APPLICATION OF STATISTICAL MODELS TO OUTPATIENT DEPARTMENT (OPD) ATTENDANCE DATA IN SALTPOND MUNICIPAL HOSPITAL OF THE CENTRAL REGION OF GHANA

By

JAPHETH ARTHUR (BSc. Statistics, HND Statistics)

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

MASTER OF PHILOSOPHY (Industrial Mathematics)

Institute of Distance Learning

MAY 2013

CERTIFICATION

I hereby declare that this submission is my own work towards the MPhil. degree 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.

Japheth Arthur	…………………	…………………
(PG 6317411)	Signature	Date

Certified by:

Dr. Richard K. Avuglah	…………………	…………………
(Supervisor)	Signature	Date
Mr. I. K. Darkwah	…………………	…………………
(HOD)	Signature	Date
Professor I. K. Dontwi	…………………	…………………
(Dean, IDL)	Signature	Date

ABSTRACT

Hospitals in Ghana are always overwhelmed by the unexpected high Outpatient Department (OPD) attendance. The Saltpond Municipal Hospital at Saltpond in the Central Region of Ghana is no exception. This reason underpins this study where OPD attendance data from 2002 to 2012 as can be seen in appendix A was gathered for developing an adequate time series model and forecasting attendance for the next five (5) years. Several time series models including AR, MA, ARMA, ARIMA and SARIMA were fitted to the data. It was revealed that the most adequate model for the data was the seasonal ARIMA (1, 1, 3) $(0, 1, 1)_{12}$. Also, there will be an increase in the OPD attendance at the hospital in the next 5 years. It is recommended that in order for the hospital to prepare adequately, the forecasted figures should be relied upon in planning its activities.

ACKNOWLEDGEMENTS

The successful completion of this work owes much to individuals who cannot go unnoticed. My heartfelt gratitude goes to Dr. Richard K. Avuglah for perusing through my work and most particularly his personal friendship with me. I sincerely appreciate the invaluable tuition and direction of all lecturers of the Department of Mathematics at the Kwame Nkrumah University of Science and Technology (KNUST), Kumasi especially Professor S. K. Amponsah, Dr. F. T. Oduro, Dr. Richard K. Avuglah and Dr. Richard Koufie. I also appreciate the invaluable assistance and morale support of my dear wife, Mrs. Beatrice Arthur, and my son, Papa Kwame Arthur. Also, I thank my colleagues especially Francis Mawuli Abude, Bernard Kudjo Nana Brany and Cosmos Agbe Todoko. My heartfelt thanks go to the officials at the Saltpond Municipal Hospital's Records and Statistics Department for freely given me the data needed for this work. Again, I appreciate the effort of the officers at the Health Information Department of the Central Regional Health Directorate, Ghana Health Service.

DEDICATION

To my mother, Madam Efua Atta Eku

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AR	–	Autoregressive
ARMA	–	Autoregressive Moving Average
ARIMA	–	Autoregressive Integrated Moving Average
CR	–	Central Region
GHS	–	Ghana Health Service
GSS	–	Ghana Statistical Service
KPSS	–	Kwiatkowski-Phillips-Schmidt-Shin test
MA	–	Moving Average
OPD	–	Outpatient Department
SARIMA	–	Seasonal Autoregressive Integrated Moving Average

CHAPTER ONE

INTRODUCTION

1.1 Background of study

The first point of call of every health facility is its Out-Patient Department (OPD) and the clients judge the standard of the facility by the standard of care they receive at the OPD. It is an accepted fact that first impressions go a long way to fashion out of people's perceptions about things and places they come into contact with. Several studies in Ghana have shown that outpatient attendance continues to increase proportionately with the increase in the population of Ghana.

It is important to note that some of the factors that affect the frequency and the length of hospitalisation include diseases that hospitalise the patient and the diagnoses given them. Recently, District Health Management Team (DHMT) of Ho District, which found out that the OPD attendance in the district has been declining for the past four years. The DHMT has attached importance to this falling OPD attendance, therefore requested that the factors influencing OPD attendance and their extent of influence be reviewed. It was found out that public health facilities were geographically and financially accessible to the people. However, a reasonable proportion of the respondents live more than 1km from where they seek treatment. Socio-cultural values and beliefs had no influence on OPD attendance, as there were no social or cultural values/beliefs preventing respondents from using orthodox medicine. In addition, outreach services and drug peddlers' activities had no influence. Although, there were different types of health providers in the communities studied, orthodox medicine was preferred. Majority of the respondents (88.4%) were satisfied with the services provided at the public health facilities. Meanwhile, a reasonable proportion (41.3%) was not issued with receipts covering their expenses at the health centre level. The users of the district hospital suggested that staff change their attitude towards patients by showing them respect and patience. Based on the results, the following conclusions were drawn. The main causes of falling OPD attendance are facility-related. These were poor attitude of health staff towards patients (14.4%), lack of valid receipts covering expenses (41.3%) and health staff not at post (26.5%). The recommendations among others provided to DHMT include close supervision to ensure that valid receipt are issued to patients, drugs as well as other materials are used for their purpose, management should encourage staff to show respect, patience and care for patients.

The frequency and seriousness of the disease may differ from male to female. Because of the basic biological differences between males and females, some diseases are protracted more in men than women and vice versa. Pregnancy among woman with its attendant labour period is one important factor that may hospitalise them. The frequency and the length of hospitalisation differ according to age, sex and genotype.

The frequency and intensity of some diseases tend to change with age. Indeed Thompson (1962) found in Ghana that while sickle cell disease children were protected against malaria, sickle cell trait adult were those who reported most often with malaria infection. According to Konotey-Ahulu (1972), "at the vulnerable age when malaria takes its heaviest toll the sickle cell trait has been found to enjoy an advantage over either individual, but at the later stages in life the sickle cell trait has been said occasionally to have more severe manifestations in certain diseases.” Another factor suspected to be affecting frequency of hospitalisation and in particular length of hospitalisation is the kind of treatment (diagnosis) given to the patients. Diagnosis, we suspect will influence frequency and length of hospitalisation because some patients were transferred from one ward to another to receive different treatment after about a week’s stay in the former. Also patients after being discharged from one ward were re-admitted to different wards to receive different treatment after a day or two at home.

Agbodza (1977) observed an increasing trend in the number of Onchocerciasis cases treated in the Kpando area of the Volta Region. Upon forecasting, he reasonably concluded that the number of Onchocerciasis cases expected in the area in 1977 would be higher than that recorded in 1976. It is, however, not the intention here to consider factor that compel people to attend hospital. What this study considers, is to determine if there has been any established pattern characterising the OPD attendance at Saltpond Municipal Hospital.

Malaria continues to be the number one public health problem in the Central Region. It accounted for about 54% of all OPD visits in the region in the year 2009. Is it followed by upper respiratory tract infection, which accounted for 9.1% of all cases Hypertension which used not to be in the top ten is gradually gaining grounds in the top 10 it is not occupying the fifth position in the top ten.

Table 1.1: Trends of Cause of OPD Visits, 2007 – 2009

Disease	No. of cases (2007)	No. of cases (2008)	No. of cases (2009)
Malaria	570,315 (45.6%)	446,075 (44.7%)	462,606 (54%)
Upper resp. tract inf.	112,377 (9%)	49,136 (7.6%)	78,000 (9.1%)
Diseases of skin	57,425 (4.6%)	35,206 (5.5%)	54,115 (6.3%)
Diarrhoea diseases	0 (0.0%)	21,081 (3.3%)	42,043 (4.9%)
Hypertension	55,126 (4.4%)	20,008 (3.1%)	31,761 (3.7%)
Rheumatism/joint pains	35,108 (2.8%)	16,464 (2.6%)	24,249 (2.8%)
Anaemia	32,327 (2.6%)	11,531 (1.8%)	20,472 (2.4%)
Accidents	22,572 (1.8%)	9,451 (1.5%)	13,564 (1.6%)
Intestinal Worms	0 (0.0%)	7,625 (1.2%)	13,480 (1.6%)
Acute Eye Infection	14,131 (1.4%)	7,390 (1.2%)	12,588 (1.5%)
All Others	217,394 (17.4%)	117,354 (27.6%)	103,157 (12.1%)

Source: RHD, Central Region, 2009.

1.2 Study area profile

This study covered all the recorded attendance experienced at the Saltpond Government Hospital for the period 2002 to 2012 as can be seen in appendix A. It is located at Saltpond, the Municipal capital of Mfantseman Municipality of the Central Region of Ghana. The Hospital was built in 1920 and it is one of the eight main hospitals in the Region. The Hospital serves a wide range of areas namely Mfantseman West and East and other surrounding districts like Ajumako-Enyan-Essiam District and some parts of the Gomoa District in the Central Region. Since the hospital is the only government hospital in the Municipality, all other health posts and clinics cases in are submitted to it (GSS, 2010).

The Hospital has three consulting rooms with two medical officers and one medical assistant. There are five wards in the hospital; maternity, male surgical, female surgical, pediatric and emergency ward. The hospital has approximately one hundred and eighty beds including cots. The other departments or units in the hospital include the Central OPD which functions in a capacity as the gateway of the hospital. In here, the patients are directed to the various consultants for consultation and treatment. It is in this same department that medical charges are paid. The hospital also has laboratory, x-ray, injection and dressing room, chest clinic, stores, records and statistics office, catering and a laundry service. Geographically, the Mfantseman Municipality is located along the Atlantic coastline of the Central Region of Akanland. The Municipal is bounded to the West and Northwest by Abura-Asebu-Kwamankese District, to the East by Gomoa District and to the South by the Atlantic Ocean.

1.3 Problem statement

According to the Ghana Health Service (2011), OPD attendance continues to increase nationally with a current OPD per capita of 0.98 in the year 2010 to 1.07 in the year 2011. This increased cannot, however, be entirely attributed to full satisfaction of clients about quality services provided at the facilities. More so, the GHS revealed that the proportion of OPD attendance by insured clients increased from 55.81% in 2010 to

82.11% in 2011, OPD per capita increased from 0.98 in 2010 to 1.07 in 2011, with CHPS contributing approximately 5% to the total OPD attendance countrywide. There has been a corresponding progressive and significant increase in IGF from increasing attendance of insured clients at GHS facilities. In 2011, attendance of insured clients at GHS facilities contributed to more than 80% of their total IGF after suffering a dip in 2010 (72%) in comparison to 77.9% in 2009.

This alarming statistics continues to be a major concern for key stakeholders (like Government, hospital administrators, doctors, nurses and pharmacists among others) in the health sector of the country in the face of scary population-to-doctor and population-to-nurse ratios of 10,032: 1 and 1240: 1 respectively. Not many studies have been conducted on OPD attendance in the country. Moreover, no statistical model is available for forecasting OPD cases for proper hospital planning and management. These among other reasons necessitated this study.

1.4 Objectives of the study

The objectives of the study are:

1. to identify the best fit times series model to the OPD attendance data,
2. to forecast attendance at the hospital for the next 5 years on monthly bases.

1.5 Methodology

In forecasting OPD attendance, a time series data is required. Therefore, secondary data would be obtained from the Records and Statistics Office of the Saltpond Municipal

Hospital from 2002 to 2012 on monthly bases. The statistical tools include tables, graphs, Moving Average (MA), Autoregressive (AR), ARIMA and seasonal Autoregressive Integrated Moving Average (SARIMA) model. A time series is a set of statistics, usually collected at regular intervals. Time series data occur naturally in many application areas like the following:

1. Economics - e.g. monthly data for unemployment, hospital admissions, etc.
2. Finance - e.g. daily exchange rate, a share price, etc.
3. Environmental - e.g. daily rainfall, air quality readings.
4. Medicine - e.g. ECG brain wave activity every 2−8 second.

The methods of time series analysis pre-date those for general stochastic processes and Markov Chains. The aims of time series analysis are to describe and summarize time series data, fit low-dimensional models, and make forecasts. We write our real-valued series of observations as $\ldots X_{-2}, X_{-1}, X_0, X_1, X_2, \ldots$ a doubly infinite sequence of real-valued random variables indexed by Z. In gathering literature on the subject, the researcher consulted the University of Cape Coast's Library, GHS Annual Reports (2002 – 2012), and the internet. The SPSS (version 17.0), and MINITAB (version 14.0) software were used for the analysis of data.

1.6 Justification

The ever increasing patronage of the Saltpond Municipal Hospital requires that adequate preparations in terms of personnel and logistics are made in advance. The findings of the study would therefore help the Hospital Management to adequately prepare for the large number of prospective patients. This is likely to help them make advanced plans in terms of manpower and logistical requirements for a better service delivery to the satisfaction and expectations of clients. Again, the Ghana Health Service can adapt same model for nationwide forecasting of OPD attendance cases.

Additionally, the Government of Ghana can also use the findings to review its financial commitment and contribution to the various health facilities throughout the country. Lastly, this study will add to the existing literature for academic purposes.

1.7 Scope and Limitation

The thesis is restricted to the objectives of the research. Research work is characterized by some constraints. Some of these setbacks include time constraints and the difficulties in obtaining relevant materials on the topic.

1.8 Thesis organisation

Chapter one is made up of introduction, which comprises the background of the study, study areas, problem statement and objectives of the study. It also presents the justification and limitations of the study. Chapter two highlights related literature on the topic with ideas of different authors whose findings have been defined in relation to the topic under study.

The third chapter focuses on methodology in the light of mathematical and statistical tools that are relevant to the analyses of the data. Basically, the study seeks to use time series model for the analyses. Chapter four deals with the data collection and analysis, and the findings from the application of the various time series models. The final chapter consists of summary, conclusion and recommendations.

CHAPTER TWO

LITERATURE REVIEW

In this section, there is a review of the work of several authors concerning concept definitions and various researches done to uncover the academic work. Researches, empirical work and authors' opinion are looked at.

2.1 Previous Studies on Outpatient Departments (OPDs) in Ghana

2.1.1 The Nature of Cases Reported at OPDs

Goka (2011) conducted a study about diseases reported at the outpatient departments (OPD) in the Greater Accra Region. The objectives of his study was to find the trend of the various diseases reported at the Outpatient Departments, forecast 2007 reported diseases and finally, to identify the most reported disease(s) within the period. Data were obtained from the Adabraka Polyclinic in Accra, Ghana which is the coordinating centre for the region's health statistics. The data covered the period 1996 to 2006. Time series (trend) analysis was the main statistical technique used in the study. It was found that malaria constituted half of all cases reported at OPD each year. Malaria, upper respiratory tract infection and skin disease formed an overwhelming majority (about 52.4%) of diseases reported at the OPD each year. It was also found that all top ten diseases exhibited upward trends. Trend analysis of these diseases yielded various forecasted values for 2007.

