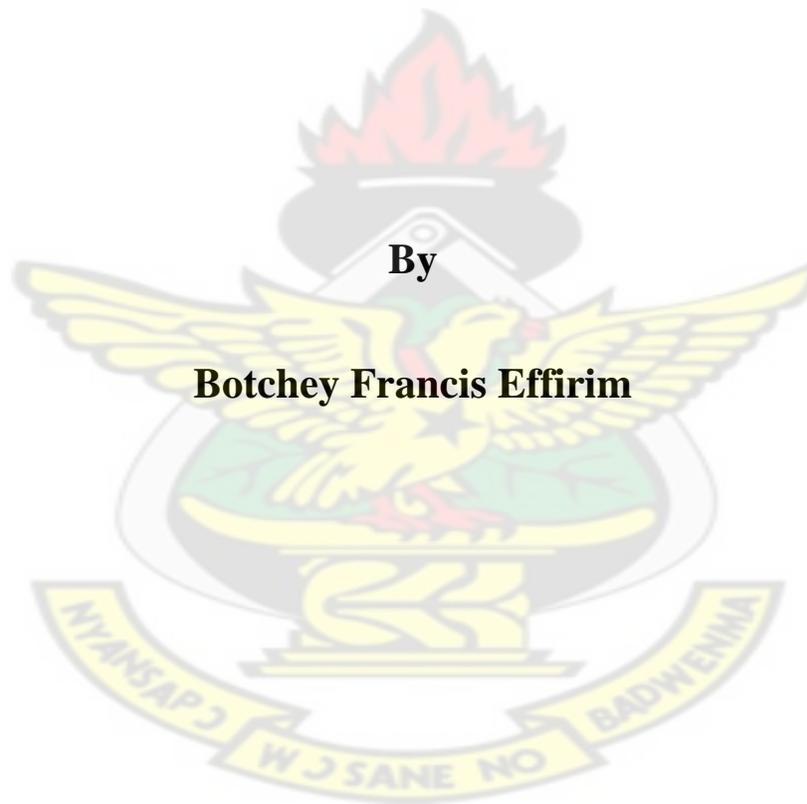


# **Modeling Demand for a Telecommunication Utility Operator Based On the Box-Jenkins Methodology**

KNUST



**May 2012**

**Kwame Nkrumah University of Science and Technology**



**College of Engineering**

**Department of Electrical/Electronic Engineering**

**Modeling Demand for a Telecommunication Utility Operator Based  
On the Box-Jenkins Methodology**

---

**Submitted in Partial Fulfillment of the Degree of MSc. Telecommunications  
Engineering**

**By**

**Botchey Francis Effirim, BSc. (Hons)**

**(PG2652408)**

**SUPERVISOR**

**Nana KwasiDiawuo (PhD.)**

**DECLARATION**

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

KNUST

.....

Student Name & ID

Signature

Date

Certified by:

.....

Supervisor

Signature

Date

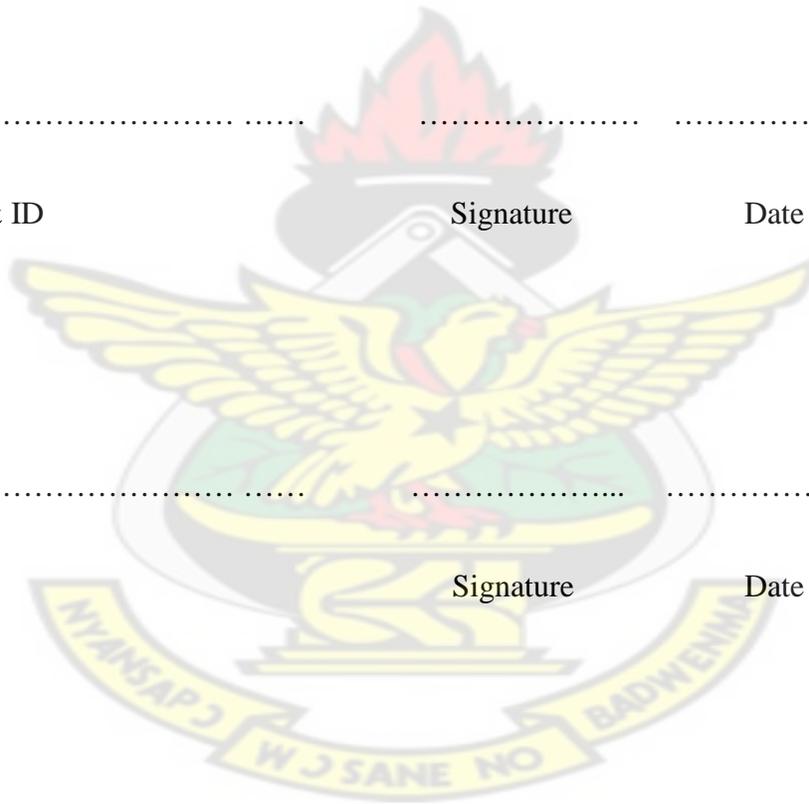
Certified by:

