm{X}_{t-1})$$

Where, S is the Seasonal period=12, Ø is the parameter estimate (Seasonal Autoregressive estimate) = 0.492, X_t is the study variable (Utilization), and **B** is the backshift operator, that is $\mathbf{B}Y_t=Y_{t-1}$, $\mathbf{B}^{12}Y_t=Y_{t-12}$ etc

Thus $X_t = X_{t-1} + 0.492(X_{t-12} - X_{t-13}) + W_t$, where W_t is the error term

This is an AR model with predictors at lags 1, 12 and 13.

5.3 RECOMMENDATIONS

From our results and discussions we make the following recommendations to stakeholders:

- The model indicates an increasing utilization pattern in the short to medium term. This may be due to growing awareness. The NHIS must increase their capacity in terms of manpower, resources and plan in order to respond to the growth.
- The time series model should be adapted and use for six months forecasting. This should continuously be updated for forecasting.
- The model indicates that the pattern of utilization is seasonal; peaking in October each year with lower figures registered in April and May. Advance plans should be put in place to speed up claims processing during the peak periods to ensure maximum output.
- Maintenance plans and leave periods for staff should be implemented during the low utilization months to ensure that claims processing is not significantly affected negatively.
- Most of the utilization (over 90%) results from the Out Patients Departments (OPD) of the various providers; it will therefore reduce the work load significantly if an alternative payment mechanism like capitation is adopted

for OPD services utilization instead of the current G-DRG system and itemized billing that involves huge paperwork for processing.

- The healthcare providers in the district should ensure that adequate medical supplies are made available especially in the month of October where huge utilization by subscribers is experience across the district.

REFERENCES:

Arsham, H. "Time-Critical Decision Making for Business Administration" http://home.ubalt.edu/ntsbarsh/stat-data/forecast.htm (accessed 2011 December 13).

Ashuri, B. and Lu, J. (2010) "Is it possible to forecast the Construction Cost Index in the USA?" Georgia Institute of Technology, USA. http://www.ricsamericas.org/files/editor/file/Various%20Media/700_Construction_Cost_AMERCIA%20v7%20web%20opt.pdf (accessed 2011 December 13)

Brisson, M., Campbell, B. and Galbraith, J.W. "Forecasting some low-predictability time series Using diffusion indices" http://www.google.com.gh/url?sa=t&rct=j&q=Major+work+in+time+series+forecasting&source=web&cd=29&ved=0CFMQFjAIOBQ&url=http%3A%2F%2Fciteseerx.ist.psu.edu%2Fviewdoc%2Fdownload%3Fdoi%3D10.1.1.132.9374%26rep%3Drep1%26type%3Dpdf&ei=XorpTtCoE4K6-AbZ7azFCg&usg=AFQjCNH2IZ2o-lkTUzHrQ3mdqHCgQPWkPQ (accessed 2011 December 15)

Contreras, J., Espínola, R., Nogales, F.J. and Conejo, A.J.(2003), "ARIMA Models to Predict Next-days Electricity Prices", IEEE Transactions on power systems, Vol. 18, No.3. http://scholar.google.com/scholar_url?hl=en&q=http://citeseerx.ist.psu.edu/viewd

oc/download%3Fdoi%3D10.1.1.198.2576%26rep%3Drep1%26type%3Dpdf&sa=X&scisig=AAGBfm2cnTbEfWJ2JaRD_BRtn_yRTX_3IQ&oi=scholarr (accessed 2012 January 6)

De Gooijer J.D and Hyndman R.J. (2006), "25 years of time series forecasting", International Journal of Forecasting 22 (2006) 443– 473. http://www.est.uc3m.es/esp/nueva_docencia/comp_col_get/lade/tecnicas_prediccion/Practicas0708/Practica1/25%20years%20of%20time%20series%20forecasting%20(Gooijer%20and%20Hyndman).pdf (accessed 2011 December 15)

Dobre, I. and Alexandru, A.A.M. (2008), Journal of Applied Quantitative Method Vol .3 No.2 Summer 2008 "Modelling unemployment rate using Box-Jenkins procedure" http://jaqm.ro/issues/volume-3,issue-2/pdfs/dobre_alexandru.pdf (accessed 2011 December 15)