Similarly, in a study dubbed "Cost of care in public health facilities – a comparative study in Kumasi" by Amankwah-Kumi (2002), he stated that the Government of Ghana has the responsibility of ensuring that all Ghanaians enjoy good quality health at an affordable cost to clients. He continued that in pursuance of this goal, the Government has invested in resourcing and expanding the health sector. However, the Ministry of Health has realised that despite these efforts, utilisation of health facilities has been low across the country. The five-year programme of work objectives for the period 1997 – 2001 therefore focused, among others, on means to optimise utilisation of health services to improve the Ghanaian health status. The study was conducted at Manhyia and Kumasi South hospitals to compare cost of care and how cost affects the OPD utilisation level of the two hospitals. Specifically, to compare the cost of managing the top four diseases as reported by the Metropolis Health Directorate in the two hospitals. Data were collected from clients in both hospitals and service providers using two instruments: exit poll interview and key informant interviews respectively. For these instruments, 50 clients and 15 service providers from each hospital were selected respectively. The clients were interviewed on their socio-economic background; the process they go through in receiving care and the affordability of the cost. Providers were also interviewed on level of OPD attendance; their knowledge of treatment procedures and how services are costed. The result gave a view of the level of utilisation of clinical services at both hospitals, the perception of treatment procedures by prescriber and affordability of cost of care by clients. Reference was made to related secondary data as well as hospital records. Analysis was undertaken with the aid of the EPI-INFO 6.04.

The findings showed that although the level of OPD attendances at both hospitals were low compared to the National target and had seen a general decline between 1997 and 1999, observed utilisation in terms of OPD attendance had not experienced a significant deviation from the expected utilisation level of 0.3 per capita OPD attendance set by the Ministry of Health. However, a longer series of OPD attendance data is required to determine the exact trend. The respondents at Manhyia Hospital averagely earn more than respondents at Kumasi South hospital and they found cost of care more affordable, though cost of managing three out of the four top diseases is higher at Manhyia Hospital. Forty per cent of respondents at Kumasi South Hospital said cost of care was not affordable whilst only 18% of respondents at Manhyia Hospital said it was not affordable. It was also found out that prescribers of both hospitals mostly stick in the treatment procedures and over 80% of their prescriptions are from EDL It was recommended that a further study should be done after five to ten years to determine the long-term trend of utilisation in order to undertake any corrective measures. Service charges should also be made more affordable to the poorest section of the communities by putting in place an affordable insurance scheme or making exemptions for paupers easier to access.

Additionally, Nyako (2002) conducted a similar study on the topic: “An assessment of utilisation of clinical services at the La Polyclinic.” He indicated that in pursuance of its goal of improving the health of all Ghanaians, the Government has invested in resourcing and expanding the health sector. The Ministry of Health has realised that despite these efforts, utilisation of health facilities has been low across the country. The Five-year

Programme of Work objectives for the period 1997–2001 therefore focused, among others, on means to optimise utilisation of health services to improve the Ghanaian health status. The La Polyclinic has undergone changes in terms of massive rehabilitation, resourcing, and growth with regard to the mix of services offered since its inception. With this, it was expected that the La Polyclinic would play a competent role in reducing the flow of patients to the Regional and Teaching Hospitals.

However, a preliminary comparison of OPD attendance per capita for 1999 (0.18) of the La Polyclinic with the 1997 national target for per capita OPD attendance (0.30) reveals that attendance is low. This means that notwithstanding all the investment, OPD attendance is still not up to par. This may be an indication of the general level of utilisation of clinical services in Ghana. The study was conducted in the Kpeshie sub-district to assess the level and pattern of utilisation of clinical services at the La Polyclinic. Specifically, to assess the level of utilisation, identify factors and client perceptions that have led to that level and to recommend any necessary ways to optimise utilisation levels. Data collected from clients of La Polyclinic on the grounds, adult community members in the Kpeshie sub district and opinion leaders within the Kpeshie sub district and service providers using four instruments: exit poll interview, community survey, focus group discussion and key informant interviews respectively, to gather data. For these instruments, 100, 100, 8 and 6 respondents were selected respectively. These groups were interviewed on issues affecting utilisation of clinical services at the La Polyclinic.

The results of the study gave a view of the level and distribution of use of clinical services, the perception of the clients and providers about the provision of clinical care and factors influencing utilisation of clinical care and the major factors influencing decision to use the La Polyclinic for clinical care. Reference was made to related secondary data as well as hospital records. Analysis was undertaken with the aid of the EPI-INFO 2000 software. The findings showed that although the level of OPD attendance at the La Polyclinic was low compared to the national target and had seen a general decline between 1993 and 1999; observed utilisation in terms of OPD attendance had not experienced a significant deviation from the expected utilisation level of 0.20 per capita OPD attendance and therefore, utilisation was presently optimal. However, a longer series of OPD attendance data is required to determine the exact trend. Non-users had significantly higher income than users and were more intolerant to long waiting time and harsh staff attitude.

It was recommended that a further study should be done after five to ten years to determine the long-term trend of utilisation in order to undertake any corrective measures. Additionally, the Saltpond District Health Management Team (SDHMT), District Health Management Team (DHMT) and Regional Health Administration (RHA) should collaborate to reduce waiting time and poor staff attitude by ensuring that staffs are in adequate numbers, motivated, well-trained to relate better to clients and more efficient in time management. Service charges should also be made more affordable to the poorest section of the communities by putting in place an affordable insurance scheme or making exemptions for paupers easier to access.

2.1.2 Factors Influencing OPD Attendance

The outpatient Department (OPD) is the eye of every health facility and the patients judge the standard of the health facility by the care they receive at the OPD. This could reflect in the utilisation of the health facility. Past studies in Ghana have shown that outpatient attendance continues to increase proportionately with the increase in the population of Ghana. However, the contrary was observed by the DHMT of Ho District, which found out that the OPD attendance in the district has been declining for the past four years. The DHMT has attached importance to this falling OPD attendance, therefore requested that the factors influencing OPD attendance and their extent of influence be reviewed. This will enable them provide recommendations to improve attendance in public health facilities. The study was a descriptive cross-sectional study. The study units were selected purposively using extreme case strategy and the respondents were selected using simple random sampling. Structured Interview guide was used to interview the selected individuals. The data were coded and analysed using SPSS. Part of the data was also analysed manually.

The study results revealed that public health facilities were geographically and financially accessible to the people. However, a reasonable proportion of the respondents live more than 1km from where they seek treatment. Socio-cultural values and beliefs had no influence on OPD attendance, as there were no social or cultural values/beliefs preventing respondents from using orthodox medicine. In addition, outreach services and drug peddlers' activities had no influence. Although, there were different types of health providers in the communities studied, orthodox medicine was preferred. Majority of the respondents (88.4%) were satisfied with the services provided at the public health facilities. Meanwhile, a reasonable proportion (41.3%) was not issued with receipts covering their expenses at the health centre level. The users of the district hospital suggested that staff change their attitude towards patients by showing them respect and patience. Based on the results, the following conclusions were drawn.

The main causes of falling OPD attendance are facility-related. These were poor attitude of health staff towards patients (14.4%), lack of valid receipts covering expenses (41.3%) and health staff not at post (26.5%). The recommendations among others provided to DHMT include close supervision to ensure that valid receipt are issued to patients, drugs as well as other materials are used for their purpose, management should encourage staff to show respect, patience and care for patients.

Patavegar et al. (2012) also studied the satisfaction of OPD patients in tertiary care hospital and to know the relationship between various determinants and OPD patient's satisfaction. The present cross-sectional study was conducted among 450 patients attending the outpatient departments (OPDs) of Sassoon General Hospital Pune during 6 months period. Systemic random sampling was used for patient selection. Maximum number of patients i.e. 197(43.78%) were in the age group of 49 and above. About 61% patients were females. About cleanliness of waiting area 44.5% patients were found unsatisfied. About explanation of treatment by pharmacist 77% patients were satisfied. 91% patient said that OPD timings were convenient. One-hundred and seventy-six (39.12%) patients had to wait less than 30 minutes before consulting doctor. According to the patient's opinion, the study showed good satisfaction with respect to registration services, doctor services, nurse services, laboratory services and pharmacy staff services.

Also, Jawahar (2007) studied Out Patient Satisfaction at a Super Specialty Hospital in India Patients and staff satisfaction is an important component of the health care industry in this competitive modern era. In the hospital, the Outpatient Department is often called "Shop Window." Patients' satisfaction leads to drift in both new and old patients, which hinders the sustainability of any hospital in long run. This study was conducted to know the satisfaction level of patients and also get a feedback about the services provided in the outpatient departments.

The patients were randomly selected and a questionnaire was developed to evaluate patient satisfaction about the outpatient department services, logistic arrangement in the outpatient departments, waiting time, facilities, perception about the performance of staff, appointment system, behaviour of staff, support service and any other suggestions of patients. Out of 200 patients surveyed, 90-95% of patients were satisfied with the service offered in the hospital. This study also showed that some of the patients waiting time were prolonged and the friendliness of the nursing staff needs to be improved.

2.2 Effects of New Service Charges on OPD Attendance

Osauga and Nordberg (1993) studied the effects of new service charges on hospital attendance in Kenya. They said that Kenyans have long enjoyed free outpatient health care at government facilities while paying for admission and for child delivery. In December 1989 user charges were introduced also for out-patient care at hospitals and health centres. This before-and-after study of one rural hospital, two health centres and two dispensaries in rural Kenya shows major and statistically significant early drops in outpatient attendance at the hospital (28%) and at the health centres (50 and 43%) followed by a slow increase during the following months. There was a modest, not significant, decline also at the dispensaries (14 and 7%) and in demand for services unaffected by the new fees and charges.

2.3 OPD, Non-Attendance Rates, and Reducing Non-Attendance Rates

According to Killaspy et al. (2000), psychiatric clinics have high non-attendance rates and failure to attend may be a sign of deteriorating mental health. They investigated why psychiatric out-patients fail to attend, and the outcome of attenders and non-attenders using prospective cohort study of randomly selected attenders and non-attenders at general adult psychiatric out-patient clinics. Subjects were interviewed at recruitment and severity of mental disorder and degree of social adjustment were measured. Six and 12 months later their engagement with the clinic and any psychiatric admissions were ascertained. The result indicated that of the 365 patients included in the study, 30 were untraceable and 224 consented to participate.

Follow-up patients were more psychiatrically unwell than new patients. For follow-up patients, non-attenders had lower social functioning and more severe mental disorder than those who attended. At 12-month follow-up patients who missed their appointment were more likely to have been admitted than those who attended.

They concluded that those who miss psychiatric follow-up out-patient appointments are more unwell and more poorly socially functioning than those who attend. They have a greater chance of drop-out from clinic contact and subsequent admission. In order to identify strategies of reducing hospital non-attendance, Stone et al. (1999) studied on the subject: “Reducing non-attendance at outpatient clinics.” They opined that outpatient non-attendance is a common source of inefficiency in a health service, wasting time and resources and potentially lengthening waiting lists. A prospective audit of plastic surgery outpatient clinics was conducted during the six months from January to June 1997, to determine the clinical and demographic profile of non-attenders. Of 6095 appointments 16% were not kept.

Using the demographic information, we changed our follow-up guidelines to reflect risk factors for multiple non-attendances, and a self-referral clinic was introduced to replace routine follow-up for high risk non-attenders. After these changes, a second audit in the same six months of 1998 revealed a non-attendance rate of 11% – i.e. 30% lower than before. Many follow-up appointments are sent inappropriately to patients who do not want further attention. This study, indicating how risk factor analysis can identify a group of patients who are unlikely to attend again after one missed appointment, may be a useful model for the reduction of outpatient non-attendance in other specialties.

Reti (2003) conducted a study on the topic: “Improving outpatient department efficiency: A randomised controlled trial comparing hospital and general-practice telephone reminders.” The study aimed to ascertain whether or not telephone reminders reduce nonattendance at hospital outpatient clinics and whether telephone reminders from general practitioners are more effective than those made from hospitals. Outpatient department appointments for three general practitioners (GPs) over a three-month period, were randomised into three groups: 'Hospital', 'GP', and 'Control'. Patients in the Hospital and the GP groups were reminded of their appointment by telephone 24 hours beforehand, by a hospital waiting-list clerk or their general practitioner respectively.

Information was recorded on appointment awareness and subsequent attendance history. A total of 109 patients were included in the study. The three study groups had 'no show' rates of 3% (GP), 8% (Hospital), and 27% (Control). The combined 'no show' rate for the groups reminded by telephone was 5%. The combined telephone-reminded group was statistically different from the Control group ($p = 0.004$). There was no statistical difference between the GP group and the Hospital group ($p=0.764$). In this study, telephone reminders significantly decreased outpatient department 'no show' rates. The source of the telephone reminder made no difference to non-attendance.

2.4 Utilisation of Health Service

Out Patient Department (OPD) attendance recorded by the District Health Directorate during the year 2005 was 30,879 compared to 26,631 in 2004. The trend of OPD attendance for four years is as indicated in the table 1.29 below. The District recorded OPD per capita of 0.33 compared to 0.29 in 2004 and 0.24 for 2003. There is an improvement which will be built upon in the subsequent years. Uptake of ante care in the district is 66.8 per cent. Postnatal coverage is 32 per cent, which is very low. This is perhaps due to transportation difficulties in traveling with babies to the health facilities DPT coverage in the district is 88.5 per cent. Proportion of supervised delivery in the district is 36 per cent. This is quite low, and is perhaps due to the people's inability to afford the delivery charges. The small proportion of supervised delivery in the district is also due to cultural beliefs.

A study by Birk et al. (2011) on patients' choice of hospital has focused on inpatients' rather than outpatients' choice of provider. We have investigated Danish outpatients' awareness and utilisation of freedom of choice of provider; which factors influence outpatients' choice of hospital, and how socio-demographic variables influence these factors in a single uptake area, where patients were free to choose any public hospital, where care was provided free at the point of delivery, and where distance to the closest hospitals were short by international standards.

Retrospective questionnaire study of 4,232 outpatients referred to examination, treatment, or follow-up at one of nine somatic outpatient clinics in Roskilde County in two months of 2002, who had not been hospitalised within the latest 12 months. The patients were asked, whether they were aware of and utilised freedom of choice of hospital. Fifty-four percent (2,272 patients) filled in and returned the questionnaire. Forty-one percent of respondents were aware of their right to choose, and 53% of those patients utilised their right to choose. Awareness of freedom of choice of provider was reported to be especially high in female outpatients, patients with longer education, salaried employees in the public sector, and in patients referred to surgical specialties. Female outpatients and students were especially likely to report that they utilised their right to choose the provider.