.....

Head of department

Signature

Date



## ABSTRACT

Telecommunication Demand Modeling has been a problem both in Ghana and the global world and has been under study for a long period of time. Developments in new communication technologies such as mobile devices and developments of new services to meet user demand such as video, multi-media and data services has led to an exponential growth in the demand for these services. In Ghana, the national regulator, the National Communications Authority puts cellular phone penetration in the country at about seventy percent (70%). The sharp increase in the number of customers over a short period and limited infrastructure has led users to raise concerns over the quality of service been offered. It is in this light that this thesis attempts to solve a case study of the telecommunication demand modeling for a leading cellular company in Ghana.

Worldwide, different approaches have been employed in solving this very sensitive demand problem. The thesis employed the Box- Jenkins methodology to develop a near perfect ARIMA (2, 1, 0) model with the corresponding equation being

$$Z_t = Z_{t-1} + .042(Z_{t-1} - Z_{t-2}) - .564(Z_{t-2} - Z_{t-3}) + 3835.022$$
 that can adequately forecast the future rise in the subscriber base so that this can be carefully factored in planning so proper decisions can be taken to minimize this demand problem.

With a mean absolute error of 8405 which is insignificant when considering the nation Ghana as a whole.

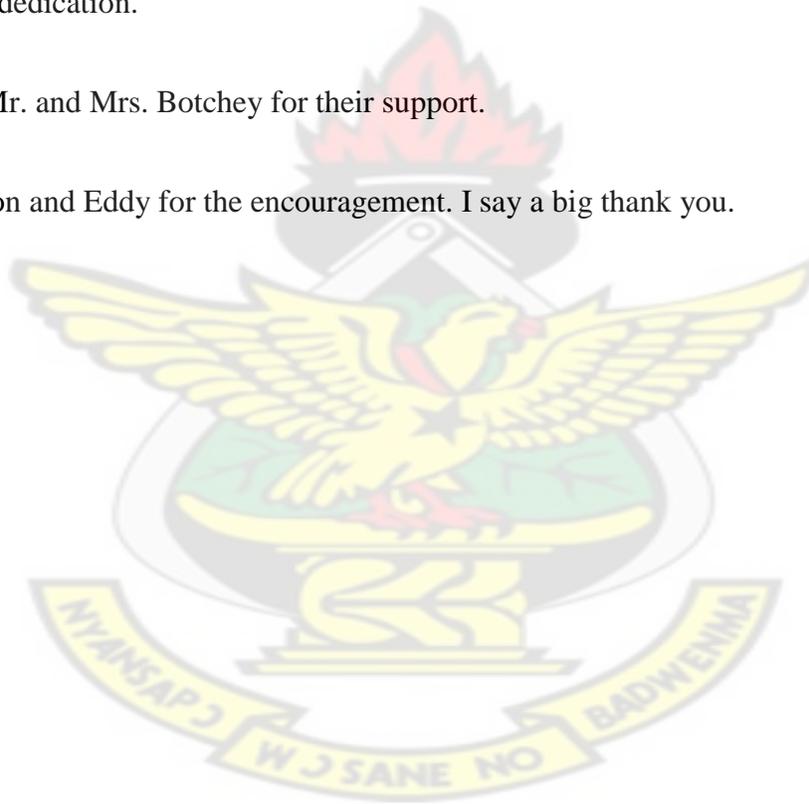
## ACKNOWLEDGEMENTS

I am very thankful and grateful to the creator and sustainer of this universe, Almighty God for making everything possible.