El-Mefleh, M.A. and Shotar, M.M. "The application of Time Series modeling to some major economic variables" http://aabss.org/Perspectives2009/AABSS2009TheApplicationOfTimeSeriesModelingToSomeMajorEconomicVariables.pdf (accessed 2012 January 6)

Generating a perfect ARIMA model. http://calcnet.mth.cmich.edu/org/spss/V16_materials/Video_Clips_v16/43TSExpert/43TSExpert.swf (accessed 2012 October 25)

Georgakarakos S., Koutsoubas D. and Valavanis V. (2006) "Time series analysis and forecasting techniques applied on loliginid and ommastrephid landings in Greek waters" http://arch.her.hcmr.gr/papers/fisres78p55.pdf (accessed 2012 January 6)

Gilmour, S., Degenhardt, L., Hall, W. and Day, C. "Using intervention time series analyses to assess the effects of imperfectly identifiable natural events: a general method and example" http://www.biomedcentral.com/1471-2288/6/16 (accessed 2012 January, 4)

Gyapong, H., Garshong, B., Akazili, J., Aikins, M., Agyepong, I., Nyanator, F. (2007) "Critical analysis of Ghana's health system with a focus on equity challenges and the national health insurance http://web.uct.ac.za/depts/heu//SHIELD/reports/Ghana1.pdf (accessed 2011 November 15)

History of Health Insurance: http://en.wikipedia.org/wiki/History_of_insurance (accessed 2011 November 15)

ILO/STEP, Health Micro-Insurance Schemes: Feasibility Study Guide, Volume 2: Tools. Geneva, International Labour Office, Strategies and Tools against social Exclusion and Poverty (STEP), 2005.

Jensen, A.L. (1999) "Using simulation to verify life history relations indicated by time series analysis" http://deepblue.lib.umich.edu/bitstream/2027.42/35235/1/348_ftp.pdf (accessed 2011 December 13)

Jones, C.B (1997) "Life history patterns of Howler Monkeys in a time-varying environment" http://testforms.uncfsu.edu/psychology/Dept_Web_Pages/LIFEHISTORY Jones.pdf (accessed 2011 December 13)

Legislative Instrument (L.I.1809), (2004), Parliament of Ghana, Government of Ghana.

Makridakis, S. and Hibon, M. "ARMA models and the Box Jenkins methodology". http://www.insead.edu/facultyresearch/research/doc.cfm?did=46900 (accessed 2012 January 4)

Mandal, B.N. (2005), "Forecasting sugarcane production in India with ARIMA model" http://interstat.statjournals.net/YEAR/2005/articles/0510002.pdf (accessed 2011 December 21)

National Health Insurance Act, Act 650 (2003) Parliament of Ghana. Government of Ghana.

Nogales F.J. and Conejo A.J. (2005), “Electricity price forecasting through transfer function models”. Journal of the Operational Research Society (2006), http://www.jstor.org/stable/pdfplus/4102386.pdf (accessed 2012 February 2)

Owusu, A.Y.(2010) “National / Mutual Health Insurance in Ghana, Introduction, Access and Framework”, http://www.microhealthinsurance-africa.org/fileadmin/user_upload/Ghana-Workshop/Owusu_2010_The_NHIS.pdf (accessed 2011 November 15)

Owusu, G.A.(2009) “Mutual Health Insurance verses National Health Insurance in Ghana” http://www.microhealthinsurance-africa.org/fileadmin/user_upload/training-material/Owusu_MHI-vs-NHIS-in-Ghana.pdf (accessed 2011 November 15)

Taylor, J.W. (2008), Management Science, 2008, Vol. 54, pp. 253-265 “A Comparison of Univariate Time Series Methods for Forecasting Intraday Arrivals at a Call Center” http://users.ox.ac.uk/~mast0315/CallCenterFcstComparison.pdf (accessed 2012 January 4)

Tulone, D. and Madden, S. “PAQ: Time Series Forecasting For Approximate Query Answering In Sensor Networks http://db.csail.mit.edu/pubs/EWSN1.pdf (accessed 2011 December 15)