Short distance was the most important reason for outpatients' choice, followed by the GP’s recommendations, short waiting time, and the patient's previous experience with the hospital. Outpatients’ awareness and utilisation of free choice of health care provider was low. Awareness of freedom of choice of provider differed significantly by specialty and patient’s gender, education and employment. Female patients and students were especially likely to choose the clinic by themselves. Most outpatients chose the clinic closest to their home, the GP’s recommendation and short waiting time being the second and third most important factors behind choice.

2.4.1 Serving More Patients Better

In a study conducted at Pushpanjali Crosslay Hospital (PCH) in North India. The study revealed that the PCH has also witnessed a sharp increase in the total number of OPD cases in the year 2011-2012, with over 1,00,000 cases as compared to 66,558 in 2010–2011 and 30,393 OPD cases in the year 2009-2010. The IPD cases in 2011-2012 rose to over 12,000 as compared to 10,262 in 2010-2012 and 4,984 in the year 2009-2010. Analysis of the extensive database of patient feedback collected systematically from all patient touch points indicates a rising quality of care, especially in the departments of nursing, housekeeping and dietetics. Translation of the feedback to efficient service delivery is achieved by regular meetings with management and department heads followed by staff training.

2.5 National OPD Per Capita

Outpatient department visits (OPD) per capita in Ghana recorded progressive increases from 2009 and 2010 to 2011 as shown in Table 2.1. The annual target for 2011 was 1.0, and following the three-year trend, OPD per capita figures increased from 0.81 (2009), 0.98 (2010) to 1.07 (2011) an increase is being driven in large part by increased registration on the National Health Insurance Scheme (NHIS). This is evident in the increased proportion of insured clients reporting at the OPD – 44.2% (2009), 55.8% (2010) and 82.0% (2011). This increase in OPD per capita was also due to the ongoing nationwide expansion and improvement of healthcare and service infrastructure at various service delivery points. Thus, gradually closing the access gaps, both financial and geographical, which are the persisting challenges to healthcare in Ghana. Guidelines to improve the referral system, gatekeeper system and free-maternal care policy have been developed and are been disseminated within the Greater Accra Region.

Table 2.1: 3-Year Trend in OPD per Capita 2009 – 2011

Region	2009	2010	2011
Ashanti	0.89	1.04	1.18
Brong-Ahafo	1.15	1.25	1.38
Central	0.71	0.81	0.83
Eastern	0.95	1.04	1.19
Greater Accra	0.51	0.97	0.66
Northern	0.53	0.54	0.62
Volta	0.69	0.76	0.87
Upper East	1.37	1.47	1.42
Upper West	0.72	0.91	1.10
Western	0.99	1.16	1.38
National (average)	**0.81**	**0.98**	**1.07**

Source: GHS, 2011.

2.5.1 Utilisation of Hospital Services in the Central Region

Facility utilisation in 2008 showed a marginal gain compared to 2007 with Out-patient attendance per capita increasing from 0.63 in 2007 against 0.68 in 2008 having remained the same from 2003 to 2006 as shown in the Figure 2.1. Out of the total attendance of 1,242,196, 67% were insured and 33% non-insured with the NHIS. District distributions showed that 11 out of the 17 districts recorded OPD per capita of 0.5 and above with Cape Coast recording the highest of 1.98 while Upper Denkyira West and Gomoa East all recorded OPD per capita of 0.3.

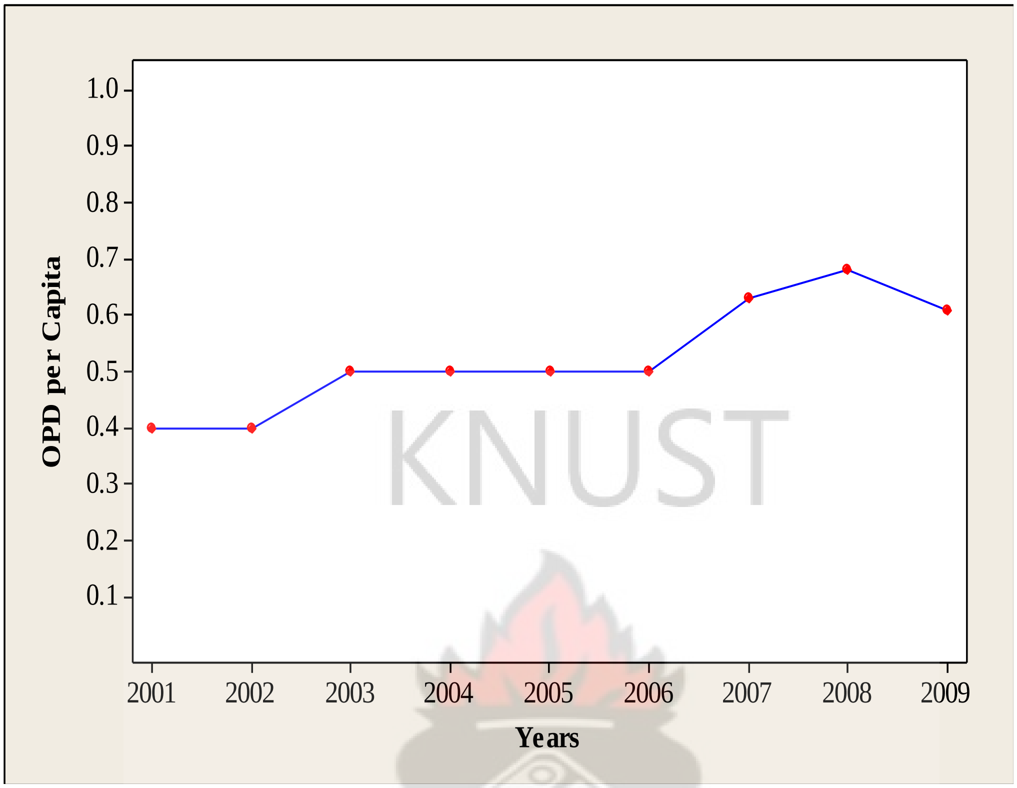

Figure 2.1: Trend in OPD Attendance per Capita, 2001 – 2009, CR.

2.6 Application of Times Series Analysis

Time series analysis has been applied in a several fields of study, namely health, economic, agriculture, meteorology, and business among others. For instance, Goka (2007) applied time series analysis on the various diseases reported at the Outpatient Departments, forecast 2007 reported diseases and finally, to identified the most reported disease(s) within the period. Data was obtained from the Adabraka Polyclinic which is the coordinating center for the region's health statistics. The data covered the period 1996 to 2006.

Trend analysis was the main statistical technique used in the study. It was found that malaria constituted half of all cases reported at OPD each year. Malaria, upper respiratory tract infection and skin disease formed an overwhelming majority (about 52.4%) of diseases reported at the OPD each year. It was also found that all top ten diseases exhibited upward trends. Trend analysis of these diseases yielded various forecasted values for 2007.

Also, Miyake et al. (2009) employed time-series analysis on the seasonal variation in liver function tests. In their study, they examined the seasonal variation in liver function tests using recently described data-mining methods.

The 'latent reference values' of aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase (ALP), gamma-glutamyltransferase (gamma GT), cholinesterase (ChE) and total bilirubin (T-Bil) were extracted from a seven-year database of outpatients (aged 20-79 yr; comprising approximately 1,270,000 test results). After calculating the monthly means for each variable, the time-series data were separated into trend and seasonal components using a local regression model (Loess method). Then, a cosine function model (cosinor method) was applied to the seasonal component to determine the periodicity and fluctuation range.

A two-year outpatient database (215,000 results) from another hospital was also analysed to confirm the reproducibility of these methods. They found that the serum levels of test results tended to increase in the winter. The increase in AST and ALT was about 6% in men and women, and was greater than that in ChE, ALP (in men and women) and gamma GT (in men). In contrast, T-Bil increased by 3.6% (men) and 5.0% (women) in the summer. The total protein and albumin concentrations did not change significantly. AST and ALT showed similar seasonal variation in both institutions in the comparative analysis.

CHAPTER THREE

METHODOLOGY

This chapter examines thoroughly the basic definitions and concepts of time series analysis, assumptions, conditions, principles and processes involved in the application of moving averages (MA), autoregressive (AR), autoregressive moving averages (ARMA), and autoregressive integrated moving average (ARIMA).

3.1 Basic Concepts and Definitions of Time Series

3.1.1 Time Series Analysis

Time series analysis comprises methods or processes that breakdown a series into components and explainable portions that allows trends to be identified, estimates and forecasts to be made. Basically time series analysis attempts to understand the underlying context of the data points through the use of a model to forecast future values based on known past values. Such time series models include MA, AR, ARIMA, GARCH, TARCH, EGARCH, FIGARCH, CGARCH, ARIMA, etc but the main focus of this study is based on MA, AR, ARMA, and ARIMA models.

3.1.2 Lag

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

3.1.3 Differencing

Differencing simply means subtracting the value of an earlier observation from the value of a later observation. Calculating differences among pairs of observations at some lag to make a non-stationary series stationary. There are possible shifts in both the mean and the dispersion over time for this series. The mean may be edging upwards, and the variability may be increasing. If the mean is changing, the trend is removed by differencing once or twice. If the variability is changing, the process may be made stationary by logarithmic transformation. Differencing the scores is the easiest way to make a non-stationary mean stationary (flat). The number of times you have to difference the scores to make the process stationary determines the value of *d*. If *d* C0, the model is already stationary and has no trend. When the series is differenced once, *d* C1 and linear trend is removed. When the difference is then differenced, *d* C2 and both linear and quadratic trend are removed. For non stationary series, *d* values of 1 or 2 are usually adequate to make the mean stationary.

3.1.4 Stationary and Non-stationary Series

Stationary series vary around a constant mean level, neither decreasing nor increasing systematically over time, with constant variance. Non-stationary series have systematic trends, such as linear, quadratic, and so on. A non-stationary series that can be made stationary by differencing is called “non-stationary in the homogenous sense.”

Stationarity is used as a tool in time series analysis, where the raw data are often transformed to become stationary. For example, economic data are often seasonal or dependant on a non-stationary price level. Using non-stationary time series produces unreliable and spurious results and leads to poor understanding and forecasting. The solution to the problem is to transform the time series data so that it becomes stationary. If the non-stationary process is a random walk with or without a drift, it is transformed to stationary process by differencing. Differencing the scores is the easiest way to make a nonstationary mean stationary (flat). The number of times you have to difference the scores to make the process stationary determines the value of *d*. If d = 0, the model is already stationary and has no trend. When the series is differenced once, d=1 and linear trend is removed. When the difference is then differenced, d =2 and both linear and quadratic trend are removed. For non-stationary series, d values of 1 or 2 are usually adequate to make the mean stationary. If the time series data analysed exhibits a deterministic trend, the spurious results can be avoided by detrending. Sometimes the non-stationary series may combine a stochastic and deterministic trend at the same time and to avoid obtaining misleading results both differencing and detrending should be applied, as differencing will remove the trend in the variance and detrending will remove the deterministic trend.

A non-stationary process with a deterministic trend becomes stationary after removing the trend, or detrending. For example, $Y_t = \alpha + \beta_t + \varepsilon_t$ is transformed into a stationary process by subtracting the trend β_t: $Yt - \beta_t = \alpha + \varepsilon_t$. No observation is lost when detrending is used to transform a non-stationary process to a stationary one. Non-stationary data, as a rule, are unpredictable and cannot be modelled or forecasted. The results obtained by using non-stationary time series may be spurious in that they may indicate a relationship between two variables where one does not exist. In order to receive consistent, reliable results, the non-stationary data needs to be transformed into stationary data. In contrast to the non-stationary process that has a variable variance and a mean that does not remain near, or returns to a long-run mean over time, the stationary process reverts around a constant long-term mean and has a constant variance independent of time.

3.2 Components of Time Series

A vital step in choosing appropriate modeling and forecasting procedure is to consider the type of data patterns exhibited from the time series graphs of the time plots. The sources of variation in terms of patterns in time series data are mostly classified into four main components:

(i) Horizontal – when data values fluctuate around constant value

(ii) Trend – when there is long term increase or decrease in the data

(iii) Seasonal – when a series is influenced by seasonal factors and recurs on a regular periodic basis.

(iv) Cyclic – when the data exhibit rises and falls that are not of a fixed period.

3.2.1 The Trend (d)

The trend is simply the underlying long term behavior or pattern of the data or series. The Australian Bureau of Statistics (ABS, 2008) defined trend as the ‘long term’ movement in a time series without calendar related and irregular effects, and is a reflection of the underlying level. It is the result of influences such as population growth, price inflation and general economic changes. A model with two trend terms (*d*C2) has to be differenced twice to make it stationary. The first difference removes linear trend, the second difference removes quadratic trend, and so on.

3.2.2 Seasonal Variation (S)

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

1. A seasonal subseries plot is a specialised technique for showing seasonality.
2. Multiple box plots can be used as an alternative to the seasonal subseries plot to detect seasonality.
3. The autocorrelation plot can help identify seasonality.

3.2.3 Cyclical Variations (C)

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

3.2.4 Irregular Variations (I)

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

3.3 Common Assumptions in Time Series Techniques

A common assumption in many time series techniques is that the data are stationary. A stationary process has the property that the mean, variance and autocorrelation structure do not change over time. Stationarity can be defined in precise mathematical terms as

1. The mean $\mu(t) = E(yt)$
2. The variance $\sigma^2(t) = Var(yt) = \gamma(0)$
3. The autocovariance $\gamma(t_1, t_2) = cov(yt_1, yt_2)$

Hence a time series is said to be strictly stationary if the joint distribution of any set of n observations $(t_1, t_2) = cov(yt_1, yt_2)$ is the same as the joint distribution of $y(t_1),\ y(t_2), y(t_3) \ldots y(t_n)$ for all n and k. If the time series is not stationary, we can often transform it to stationary with one of the following techniques:

1. We can difference the data. That is, given the series Z_t, the differenced data will contain one less point than the original data. Although you can difference the data more than once, one difference is usually sufficient.
2. If the data contain a trend, we can fit some type of curve to the data and then model the residuals from that fit. Since the purpose of the fit is to simply remove long term trend, a simple fit, such as a straight line, is typically used.
3. For non-constant variance, taking the logarithm or square root of the series may stabilize the variance. For negative data, you can add a suitable constant to make the entire data positive before applying the transformation. This constant can then be subtracted from the model to obtain predicted (i.e., the fitted) values and forecasts for future points.