I would like to express my profound gratitude to Nana KwasiDiawuo (PhD) the one who supervised this enormous task and Dr. J.D. Gadze for the guidance. I say thank you for your supervision and dedication.

To my parents Mr. and Mrs. Botchey for their support.

To Maria, Gordon and Eddy for the encouragement. I say a big thank you.

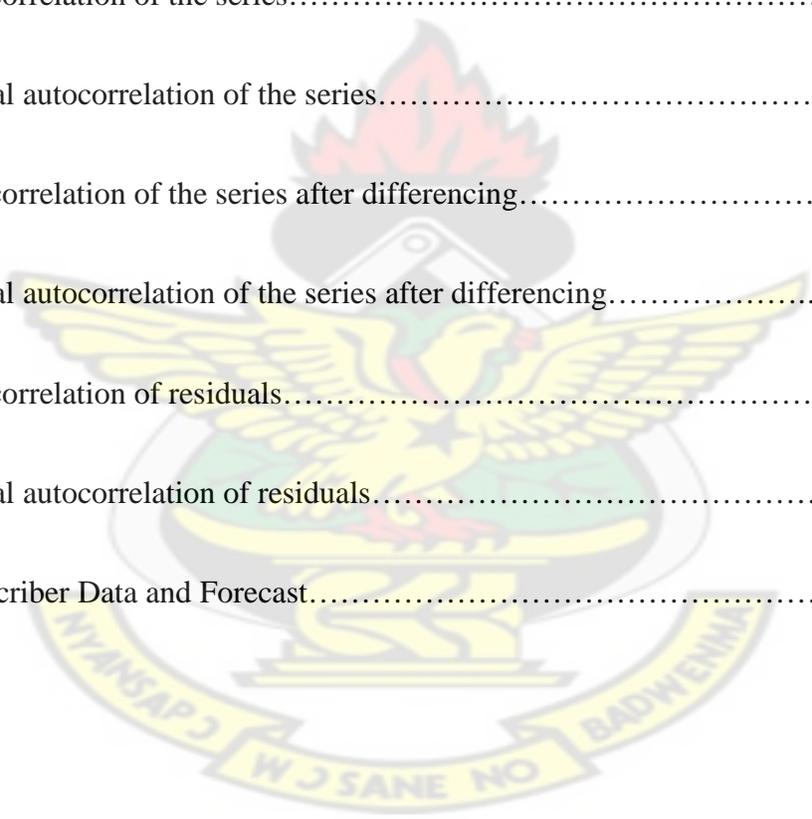


## TABLE OF CONTENTS

<b>Chapter 1 Introduction</b> .....	<b>1</b>
1.1 Background To Study.....	1
1.2 Problem Statement.....	3
1.3 General Objectives.....	4
1.4Relevance of Reserch.....	4
1.5Brief History of The Telecommunications Services in Ghana.....	4
1.6Thesis Organisation.....	8
<b>Chapter 2 Literature Review</b> .....	<b>9</b>
2.1 Review of Related Works.....	9
2.2 Approaches to Time Series Forecasting .....	10
2.2.1 Introduction to the Box-Jenkins Methodology.....	10
2.2.2 Components of Time Series.....	11
2.2.3 Lag or Backward Linear Operator.....	12
2.2.4 The difference operator.....	13
2.2.5 Purely Seasonal model.....	18
2.2.6 Seasonal Autoregressive model.....	18
2.2.7 Seasonal Moving Average.....	19
2.2.8 The Yule Walker Equation.....	19