Webster, G.D. (2007), “What’s in a Name: Is “Evolutionary Psychology” Eclipsing “Sociobiology” in the Scientific Literature?”

http://www.epjournal.net/filestore/ep05683695.pdf (accessed 2011 December 14)

Wongkoon, S., Pollar, M., Jaroensutasinee, M. and Jaroensutasinee, K. (2007), Predicting DHF Incidence in Northern Thailand using Time Series Analysis Technique http://www.waset.org/journals/waset/v32/v32-42.pdf (accessed 2011 December 15)

Zhang, G.P. and Kline, D.M. (2007) "Quarterly Time-Series Forecasting With Neural Networks" IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 18, NO. 6, NOVEMBER 2007 http://forecasters.org/pdfs/IEEE_TNN_2007.pdf (accessed 2011 December 15)

APPENDICES

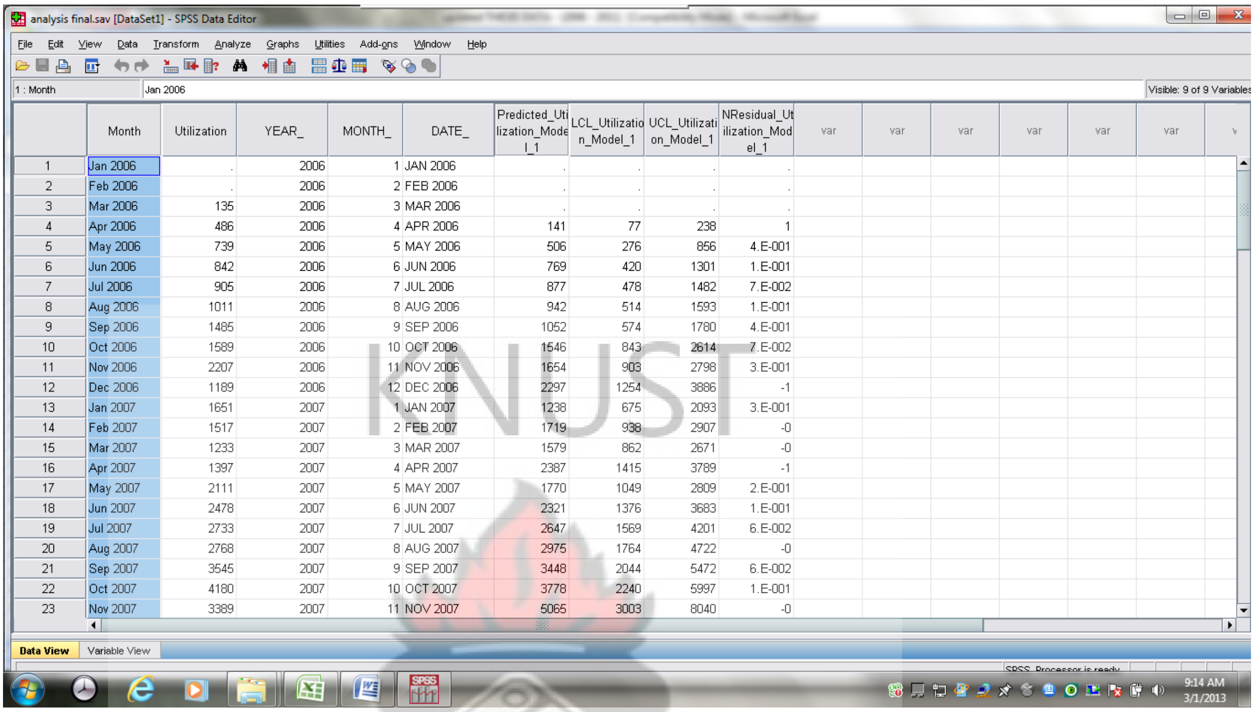
APPENDIX 1

Data on Healthcare Utilization by Subscribers of the Nadowli DMHIS

MONTH	YEAR						
	2006	2007	2008	2009	2010	2011	2012
JAN	-	1,651	3,270	7,567	6,816	8,654	9,538
FEB	-	1,517	3,386	7,442	6,484	9,481	10,021
MAR	135	1,233	3,680	7,580	7,331	9,170	11,872
APR	486	1,397	4,041	6,675	6,354	8,579	9,277
MAY	739	2,111	3,201	6,767	8,225	9,167	10,424
JUN	842	2,478	4,628	6,605	11,646	12,274	14,157
JUL	905	2,733	5,948	8,536	11,796	13,866	
AUG	1,011	2,768	7,568	9,450	9,827	13,868	
SEP	1,485	3,545	7,490	8,200	10,087	14,622	
OCT	1,589	4,180	8,325	9,188	12,832	25,647	
NOV	2,207	3,389	7,505	6,743	10,938	14,701	
DEC	1,189	3,145	6,672	6,228	8,934	9,697	