3.4 Autocorrelation Function (ACF)

Autocorrelation refers to the correlation of a time series with its own past and future values. Autocorrelation is also sometimes called "lagged correlation" or "serial correlation", which refers to the correlation between members of a series of numbers arranged in time. The pattern of autocorrelations in a time series at numerous lags; the correlation at lag 1, then the correlation at lag 2, and so on. Correlations among sequential scores at different lags. The lag 1 autocorrelation coefficient is similar to correlation between the pairs of scores at adjacent points in time, rY_t, Y_{t-1} (e.g., the pair at time 1 and time 2, the pair at time 2 and time 3, and so on). The lag 2 autocorrelation coefficient is similar to correlation between the pairs of scores two time periods apart, rY_t, Y_{t-2} (e.g., the pair at time 1 and time 3, the pair at time 2 and time 4, and so on). Positive autocorrelation might be considered a specific form of "persistence", a tendency for a system to remain in the same state from one observation to the next. Three tools for assessing the autocorrelation of a time series are:

1. The time series plot,
2. The lagged scatter plot, and
3. The autocorrelation function.

An important guide to the persistence in a time series is given by the series of quantities called the sample autocorrelation coefficients, which measure the correlation between observations at different times. The set of autocorrelation coefficients arranged as a function of separation in time is the sample autocorrelation function, or the acf.

The first- order autocorrelation coefficient is the simple coefficient of the first N–1 observations, t=1, 2,…, N-1 and the next N - 1 observations, X_t, t=2,3,…, N. The correlation between X_t and X_{t+1} is given by,

$$r_1 = \frac{\sum_{t=1}^{N-1}(X_t - \bar{X}_1)(X_{t+1} - \bar{X}_2)}{[\sum_{t=1}^{N-1}(X_t - \bar{X}_1)^2]^2[X_{t+1} - \bar{X}_2]^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.1)$$

Where $\bar{x}1$ is the mean of the first N – 1 observations and is the mean of the last N-1 observation. As the correlation coefficient given above measure correlation between successive observations it is called the autocorrelation coefficient or serial correlation coefficient. For *N* reasonably large, the difference between the sub-period means $\bar{X}_1$ and $\bar{X}_2$ can be ignored and r_1 can be approximated as by

$$r_1 = \frac{\sum_{t=1}^{N-1}(X_t - \bar{X})(X_{t+1} - \bar{X})}{\sum_{t=1}^{N}(X_t - \bar{X})^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.2)$$

Equation (3.2) can be generalised to give the correlation between observations separated by k years:

$$r_k = \frac{\sum_{t=1}^{N-k}(X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^{N}(X_t - \bar{X})^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.3)$$

The quantity is called the autocorrelation coefficient at lag k. The plot of the autocorrelation function as a function of lag is also called the correlogram. The autocorrelation function can be used for the following two purposes:

- To detect non-randomness in data.
- To identify an appropriate time series model if the data are not random.

Autocorrelation plots are formed by:

- Vertical axes: autocorrelation coefficient.
- Horizontal axes: Time lag k= 1,2,3…………
- Confidence band

The confidence band uses the following formula if the autocorrelation plot is used to check for randomness in the data.

$$\frac{\pm z1-{}^{\alpha}/_{2}}{\sqrt{N}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (3.4)$$

Where N is the sample size, *z* is the percent point function of the standard normal distribution and α is the significance level. If autocorrelation plots are also used in the model identification stage for fitting ARIMA models, the confidence band uses the following formula:

$$\pm z_{1-}{}^{\alpha}/_{2}\sqrt{\frac{1}{N}(1+2\sum_{i=1}^{K} y2i)} \ldots\ldots\ldots\ldots\ldots\ldots(3.5)$$

Where *k* is the lag, *N* is the sample size; z is the percent point function of the standard normal distribution and α is the significance level.

3.5 Partial Autocorrelation Function

Partial autocorrelation function measures the degree of association between Y_t and Y_{t+k} when the effect of other time lags on Y are held constant. The partial autocorrelation function PACF denoted by the set of partial autocorrelations at various lags *k* are defined by ($r_{kk,}$ k=1, 2, 3…). The set of partial autocorrelations at various lags *k* are defined by

$$r_{kk} = \frac{rk-\sum_{j=1}^{k-1} rk-1{,}j\ rk-1}{1-\sum_{j=1}^{k-1} rk-1{,}j\ rj} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (3.6)$$

where,

$r_{k,\,j}=r_{k-1,j} - r_{kk}\ r_{k-1,k-1}$ j=1,2……………….k-1

Specifically, partial autocorrelations are useful in identifying the order of an autoregressive model. The partial autocorrelation of an AR (*p*) process is zero at lag *p*+1 and greater. The approximate 95% confidence interval for the partial autocorrelations is at $^{+2}/_{N}$. Partial autocorrelation plots are formed by:

- Vertical axes: partial autocorrelation at coefficient at lag, *k*,
- Horizontal axes: time lag *k* (*k* = 0, 1, 2...)

In addition, 95% confidence interval bands are typically included on the plot.

3.6 Common Approaches to Univariate Time Series

There are a number of approaches to modeling time series. Few of the most common approaches are below:

3.6.1 Decomposition

One approach is to decompose the time series into a trend, seasonal, and residual component. In other words decomposition refers to separating a time series into trend, cyclical, and irregular effects. Decomposition may be linked to de-trending and de-seasonalising data so as to leave only irregular effects, which are the main focus of time series analysis. Triple exponential smoothing is an example of this approach. Another example, called seasonal loess, is based on locally weighted least squares and is discussed by Cleveland (1993).

3.6.2 Autoregressive (AR) Models

An autoregressive model is simply a linear regression of the current value of the series against one or more prior values of the series. The value of *p* is called the order of the AR model. AR models can be analyzed with one of various methods, including standard linear least squares techniques. They also have a straightforward interpretation. A common approach for modeling univariate time series is the autoregressive (AR) model:

$$X_t = \delta + \emptyset_1 X_{t-1} + \emptyset_2 X_{t-2} + \ldots\ldots\ldots \emptyset_p X_{t-p} + A_t \ldots\ldots \quad (3.7)$$

where X_t is the time series, A_t is white noise, and

with μ denoting the process mean. An autoregressive model of order p, denoted by AR(p) with mean zero is generally given by the equation:

$$X_t = \phi_1 X_{(t-1)} + \phi_2 X_{(t-2)} + \phi_3 X_{(t-3)} + \ldots\ldots + \phi_p X_{(t-p)} + \varepsilon_t \ldots\ldots \quad (3.8)$$

Or

$$X_t = (\phi_1 L + \phi_2 L^2 + \phi_3 L^3 + \ldots\ldots + \phi^P L^P) X_t + \varepsilon_t \ldots\ldots \quad (3.9)$$

Where $\phi(L) = \varepsilon_t$

$$\phi(L) = (1 - \phi_1 L + \phi_2 L^2 + \phi_3 L^3 + \ldots\ldots + \phi^P L^P) \ldots\ldots \quad (3.10)$$

Where L, is the lag operator

$\phi_1, \phi_2, \phi_3 \ldots\ldots \phi_p (\phi_{p \neq} 0)$ are the autoregressive model parameters and ε_t is the random shock or white noise process, with mean zero and variance σ_ε^2. The mean of X_t is zero. If the mean, μ of X_t is not zero, replace X_t by $X_t - \mu$. That is

$$X_t - \mu = \phi_1 (X_{t-1} + \mu) + \phi_2 (X_{t-2} + \mu) + \ldots + \phi_p (X_{t-p} + \mu) + \varepsilon_t \ldots\ldots \quad (3.11)$$

Or write

$$X_{t=} \alpha + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots\ldots \phi_P X_{t-p} + \varepsilon_t \ldots\ldots \quad (3.12)$$

Where $\alpha = \mu(1 - \phi_1 - \phi_3 \ldots\ldots, \phi_p)$

In this general case, the ACF damps down and the PACF cuts off after p lags. An AR (p) model is stationary if the roots of $\phi(L) = 0$ all lie outside the unit circle. A necessary condition for stationary is that $r_k=0$ as $k \longrightarrow \infty$.

3.6.3 Moving Average (MA) Models

Moving average model is conceptually a linear regression of the current value of the series against the white noise or random shocks of one or more prior values of the series. The random shocks at each point are assumed to come from the same distribution, typically a normal distribution, with location at zero and constant scale. The distinction in this model is that these random shocks are propagated to future values of the time series. Fitting the MA estimates is more complicated than with AR models because the error terms are not observable. This means that iterative non-linear fitting procedures need to be used in place of linear least squares. MA models also have a less obvious interpretation than AR models. Moving Average (MA) is another common approach for modeling univariate time series models is the moving average (MA) model:

$$X_t=\mu+A_t-\theta_t A_{t-1}-\theta_2 A_{t-2}-\dots-\theta_q A_{t-q} \qquad (3.13)$$

Where, X_t is the time series, μ is the mean of the series, A_{t-i} are white noise, and θ_1, ... , θ_q are the parameters of the model. The value of q is called the order of the MA model.

A moving average model of order q, with mean zero, denoted by MA (q) is generally given by:

$$X_t =\theta_1\varepsilon_{(t-1)}+\theta_2\varepsilon_{(t-2)}+\theta_3\varepsilon_{(t-3)}+\dots+\theta_q X_{(t-q)}+\varepsilon_t \qquad (3.14)$$

Or

$$\theta(L)=(\theta_1 L+\theta_2 L^2+\theta_3 L^3+\dots+\theta_q L^q)X_t+\varepsilon_t \qquad (3.15)$$

Where $X_t = \theta(L)\,\varepsilon_t$

$$\theta(L)=(\theta_1L+\theta_2L^2+\theta_3L^3+\ldots\ldots\ldots+\theta_qL^q)\ldots\ldots\ldots\ldots\ldots\ldots(3.16)$$

Where L, is the lag operator and ε_t is the random shock or white noise process and $\theta_1, \theta_2, \theta_3, \ldots\ldots\ldots\ldots \theta_q$ are the moving average model parameters. An MA (q) is said to be invertible if $\phi(L)$ can be inverted, in other words if it can be expressed as an AR. An MA (q) is invertible if the roots of $\phi(L)=0$ all lie outside the unit circle. A finite AR is always invertible. The random shocks at each point are assumed to come from the same distribution, typically a normal distribution, with location at zero and constant scale. The distinction in this model is that these random shocks are propagated to future values of the time series. Sometimes the ACF and PACF will suggest that a MA model would be a better model choice and sometimes both AR and MA terms should be used in the same model. It is also important to note, however, that the error terms after the model is fit should be independent and follow the standard assumptions for a univariate process. Box and Jenkins popularized an approach that combines the moving average and the autoregressive approaches (Box, Jenkins, and Reinsel, 1994). This resulted in autoregressive moving average model (ARMA). The Box-Jenkins model assumes that the time series is stationary. Box and Jenkins recommend differencing non-stationary series one or more times to achieve stationarity. Doing so produces an ARIMA model, with the "I" standing for "Integrated". This is described in detail below since it is the main method used in the analysis of data in this research.

3.6.4 Autoregressive Moving Average (ARMA) Models

Autoregressive and Moving Average processes can be combined to obtain a very flexible class of univariate processes (proposed by Box and Jenkins), known as ARMA processes. The time series X_t is an ARMA (p, q) process, if it is stationary and

$$X_t=\phi_1X_{(t-1)}+\phi_2X_{(t-2)}+\phi_3X_{(t-3)}+...+\phi_pX_{(t-p)}+\varepsilon_t+\theta_1\varepsilon_{(t-1)}+\theta_2\varepsilon_{(t-2)}+...+\theta_qX_{(t-q)} \quad (3.14)$$

$$X_t=\phi_1X_{t-1}+\theta(L)\varepsilon_t \quad (3.15)$$

$$\phi(L)X_t=\theta(L)\varepsilon_t \quad (3.16)$$

Where θ,ϕ,ε and L are as defined above with $\phi_p \neq 0$ and $\theta_q \neq 0$

An ARMA process is stationary if the roots of $\phi(L)$ all lie outside the unit circle and invertible if the roots of $\theta(L)$ all lie outside the unit circle.

The acronym for an auto-regressive integrated moving average model. The three terms to be estimated in the model are auto-regressive (*p*), integrated (trend—*d*), and moving average (*q*).The ARIMA (auto-regressive, integrated, moving average) model of a time series is defined by three terms (*p, d, q*). Identification of a time series is the process of finding integer, usually very small (e.g., 0, 1, or 2), values of *p, d,* and *q* that model the patterns in the data. When the value is 0, the element is not needed in the model. The middle element, *d*, is investigated before *p* and *q*. The goal is to determine if the process is stationary and, if not, to make it stationary before determining the values of *p* and *q*. Recall that a stationary process has a constant mean and variance over the time period of the study.

3.7 Box-Jenkins ARIMA Process

In statistics and econometrics, and in particular in time series analysis, an autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. The Box–Jenkins methodology, named after the statisticians George Box and Gwilym Jenkins, applies ARIMA models to find the best fit of a time series to past values of this time series, in order to make forecasts. They are applied in some cases where data show evidence of non-stationarity, where an initial differencing step (corresponding to the "integrated" part of the model) can be applied to remove the non-stationarity. The model is generally referred to as an ARIMA (p, d, q) model where p, d, and q are non-negative integers that refer to the order of the autoregressive, integrated, and moving average parts of the model respectively.

3.7.1 Box-Jenkins Modeling Approach

The Box-Jenkins model uses iterative three-stage modeling approach which is:

1. Model identification and model selection: making sure that the variables are stationary, identifying seasonality in the dependent series (seasonally differencing it if necessary), and using plots of the autocorrelation and partial autocorrelation functions of the dependent time series to decide which (if any) autoregressive or moving average component should be used in the model.
2. Parameter estimation using computation algorithms to arrive at coefficients which best fit the selected ARIMA model. The most common methods use maximum likelihood estimation or non-linear least-squares estimation.

3. Model checking by testing whether the estimated model conforms to the specifications of a stationary univariate process. In particular, the residuals should be independent of each other and constant in mean and variance over time (plotting the mean and variance of residuals over time and performing a Ljung-Box test or plotting autocorrelation and partial autocorrelation of the residuals are helpful to identify misspecification). If the estimation is inadequate, we have to return to step one and attempt to build a better model.

3.8 Box-Jenkins Model Identification

3.8.1 Stationarity and Seasonality

The first step in developing a Box–Jenkins model is to determine if the time series is stationary and if there is any significant seasonality that needs to be modeled.

3.8.2 Detecting Stationarity

Stationarity can be assessed from a run sequence plot. The run sequence plot should show constant location and scale. It can also be detected from an autocorrelation plot. Specifically, non-stationarity is often indicated by an autocorrelation plot with very slow decay. Finally, unit root tests provide a more formal approach to determining the degree of differencing such as Kwiatkowski-Phillips-Schmidt-Shin (KPSS) and Phillips-Perron Unit Root Tests are carried out employing the unit root testing procedures of Hamilton (1994). The KPSS test for the null hypothesis of a level stationary against an alternative of unit root together with the Philips-Peron test for the null hypothesis of a unit root against the alternative of a stationary series. The decision rule is that for the KPSS test if the p-value of its test statistic is greater than the critical value, say 0.05, then reject the null hypothesis of having a level stationary series and therefore conclude the alternate hypothesis that it has a unit root. The Philips-Peron Test, on the other hand, test for the null hypothesis of unit root against an alternative hypothesis of stationarity by rejecting the null hypothesis if its *p*-value is less than the critical value chosen.

3.8.3 Differencing to achieve Stationarity

Box and Jenkins recommend the differencing approach to achieve stationarity. However, fitting a curve and subtracting the fitted values from the original data can also be used in the context of Box-Jenkins models.

3.8.4 Seasonal Differencing

At the model identification stage, the goal is to detect seasonality, if it exists, and to identify the order for the seasonal autoregressive and seasonal moving average terms. For many series, the period is known and a single seasonality term is sufficient. For example, for monthly data one would typically include either a seasonal AR 12 term or a seasonal MA 12 term. For Box–Jenkins models, one does not explicitly remove seasonality before fitting the model. Instead, one includes the order of the seasonal terms in the model specification to the ARIMA estimation software. However, it may be helpful to apply a seasonal difference to the data and regenerate the autocorrelation and partial autocorrelation plots. This may help in the model identification of the non-seasonal component of the model. In some cases, the seasonal differencing may remove most or all of the seasonality effect.

3.8.5 Identifying *p* and *q*

Once stationarity and seasonality have been addressed, the next step is to identify the order (i.e. the *p* and *q*) of the autoregressive and moving average terms. These are determined by examining the values of the autocorrelations and the partial autocorrelations with their corresponding plots as explained below.

3.8.6 Autocorrelation and Partial Autocorrelation Plots

The primary tools for doing this are the autocorrelation plot and the partial autocorrelation plot. The sample autocorrelation plot and the sample partial autocorrelation plot are compared to the theoretical behaviour of these plots when the order is known.

3.9 Best Model Identification and Selection Criteria

The Ljung-Box statistic would be used to identify the best model. The Ljung-Box statistic, also called the modified Box-Pierce statistic, is a function of the accumulated sample autocorrelations, r_j, up to any specified time lag *m*. As a function of *m*, it is determined as

$$Q(m) = n(n+1)\sum_{j=1}^{m}\frac{r_j^2}{n-j}$$

where *n* = number of usable data points after any differencing operations.

The Ljung-Box test can be defined as follows:

> **H_0:** The data are independently distributed (i.e. the correlations in the population from which the sample is taken are 0, so that any observed correlations in the data result from randomness of the sampling process).

H_a: The data are not independently distributed.

The choice of a plausible model depends on its p-value for the modified Box-Pierce if is well above .05, indicating "non-significance." In other words, the bigger the p-value, the better the model.

3.10 Some Applications of ARIMA Model in Real Life Situations

ARIMA modeling techniques have been applied in many fields of research. For example, Aidoo (2010) applied ARIMA model on the monthly inflationary rates in Ghana. He indicated that Ghana faces a macroeconomic problem of inflation for a long period of time. The problem in somehow slows the economic growth in this country. Using monthly inflation data from July 1991 to December 2009, we find that ARIMA $(1,1,1)(0,0,1)_{12}$ can represent the data behaviour of inflation rate in Ghana well. Based on the selected model, we forecast seven (7) months inflation rates of Ghana outside the sample period (i.e. from January 2010 to July 2010). The observed inflation rate from January to April which was published by Ghana Statistical Service Department fall within the 95% confidence interval obtained from the designed model. The forecasted results show a decreasing pattern and a turning point of Ghana inflation in the month of July.

Again, Cui (2011) researched on the topic: "ARIMA Models for Bank Failures: Prediction and Comparison." They said that the number of bank failures has increased dramatically over the last twenty-two years. A common notion in economics is that some banks can become "too big to fail." Is this still a true statement? What is the relationship, if any, between bank sizes and bank failures? In this thesis, the proposed modeling techniques are applied to real bank failure data from the FDIC. In particular, quarterly data from 1989:Q1 to 2010:Q4 are used in the data analysis, which includes three major parts: 1) pairwise bank failure rate comparisons using the conditional test (Przyborowski & Wilenski, 1940), 2) development of the empirical recurrence rate (Ho, 2008) and the empirical recurrence rates ratio time series; and 3) the Autoregressive Integrated Moving Average (ARIMA) model selection, validation, and forecasting for the bank failures classified by the total assets.

3.10.1 ARIMA Modeling in Disease Surveillance: Advantages and Disadvantages

According to Reis and Mandl (2003), one of the primary benefits of ARIMA models is their ability to correct for local trends in the data – what has happened on the previous day is incorporated into the forecast of what will happen today. This works well, for example, during a particularly severe flu season, where prolonged periods of high visit rates are adjusted to by the ARIMA model, thus preventing the alarm from being triggered every day throughout the flu season. However, if the ARIMA model “adjusts” to an actual outbreak instead of detecting it, a slowly spreading outbreak or attack might be missed because of this correction. This correction is most likely to affect detection of outbreaks occurring over several days, rather than those that occur suddenly. It is therefore also important to rely on the non-ARIMA or non-classical ARIMA models for outbreak detection.

ARMA models in Reis and Mandl (2003) and Reis, Pagano, and Mandl (2003) require large historic records of patient visits in order to begin surveillance. This is a substantial disadvantage. As can be seen from Moore et al. (2002), in some cases long historical data are not available and not necessary. Also, combining both historical and recent trends is quite realistic. Another disadvantage of ARMA is that the corresponding detector is not sensitive to the slow growth. According to Rizzo et al. (2005), outbreaks that evolve over a matter of days, for example, can often be detected with ARMA models that generate single-day predictions based on historical data. More gradually developing outbreaks are generally easier to detect by using such techniques as CuSum (Hawkins & Olwell, 1998).

CHAPTER FOUR

DATA COLLECTION, ANALYSIS AND RESULTS

4.1 Introduction

This chapter introduces the analysis of the various models and discussion of findings. The AR, MA, ARMA and ARIMA model shall be used in the modelling. The data covered OPD attendance data from 2002 to 2012 gathered on monthly bases.

4.2 Preliminary Analysis

It is recommended that a lengthy time series data is required for univariate time series forecasting. Meyler et al. (1988) recommended that at least 50 observations should be used for such a univariate time series forecasting. This could be problematic if few observations are used. However, when using a long time series data, it could be possible that the series contains a structural break which may necessitate only examining a sub-section of the entire data series or alternatively using intervention analysis or dummy variables. This is because there may be some conflict between the 50 needed for sufficient degrees of freedom for statistical robustness and having a shorter data sample to avoid structural breaks. The series should be plotted against time to assess whether any structural breaks, outliers or data errors occurred. This step may also reveal whether there is significant seasonal pattern in the times series or not. A dimension of the preliminary analysis for examining non-stationarity of the data is by considering the time series plot of OPD cases 2002 to 2012 as shown in Figure 4.1.

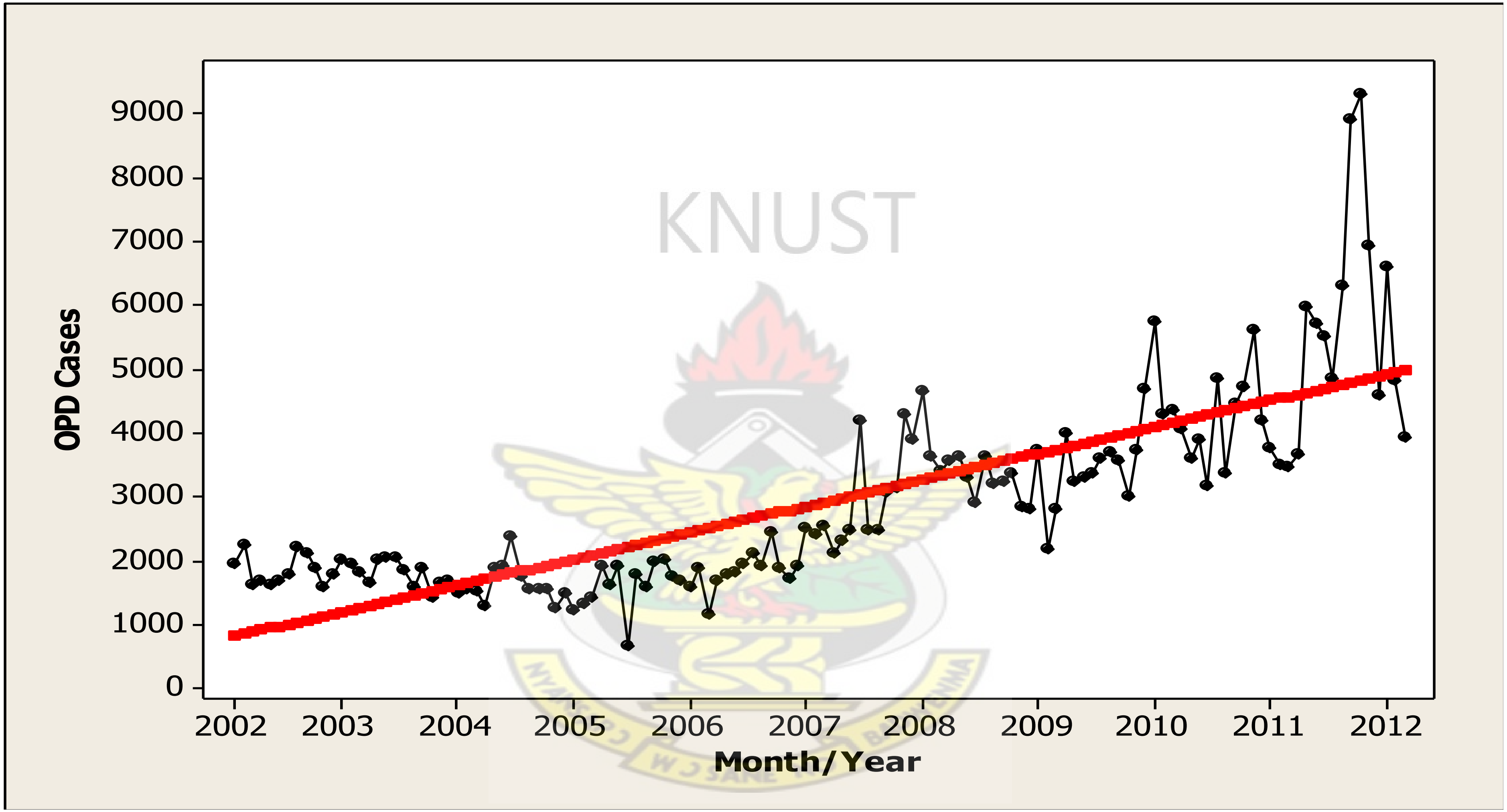

Figure 4.1: Trend in Monthly/Yearly OPD Attendance between 2002 and 2012

Figure 4.1 revealed that OPD cases from 2002 to 2012 have been largely seasonal. The mean is not constant throughout the series as it assumes a fairly stable mean till 2007 and thereafter assumes increasing values. Years 2005 and 2012 recorded the least and highest OPD attendance respectively.

The Moving Average (MA) analyses for lags 2, 4 and 12 are in Figures 4.2, 4.3, and 4.4 below. A comparison of their respective accuracy measures indicates that MA (2) better fits the OPD attendance data than the others.