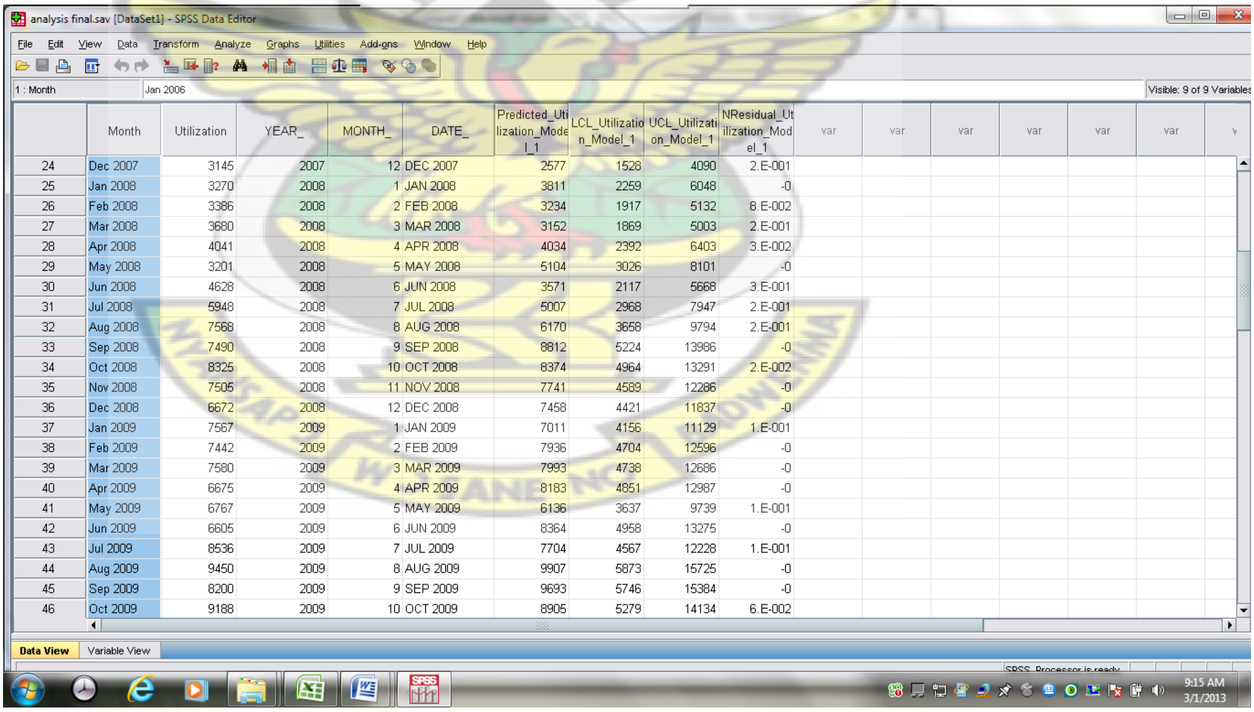
APPENDIX 2

Screen Print of Analysis



analysis final.sav [DataSet1] - SPSS Data Editor

	Month	Utilization	YEAR_	MONTH_	DATE_	Predicted_Utilization_Model_1	LCL_Utilization_Model_1	UCL_Utilization_Model_1	NResidual_Utilization_Model_1
1	Jan 2006	.	2006	1	JAN 2006	.	.	.	.
2	Feb 2006	.	2006	2	FEB 2006	.	.	.	.
3	Mar 2006	135	2006	3	MAR 2006	.	.	.	.
4	Apr 2006	486	2006	4	APR 2006	141	77	238	1
5	May 2006	739	2006	5	MAY 2006	506	276	856	4.E-001
6	Jun 2006	842	2006	6	JUN 2006	769	420	1301	1.E-001
7	Jul 2006	905	2006	7	JUL 2006	877	478	1482	7.E-002
8	Aug 2006	1011	2006	8	AUG 2006	942	514	1593	1.E-001
9	Sep 2006	1485	2006	9	SEP 2006	1052	574	1780	4.E-001
10	Oct 2006	1589	2006	10	OCT 2006	1546	843	2614	7.E-002
11	Nov 2006	2207	2006	11	NOV 2006	1654	903	2798	3.E-001
12	Dec 2006	1189	2006	12	DEC 2006	2297	1254	3886	-1
13	Jan 2007	1651	2007	1	JAN 2007	1238	675	2093	3.E-001
14	Feb 2007	1517	2007	2	FEB 2007	1719	938	2907	-0
15	Mar 2007	1233	2007	3	MAR 2007	1579	862	2671	-0
16	Apr 2007	1397	2007	4	APR 2007	2387	1415	3789	-1
17	May 2007	2111	2007	5	MAY 2007	1770	1049	2809	2.E-001
18	Jun 2007	2478	2007	6	JUN 2007	2321	1376	3683	1.E-001
19	Jul 2007	2733	2007	7	JUL 2007	2647	1569	4201	6.E-002
20	Aug 2007	2768	2007	8	AUG 2007	2975	1764	4722	-0
21	Sep 2007	3545	2007	9	SEP 2007	3448	2044	5472	6.E-002
22	Oct 2007	4180	2007	10	OCT 2007	3778	2240	5997	1.E-001
23	Nov 2007	3389	2007	11	NOV 2007	5065	3003	8040	-0