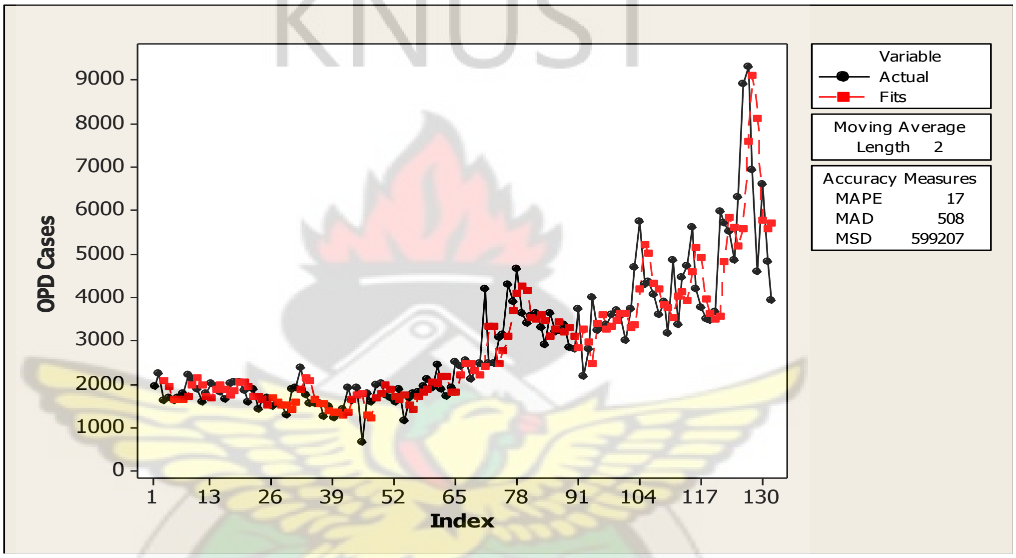

Figure 4.2: Moving Average (MA) with 2 Averages

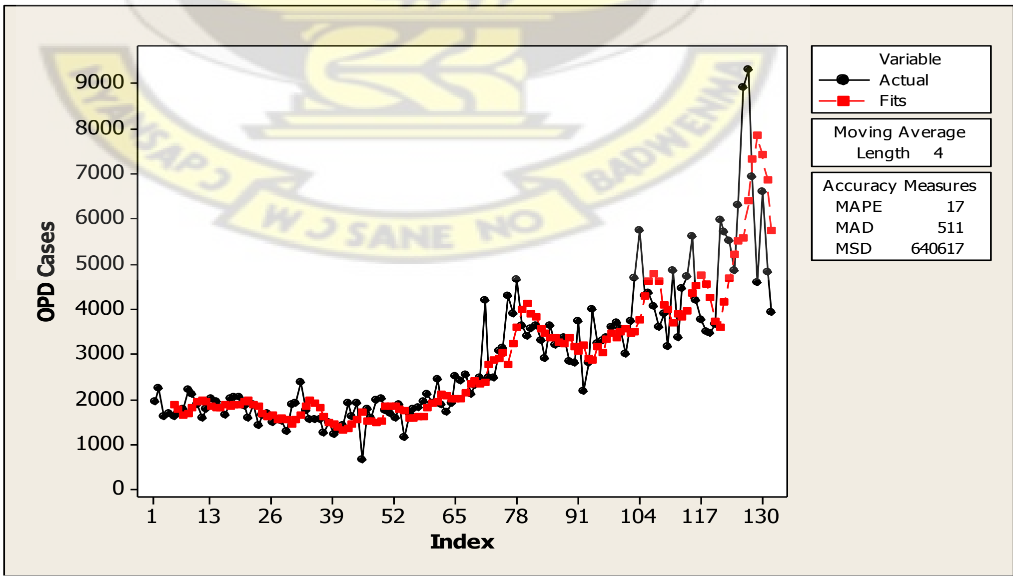

Figure 4.3: Moving Average (MA) with 4 Averages

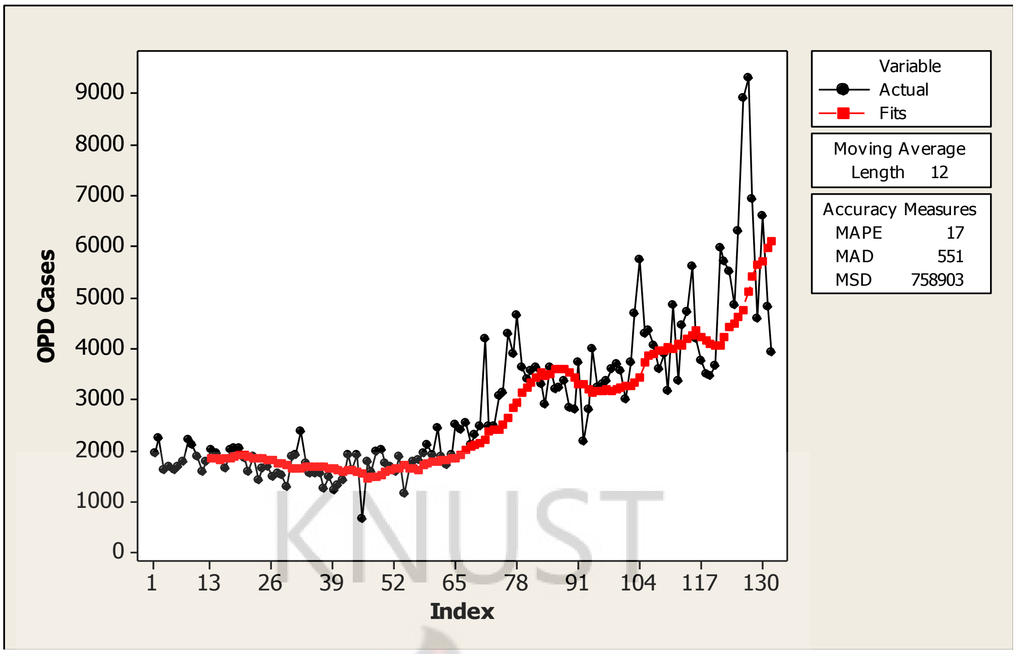

Figure 4.4: Moving Average (MA) with 12 Averages

The next step in the model building procedure is to determine the order of the AR and MA for both seasonal and non-seasonal components. This was suggested by the sample ACF and PACF plots based on the Box-Jenkins approach. From Figure 4.2, the correlations are significant for a large number of lags but perhaps the autocorrelations at lags 2 or and above are merely due to the propagation of the autocorrelation at lag 1. This is confirmed by the PACF plot.

The ACF and PACF plots respectively suggest that q = 2 or 3, and p =1 would be needed to describe this data set as coming from a non-seasonal moving average and autoregressive process respectively.

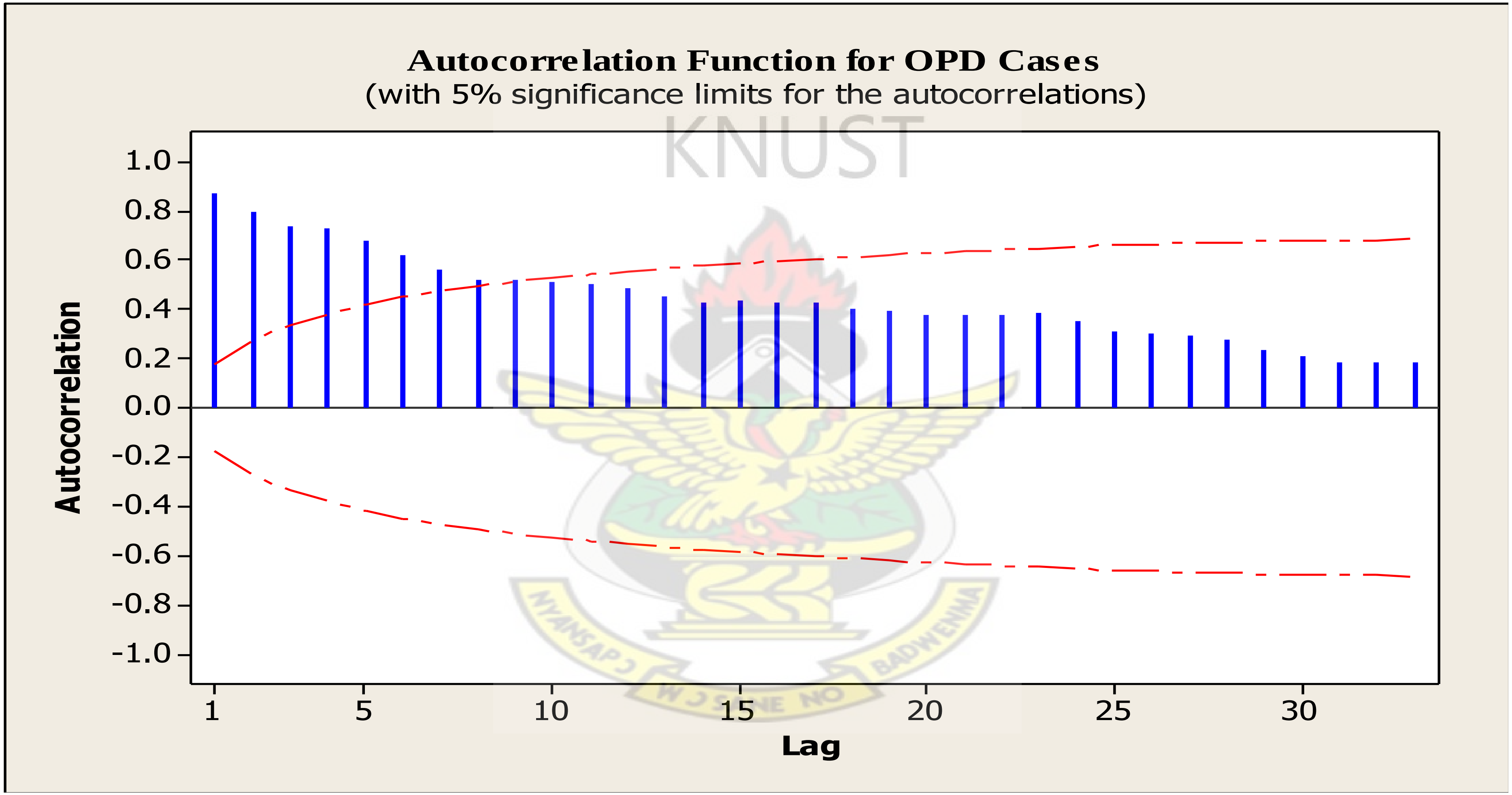

Figure 4.5: ACF for First Order Differencing

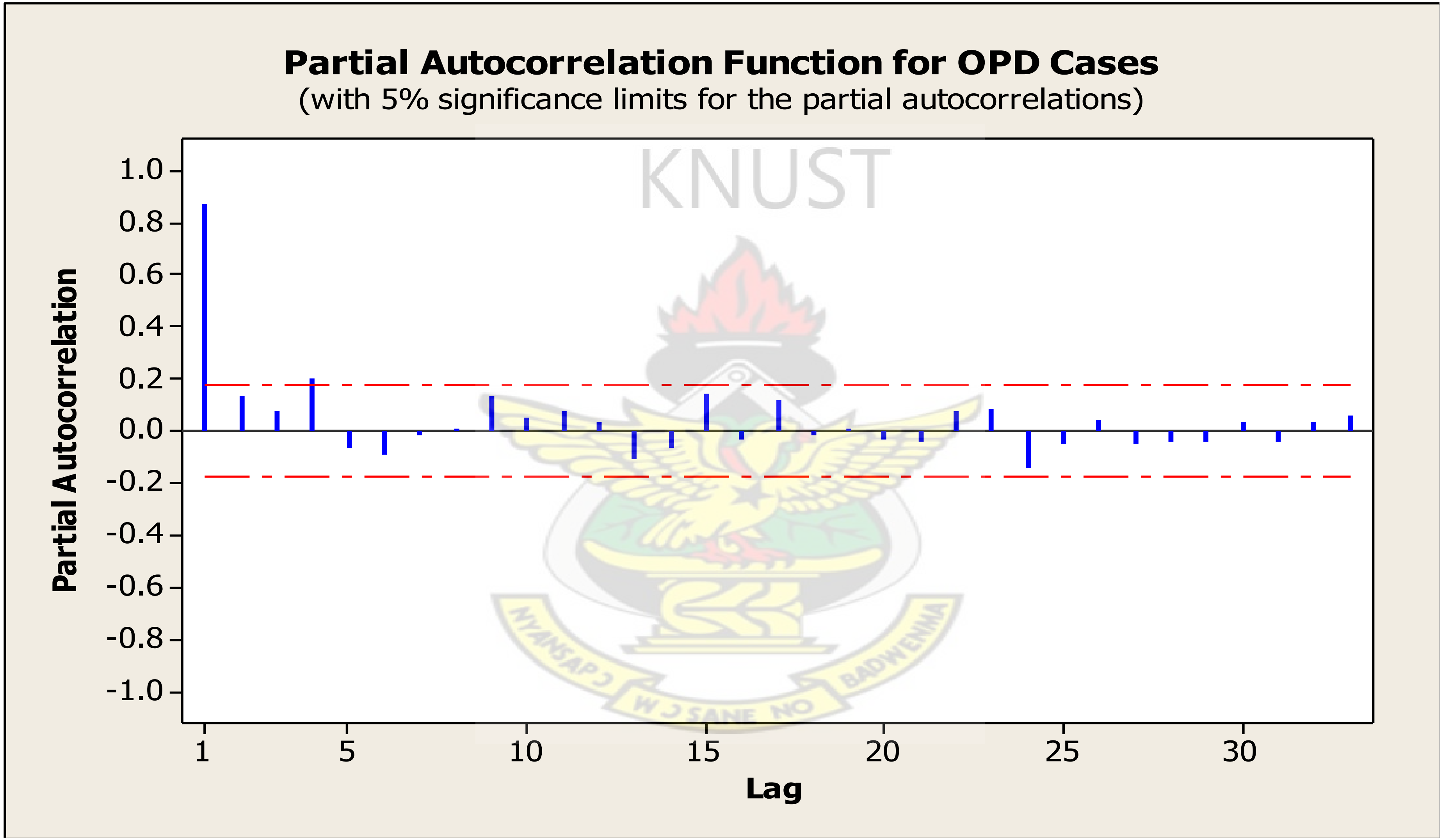

Figure 4.6: PACF for First Order Differencing

4.3 ARIMA Model Estimations

Several non-seasonal ARIMA models are constructed as follows:

Table 4.1: ARIMA (1, 1, 2)$_{12}$

Type	Coefficient	SE	t	P
Constant	13.551	5.675	2.39	0.018
AR 1	0.5442	0.1354	4.02	0.000
MA 1	0.9119	0.1348	6.76	0.000
MA 2	0.0162	0.1165	0.14	0.890

$\chi^2 = 5.8$; $p = 0.675$; df = 8

Table 4.2: ARIMA (1, 1, 1)$_{12}$

Type	Coefficient	SE	t	P
Constant	12.958	4.577	2.83	0.005
AR 1	0.5641	0.0985	5.73	0.000
MA 1	0.9344	0.0414	22.57	0.000

$\chi^2 = 5.9$; $p = 0.751$; df = 9

Table 4.3: ARIMA (1, 1, 3)$_{12}$

Type	Coefficient	SE	t	P
Constant	13.551	5.675	2.39	0.018
AR 1	0.5442	0.1354	4.02	0.000
MA 1	0.9119	0.1348	6.76	0.000
MA 2	0.0162	0.1165	0.14	0.890

$\chi^2 = 5.3$; $p = 0.628$; df = 7

Table 4.4: ARIMA (1, 2, 3)$_{12}$

Type	Coefficient	SE	*t*	*P*
Constant	0.027	5.046	0.01	0.996
AR 1	-0.6533	1.4778	-0.44	0.659
MA 1	0.5603	1.5196	0.37	0.713
MA 2	0.5629	1.7564	0.32	0.749
MA 3	-0.1462	0.3402	-0.43	0.668

$\chi^2 = 14.5$; $p = 0.043$; df = 7

In comparing the p-values and Chi-square values of the four likely non-seasonal ARIMA, it can be concluded that model ARIMA (1, 1, 1)$_{12}$ has the highest p and a relatively low Chi-square values of 0.751 and 5.9. This indicates that it is the best non-seasonal model for the data.

Table 4.5: Comparison of Non-Seasonal Models

Model	p-value	Chi-Square	Df
ARIMA (1, 1, 1)$_{12}$	**0.751**	**5.9**	**9**
ARIMA (1, 1, 2)$_{12}$	0.675	5.8	8
ARIMA (1, 1, 3)$_{12}$	0.628	5.3	7
ARIMA (1, 2, 3)$_{12}$	0.043	14.5	7

At this stage, it is important to also consider the seasonality of the data by adopting the seasonal ARIMA models. Looking at the seasonal lags, both ACF and PACF spike at seasonal lag 12 and drop to zero for other seasonal lags suggesting that Q =1 or 2 and $P = 0$ or 1 with $d = 1$ would be needed to describe these data as coming from a seasonal moving average and autoregressive process.

4.4 Model Evaluation and Selection

After the model has been identified, we use conditional-sum-of-squares to find starting values of parameters, then do the maximum likelihood estimate for the proposed models. The procedure for choosing these models relies on choosing the model with the maximum p-values for the Ljung-Box statistic (more than 5% as a rule of thumb) and minimum Chi-Square values. The 10 proposed SARIMA models are presented in Table 4.6 with their corresponding p-values.

Table 4.6: Suggested SARIMA Models

Model	p-value	Chi-Square	Df
ARIMA (1, 1, 2) $(0, 1, 1)_{12}$	0.613	5.4	7
ARIMA (1, 1, 3) $(0, 1, 1)_{12}$	**0.768**	**3.3**	**6**
ARIMA (1, 1, 1) $(0, 0, 1)_{12}$	0.711	5.4	8
ARIMA (1, 1, 2) $(1, 1, 1)_{12}$	0.693	3.0	5
ARIMA (1, 1, 2) $(1, 1, 0)_{12}$	**0.735**	**4.4**	**7**
ARIMA (1, 1, 1) $(0, 1, 1)_{12}$	**0.722**	**5.3**	**8**
ARIMA (1, 1, 3) $(0, 1, 2)_{12}$	0.630	3.5	5
ARIMA (1, 2, 1) $(1, 0, 1)_{12}$	0.023	16.2	7
ARIMA (1, 2, 1) $(1, 1, 0)_{12}$	0.055	14.7	8
ARIMA (2, 2, 1) $(1, 2, 1)_{12}$	0.006	18.2	6

A critical comparison of the models based on their respective *p*-values and Chi-Square values shows that SARIMA (1, 1, 3)(0, 1, 1)$_{12,}$ SARIMA (1, 1, 2)(1, 1, 0)$_{12}$ and SARIMA (1, 1, 1)(0, 1, 1)$_{12}$ are the appropriate models that well fit the OPD data. In time series modeling, the selection of a best model fit to the data is directly related to whether residual analysis is performed well. One of the assumptions of ARIMA model is that for a good model, the residuals must follow a white noise process. That is, the residuals have zero mean, constant variance and also uncorrelated. Tables 4.7, 4.8 and 4.9 present the parameter estimates of the three possible SARIMA models.

Table 4.7: Estimates of Parameters for SARIMA (1, 1, 2) (1, 1, 0)$_{12}$

Variable	Coefficients	SE	*t*	*P*
AR (1)	0.5838	0.0809	7.22	0.000
SAR (12)	-0.5103	0.1040	-4.91	0.000
MA (1)	1.0765	0.0164	65.46	0.000
MA (2)	-0.0927	0.0317	-2.93	0.004
Constant	5.6690	1.6590	3.42	0.001

p=0.735, χ^2=4.4

Table 4.8: Estimates of Parameters for SARIMA (1, 1, 1) (0, 1, 1)$_{12}$

Variable	Coefficients	SE	*t*	*P*
AR (1)	0.5041	0.1003	5.03	0.000
MA (1)	0.9387	0.0534	17.57	0.000
SMA (12)	0.7313	0.1198	6.11	0.000
Constant	4.5330	1.8510	2.45	0.016

p=0.722, χ^2=5.3

Although ARIMA models $(1, 1, 1)_{12}$, (1, 1, 2) $(1, 1, 0)_{12}$ and (1, 1, 1) $(0, 1, 1)_{12}$ somewhat adequate, the SARIMA (1, 1, 3) $(0, 1, 1)_{12}$ is most plausible because of its high p and least Chi-Square values. More so, the diagnostic analyses using the ACF of residuals, PACF residuals, and the normal probability plot of the residuals as shown in Figures 4.7, 4.8 and 4.9 reveal that the residuals of the model have zero mean and constant variance. Also, the ACF of the residuals depicts that the autocorrelation of the residuals are all zero, that is to say they are uncorrelated. Hence, it can be concluded that there is a constant variance among residuals of the selected model and the true mean of the residuals is approximately equal to zero. Thus, the selected model satisfies all the model assumptions. Since the SARIMA (1, 1, 3)$(0, 1, 1)_{12}$ satisfies all the necessary assumptions, it can be inferred that the model provides an adequate representation of the data. Hence, the predictive model would be formulated from the parameter estimates in Table 4.9.

Table 4.9: Estimates of Parameters for SARIMA (1, 1, 3) $(0, 1, 1)_{12}$

Variable	**Coefficients**	**SE**	***t***	***P***
AR (1)	-0.2441	0.3099	-0.79	0.433
MA (1)	0.1569	0.2933	0.54	0.594
MA (2)	0.3190	0.1731	1.84	0.038
MA (3)	0.3818	0.1239	3.08	0.003
SMA (12)	0.7537	0.1176	6.41	0.000
Constant	11.3510	3.1830	3.57	0.001

***p=0.768*, χ^2=3.3**

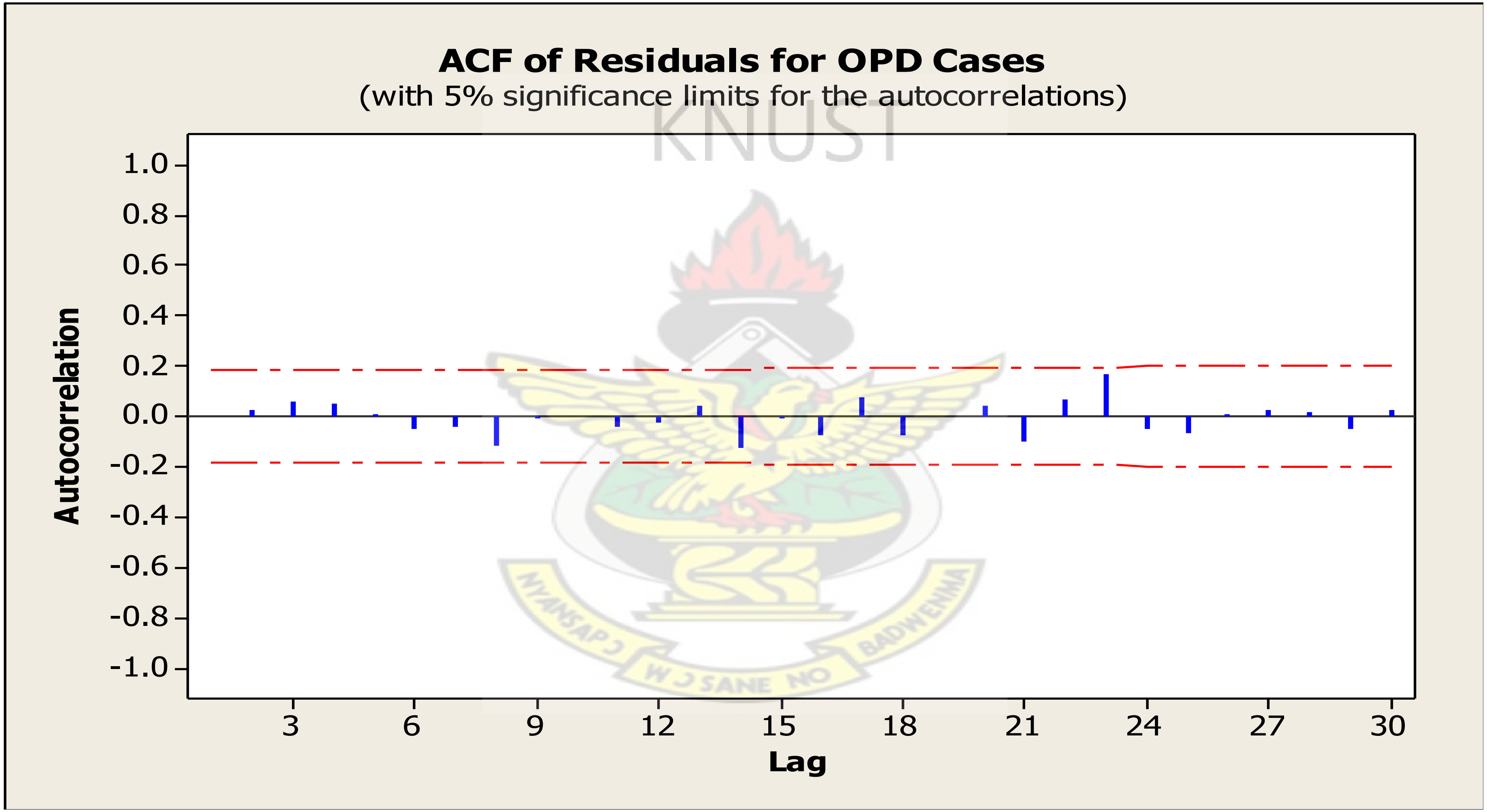

Figure 4.7: ACF Diagnostic Plot of the Residuals for SARIMA (1, 1, 3) (0, 1, 1)$_{12}$ Model

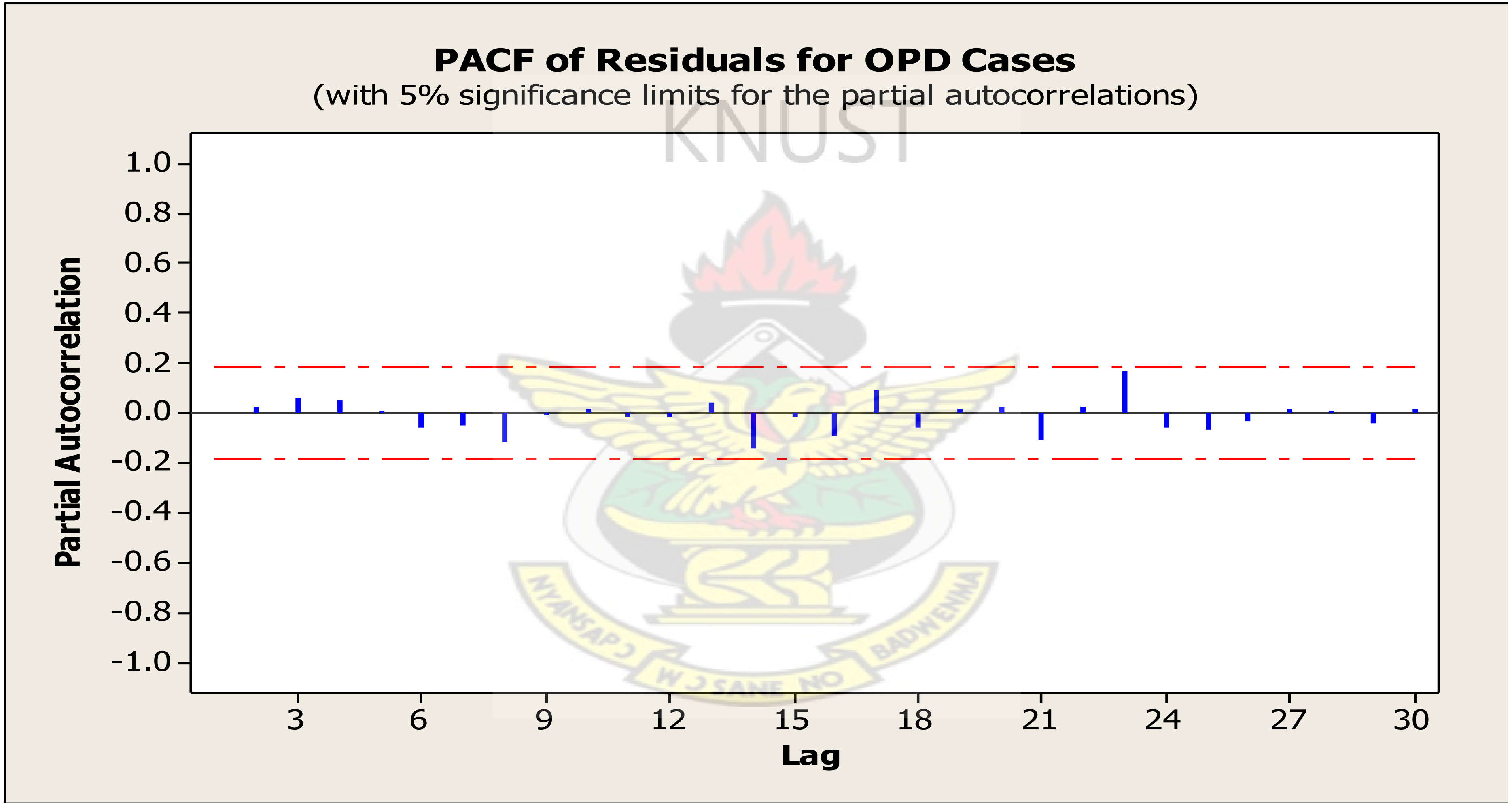

Figure 4.8: PACF Diagnostic Plot of the Residuals for SARIMA (1, 1, 3) (0, 1, 1)$_{12}$ Model

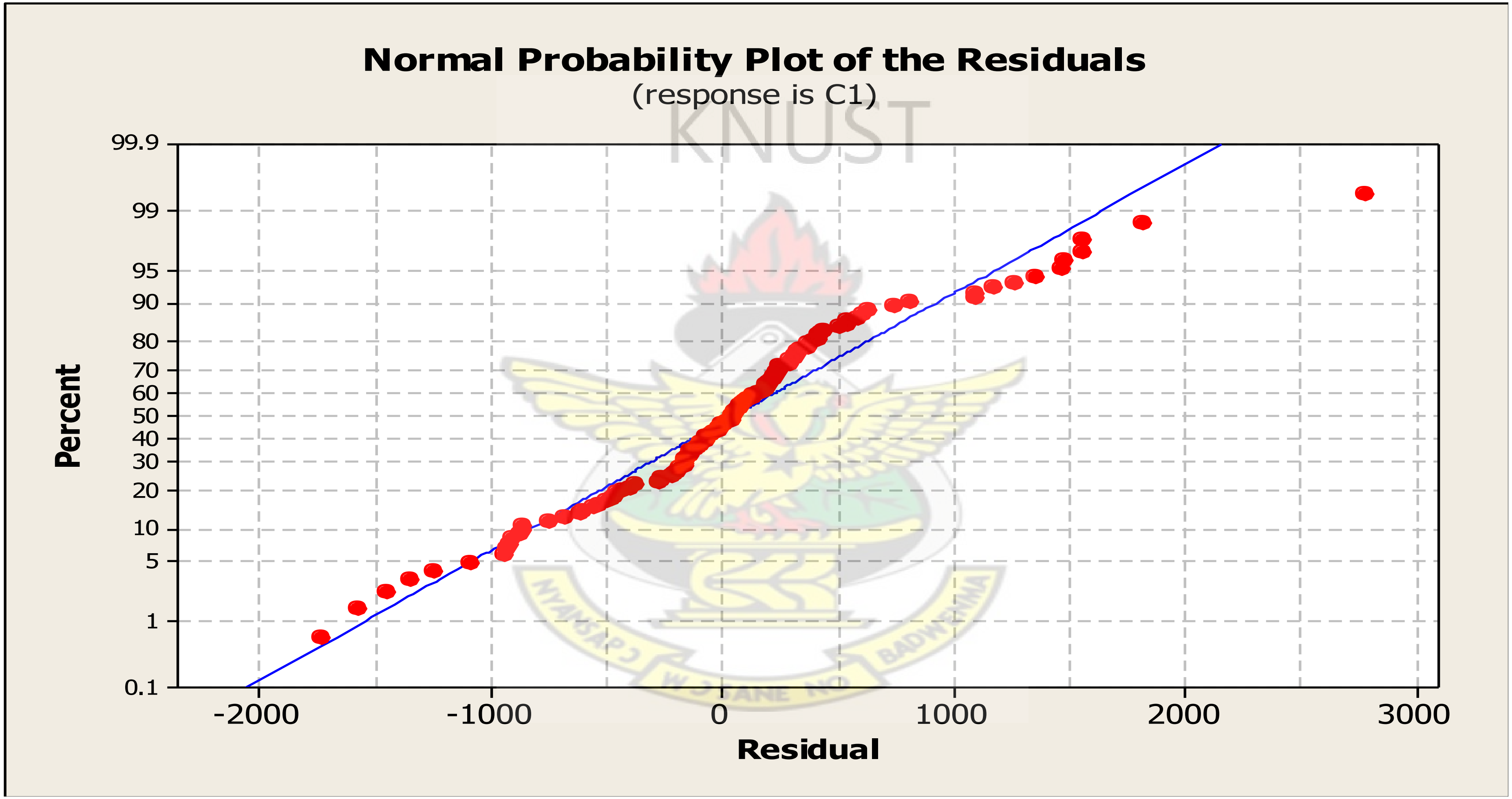

Figure 4.9: Normal Probability Plot of the Residuals for SARIMA (1, 1, 3) (0, 1, 1)$_{12}$ Model