analysis final.sav [DataSet1] - SPSS Data Editor

	Month	Utilization	YEAR_	MONTH_	DATE_	Predicted_Utilization_Model_1	LCL_Utilization_Model_1	UCL_Utilization_Model_1	NResidual_Utilization_Model_1
24	Dec 2007	3145	2007	12	DEC 2007	2577	1528	4090	2.E-001
25	Jan 2008	3270	2008	1	JAN 2008	3811	2259	6048	-0
26	Feb 2008	3386	2008	2	FEB 2008	3234	1917	5132	8.E-002
27	Mar 2008	3680	2008	3	MAR 2008	3152	1869	5003	2.E-001
28	Apr 2008	4041	2008	4	APR 2008	4034	2392	6403	3.E-002
29	May 2008	3201	2008	5	MAY 2008	5104	3026	8101	-0
30	Jun 2008	4628	2008	6	JUN 2008	3571	2117	5668	3.E-001
31	Jul 2008	5948	2008	7	JUL 2008	5007	2968	7947	2.E-001
32	Aug 2008	7568	2008	8	AUG 2008	6170	3658	9794	2.E-001
33	Sep 2008	7490	2008	9	SEP 2008	8812	5224	13986	-0
34	Oct 2008	8325	2008	10	OCT 2008	8374	4964	13291	2.E-002
35	Nov 2008	7505	2008	11	NOV 2008	7741	4589	12286	-0
36	Dec 2008	6672	2008	12	DEC 2008	7458	4421	11837	-0
37	Jan 2009	7567	2009	1	JAN 2009	7011	4156	11129	1.E-001
38	Feb 2009	7442	2009	2	FEB 2009	7936	4704	12596	-0
39	Mar 2009	7580	2009	3	MAR 2009	7993	4738	12686	-0
40	Apr 2009	6675	2009	4	APR 2009	8183	4851	12987	-0
41	May 2009	6767	2009	5	MAY 2009	6136	3637	9739	1.E-001
42	Jun 2009	6605	2009	6	JUN 2009	8364	4958	13275	-0
43	Jul 2009	8536	2009	7	JUL 2009	7704	4567	12228	1.E-001
44	Aug 2009	9450	2009	8	AUG 2009	9907	5873	15725	-0
45	Sep 2009	8200	2009	9	SEP 2009	9693	5746	15384	-0
46	Oct 2009	9188	2009	10	OCT 2009	8905	5279	14134	6.E-002

	Month	Utilization	YEAR_	MONTH_	DATE_	Predicted_Utilization_Model_1	LCL_Utilization_Model_1	UCL_Utilization_Model_1	NResidual_Utilization_Model_1
47	Nov 2009	6743	2009	11	NOV 2009	9001	5336	14286	-0
48	Dec 2009	6228	2009	12	DEC 2009	6561	3889	10413	-0
49	Jan 2010	6816	2010	1	JAN 2010	6831	4049	10842	3.E-002
50	Feb 2010	6484	2010	2	FEB 2010	6969	4132	11062	-0
51	Mar 2010	7331	2010	3	MAR 2010	6745	3999	10706	1.E-001
52	Apr 2010	6354	2010	4	APR 2010	7099	4209	11268	-0
53	May 2010	8225	2010	5	MAY 2010	6595	3909	10467	3.E-001
54	Jun 2010	11646	2010	6	JUN 2010	8379	4967	13299	4.E-001
55	Jul 2010	11796	2010	7	JUL 2010	13621	8075	21619	-0
56	Aug 2010	9827	2010	8	AUG 2010	12785	7579	20292	-0
57	Sep 2010	10087	2010	9	SEP 2010	9448	5601	14996	1.E-001
58	Oct 2010	12832	2010	10	OCT 2010	10998	6519	17455	2.E-001
59	Nov 2010	10938	2010	11	NOV 2010	11361	6735	18032	-0
60	Dec 2010	8934	2010	12	DEC 2010	10844	6428	17212	-0
61	Jan 2011	8654	2011	1	JAN 2011	9628	5708	15282	-0
62	Feb 2011	9481	2011	2	FEB 2011	8705	5160	13817	1.E-001
63	Mar 2011	9170	2011	3	MAR 2011	10383	6155	16480	-0
64	Apr 2011	8579	2011	4	APR 2011	8811	5223	13985	4.E-003
65	May 2011	9167	2011	5	MAY 2011	10042	5953	15938	-0
66	Jun 2011	12274	2011	6	JUN 2011	11214	6648	17799	1.E-001
67	Jul 2011	13866	2011	7	JUL 2011	12733	7549	20211	1.E-001
68	Aug 2011	13868	2011	8	AUG 2011	13066	7746	20739	9.E-002
69	Sep 2011	14622	2011	9	SEP 2011	14482	8585	22985	4.E-002

	Month	Utilization	YEAR_	MONTH_	DATE_	Predicted_Utilization_Model_1	LCL_Utilization_Model_1	UCL_Utilization_Model_1	NResidual_Utilization_Model_1
70	Oct 2011	25647	2011	10	OCT 2011	16969	10060	26934	4.E-001
71	Nov 2011	14701	2011	11	NOV 2011	24442	14490	38795	-0
72	Dec 2011	9697	2011	12	DEC 2011	13719	8133	21775	-0
73	Jan 2012	9538	2012	1	JAN 2012	9841	5834	15620	.
74	Feb 2012	10021	2012	2	FEB 2012	10612	4976	20034	.
75	Mar 2012	11872	2012	3	MAR 2012	10762	4186	23047	.
76	Apr 2012	9277	2012	4	APR 2012	10737	3550	25450	.
77	May 2012	10424	2012	5	MAY 2012	11436	3265	29535	.
78	Jun 2012	14157	2012	6	JUN 2012	13610	3394	37875	.
79	Jul 2012	.	2012	7	JUL 2012	14898	3271	44298	.
80	Aug 2012	.	2012	8	AUG 2012	15360	2990	48471	.
81	Sep 2012	.	2012	9	SEP 2012	16253	2821	54136	.
82	Oct 2012	.	2012	10	OCT 2012	22091	3433	77313	.
83	Nov 2012	.	2012	11	NOV 2012	17319	2420	63438	.
84	Dec 2012	.	2012	12	DEC 2012	14549	1834	55588	.
85									
86									
87									
88									
89									
90									
91									
92									