4.5 Forecasting

Forecasting plays an important role in decision making process. It is a planning tool which helps decision makers to foresee the future uncertainty based on the behaviour of past and current observations. Forecasting as described by Box and Jenkins (1976), provide the basis for economic and business planning, inventory and production control and control and optimisation of industrial processes. Forecasting is the process of predicting some unknown quantities. From previous studies, most research work has found that the selected model is not necessary the model that provides best forecasting. In this sense, further forecasting accuracy test such as Mean Error (ME), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) must be performed on the model. Table 4.10 summarises the forecasted values of OPD cases over the period of January 2013 to December 2013 with 95% confidence level using the SARIMA (1, 1, 3) $(0, 1, 1)_{12}$ model which has higher *p*-value of 0.768 (thus, greater than alpha value of 0.05) indicating that it is the best model according to Modified Box-Pierce Chi-Square statistic.

Table 4.10: Forecasted OPD Attendance for the Next 5 Years

Year	Month	Forecast	Lower Limit	Upper Limit
2013	**January**	**5028.0**	3655.5	6400.4
2013	**February**	**5801.0**	4201.1	7400.8
2013	**March**	**6374.5**	4692.7	8056.3
2013	**April**	**5996.2**	4313.0	7679.4
2013	**May**	**6497.0**	4804.4	8189.6
2013	**June**	**7371.9**	5672.5	9071.3

2013	**July**	**7862.6**	6155.9	9569.4
2013	**August**	**7010.4**	5296.4	8724.3
2013	**September**	**6198.1**	4477.0	7919.2
2013	**October**	**6878.1**	5149.8	8606.3
2013	**November**	**6398.2**	4662.9	8133.6
2013	**December**	**6071.0**	4328.5	7813.4
2014	January	**6936.2**	5124.8	8747.5
2014	February	**7001.7**	5155.0	8848.3
2014	March	**7371.3**	5502.8	9239.8
2014	April	**7054.1**	5177.6	8930.7
2014	May	**7551.4**	5664.1	9438.7
2014	June	**8438.5**	6541.2	10335.7
2014	July	**8937.6**	7030.3	10844.9
2014	August	**8094.6**	6177.3	10012.0
2014	September	**7291.5**	5364.2	9218.8
2014	October	**7980.6**	6043.4	9917.8
2014	November	**7509.8**	5562.8	9456.9
2014	December	**7191.7**	5234.9	9148.5
2015	**January**	**8066.0**	6037.8	10094.3
2015	**February**	**8140.6**	6073.7	10207.6
2015	**March**	**8519.4**	6427.3	10611.5
2015	**April**	**8211.3**	6108.5	10314.2
2015	**May**	**8717.7**	6601.3	10834.2
2015	**June**	**9614.0**	7484.7	11743.2

2015	**July**	**10122.2**	7980.1	12264.3
2015	**August**	**9288.4**	7133.5	11443.2
2015	**September**	**8494.3**	6326.8	10661.9
2015	**October**	**9192.6**	7012.4	11372.7
2015	**November**	**8730.9**	6538.2	10923.7
2015	**December**	**8421.9**	6216.7	10627.1
2016	January	**9305.4**	7027.1	11583.6
2016	February	**9389.1**	7069.4	11708.8
2016	March	**9777.0**	7429.2	12124.8
2016	April	**9478.1**	7116.9	11839.2
2016	May	**9993.6**	7616.1	12371.1
2016	June	**10898.9**	8506.0	13291.9
2016	July	**11416.3**	9007.8	13824.8
2016	August	**10591.6**	8167.7	13015.5
2016	September	**9806.7**	7367.5	12245.9
2016	October	**10514.0**	8059.6	12968.5
2016	November	**10061.5**	7592.0	12531.1
2016	December	**9761.6**	7277.0	12246.2
2017	**January**	**10654.2**	8095.6	13212.8
2017	**February**	**10747.1**	8144.7	13349.5
2017	**March**	**11144.1**	8511.1	13777.1
2017	**April**	**10854.3**	8205.4	13503.2
2017	**May**	**11378.9**	8711.1	14046.7
2017	**June**	**12293.4**	9607.6	14979.2

2017	**July**	**12819.9**	10116.1	15523.7
2017	**August**	**12004.3**	9282.6	14726.0
2017	**September**	**11228.5**	8489.0	13968.0
2017	**October**	**11945.0**	9187.8	14702.1
2017	**November**	**11501.6**	8726.9	14276.3
2017	**December**	**11210.8**	8418.7	14003.0

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.1 Introduction

This chapter presents conclusions drawn from the study and some recommendations made to inform policy at the Saltpond Hospital at Saltpond in the Central Region of Ghana.

5.2 Conclusion

The objective of this research was to develop a time series model and forecast OPD attendance data for the next 5 years (January 2013 to December 2017). Data from 2002 to 2012 were collated from the Health Information Department of the Hospital. The preliminary analysis revealed that the Hospital recorded its highest attendance in July 2012 with 9320 cases.

Several time series models including AR, MA, Autoregressive Integrated Moving Average (ARIMA) and seasonal ARIMA were used in modelling the data in the Minitab 14. The study identified 14 'candidate' models which best fitted the data. However, with the use of the Modified Box-Pierce (Ljung-Box) Chi-square statistic criteria of the "largest *p*-value and minimum Chi-Square value," four best-fitted ARIMA models tend to be ARIMA (1, 1, 1)$_{12,}$SARIMA (1, 1, 3) (0, 1, 1)$_{12,}$ SARIMA (1, 1, 2) (1, 1, 0)$_{12}$ and SARIMA (1, 1, 1) (0, 1, 1)$_{12}$ were selected and evaluated.

After the estimation of the parameters of selected models, a series of diagnostic and forecasting accuracy tests were performed. Having satisfied all the model assumptions, SARIMA (1, 1, 3) (0, 1, 1)$_{12}$ model was adjudged to be the best and most plausible model for forecasting OPD cases at Saltpond Hospital.

With reference to the findings of the research, it can be concluded that:

1. The most adequate model for the data was SARIMA (1, 1, 3) (0, 1, 1)$_{12}$.
2. There will be an increase in OPD attendance at the hospital in the next 5 years as the forecasted values fell in the neighbourhood of 5,028 in January 2013 and 11,211 in year 2017.

5.3 Recommendations

On the basis of the findings of the research, the following recommendations were made:

1. The Saltpond Municipal Hospital authorities should use the seasonal ARIMA (1,1,3) (0,1,1)$_{12}$ model in their outpatient department (OPD) attendance planning activities.
2. In order to prepare adequately for the further overwhelming OPD attendance at the hospital, the hospital authorities should rely upon the forecasted figures in its planning activities.
3. Government must continue to support health facilities like the Saltpond Hospital in terms of personnel and logistics in order to provide quality health care for the citizenry.
4. The hospital must investigate the cause of high attendance in the second and third quarters of every year, and embark on rigorous health and environmental campaigns in its catchment areas.

REFERENCES

Abdel, H. A. (2005). Stats methods of predicting fatalities of road traffic accidents in Kuwait. *Accident Analysis and Prevention, 18*(2), 119-127.

Agbodza, R. M. (1977). Trend analysis of onchocerciasis in the Kpando area of the Volta Region of Ghana. *Ghana Medical Journal*, *15,* 32-56.

Aidoo, E. (2010). *Modeling and forecasting inflation rates in Ghana: An application of SARIMA models.* Högskolan Dalarna, School of Technology and Business Studies.

Amankwah-Kumi, G. (2003). *Cost of care in public health facilities - a comparative study in Kumasi*. A Thesis submitted to the Kwame Nkrumah University of Science and Technology, Kumasi.

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

Birk, H. O., Gut, R., and Henriksen, L. O. (2011). Patients' experience of choosing an outpatient clinic in one county in Denmark: Results of a patient survey. *BMC Health Services Research*, **11**, 262.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994). *Time series analysis, forecasting and control* (3rd ed.). New Jersey: Prentice Hall, Englewood Clifs.

Cui, F. (2011). *ARIMA models for bank failures: Prediction and comparison*. University of Nevada, Las Vegas.

Fama, E. F., and Gibbons S. W. (1997). Asset Returns and inflation. *Journal of financial Economics, 5*(2), 115-146.

GHS (2011). *Central Region Annual Report.* Cape Coast: RHD.

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

GSS (2010). *Ghana Demographic and Health Survey*. Accra: GSS.

Hamilton, J. D. (1994). *Time series analysis*. New Jersey: Princeton University Press.

Hawkins, M. D., and Olwell, D. H. (1998). *Cumulative sum charts and charting for quality improvement*. Berlin/Heidelberg: SpringerVerlag.

Ho, C.-H. (2008). Empirical recurrent rate time series for Volcanism: Application to Avachinsky Volcano, Russia. *Volcanol Geotherm Res, 173*, 15-25.

Jawahar, S. K. (2007). A study on out patient satisfaction at a Super Specialty Hospital in India. *Internet Journal of Medical Update, 2*(2), 13-17.

Killaspy, H., Banerjee, S., King, M., and Lloyd, M. (2000). Prospective controlled study of psychiatric out-patient non-attendance. Characteristics and outcome. *Br J Psychiatry, 1*, 160-165.

Konotey-Ahulu, F. I. D. (1972). Definition of sickle cell trait and sickle cell disease. *Ghana Medical Journal, 11,* 417-420.

Meyler, A., Kenny, G., and Quinn, T. (1998). Forecasting Irish inflation using ARIMA models. Technical Paper, 3/RT/98.

Miyake, K., Miyake, N., Kondo, S., Tabe, Y., Ohsaka, A., and Miida, T. (2009). Seasonal variation in liver function tests: A time-series analysis of outpatient data. *Ann Clin Biochem*., 46(5), 377-384.

Nyako, L. A. (2002). *An assessment of utilization of clinical services at the La Polyclinic.* A Thesis submitted to the Kwame Nkrumah University of Science and Technology, Kumasi.

Patavegar, B. N., Shelke, S. C., Adhav, P., and Kamble, M. S. (2012). A cross-sectional study of patient's satisfaction towards services received at Tertiary Care Hospital on OPD basis. *National Journal of Community Medicine, 3*(2), 232-237.

Przyborowski, J., and Wilenski, H. (1940). Homogeneity of results in testing samples from Poisson series. *Biometrika, 31*, 313-323.

Reis, B. Y., Pagano, M., and Mandl, K. D. (2003). Using temporal context to improve biosurveillance. *Proceedings National Academy of Sciences*, *100*(4), 1961-1965.

Reti, S. (2003). Improving outpatient department efficiency: a randomised controlled trial comparing hospital and general-practice telephone reminders. *N Z Med J.*, *116*(1175), 34-56.

Rizzo, S. L., Grigoryan, V. V., Wallstrom, G. L., Hogan, W. R., and Wagner M. M. (2005). *The use of case studies for the evaluation of surveillance systems.* RODS Technical Report Centre for Biomedical Informatics, University of Pittsburgh.

Stone, C. A., Palmer, J. H., Saxby, P. J., and Devaraj, V. S. (1999). Reducing non-attendance at outpatient clinics. *Journal of the Royal Society of Medicine, 92*(3), 114–118.

Thompson, M. (1962). *Effects of malaria in sickle cell patients*. Medical Research.

APPENDIX A

RAW DATA ON THE OPD ATTENDANCE FOR THE 11 - YEAR PERIOD (2002-2012)

MONTH	YEARS										
	2002	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
JANUARY	1939	1998	1685	1232	2003	2442	2461	3636	3367	3904	5962
FEBRUARY	2241	1944	1479	1466	1759	1880	3053	3196	3583	3162	5698
MARCH	1613	1826	1538	1227	1693	1722	3116	3218	3703	4861	5498
APRIL	1684	1652	1512	1303	1590	1925	4272	3369	3558	3374	4852
MAY	1609	2018	1294	1418	1892	2513	3879	2829	3012	4450	6297
JUNE	1676	2045	1860	1898	1141	2404	4657	2808	3712	4703	8901
JULY	1762	2030	1906	1614	1672	2546	3635	3731	4693	5610	9320
AUGUST	2217	1836	2373	1921	1765	2091	3398	2173	5738	4195	6927
SEPTEMBER	2091	1574	1751	669	1824	2298	3577	2785	4303	3743	4591
OCTOBER	1870	1876	1558	1786	1958	2485	3621	3982	4347	3505	6593
NOVEMBER	1575	1398	1544	1581	2109	4198	3301	3220	4058	3459	4813
DECEMBER	1772	1642	1536	1970	1924	2459	2896	3313	3609	3651	3921
TOTAL	**22049**	**21839**	**20036**	**18085**	**21330**	**28,959**	**41,866**	**38,266**	**47,683**	**48,660**	**73,373**

Source: Saltpond Municipal Hospital, C/R, 2