2.2.9 Akaike Information Criteria/Bayesian Information Criteria/Schwarz -	
Bayesian Information Criteria.....	20
2.2.10 The Mean Absolute Error (MAE).....	21
2.3 Methodology.....	21
<b>Chapter 3 Model Development and Implementation.....</b>	<b>23</b>
3.1 Data for Model.....	23
3.2 Model Identification.....	24
3.2.1 Selecting from competing Models.....	29
3.3 Parameter Estimation.....	29
3.4 Model Diagnostics and Formulation.....	30
3.5 Forecasting.....	32
3.5.1 Upper and Lower Confidence Limit.....	32
3.5.2 Sample Period Forecasts.....	34
3.5.3 Post Sample Forecasts.....	34
3.6 Simulated Results Discussion.....	35
<b>Chapter 4 Conclusions and Recommendations.....</b>	<b>37</b>
4.1 Conclusion .....	37

4.2Reccomendations and Future Works.....	41
<b>References.....</b>	<b>43</b>
<b>Appendix A Data.....</b>	<b>47</b>
A.1 data for the number of subscribers.....	47
<b>Appendix B Tables.....</b>	<b>48</b>
B.1 autocorrelation of the series.....	48
B.2 partial autocorrelation of the series.....	49
B.3 autocorrelation of the series after differencing.....	50
B.4 partial autocorrelation of the series after differencing.....	51
B.5 autocorrelation of residuals.....	52
B.6 partial autocorrelation of residuals.....	53
B.7 Subscriber Data and Forecast.....	54



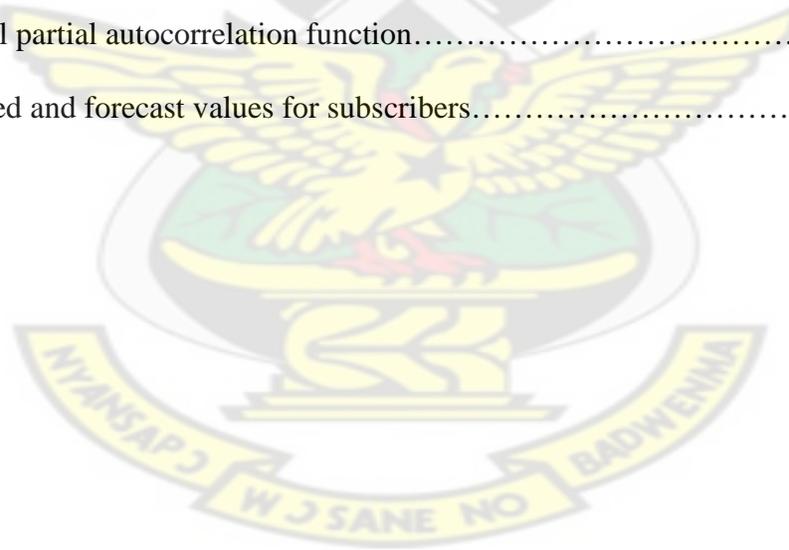
## LIST OF FIGURES

1.1 The box-Jenkins procedure. ....	15
3.1 Time plot of the number of subscribers.....	24
3.2 Plots of the autocorrelation of the series.....	26
3.3 Plot of the series after differencing.....	27
3.4 Plot of autocorrelation of the series after difference.....	28
3.5 Plot of partial autocorrelation of the series after differencing.....	28
3.6 Plot of residual autocorrelation function.....	31
3.7 Plot of residual partial autocorrelation function.....	32
3.8 Plot of the observed and forecast values with their confidence limits.....	36



## LIST OF TABLES

Table 3.1 Theoretical ACFs and partial ACFs.....	25
Table 3.2 ARIMA Models and their corresponding AIC and BIC.....	29
Table 3.3 ARIMA Model Parameters.....	30
Table 3.4 Confidence limits and their z values.....	33
Table A.1 Monthly Number of Subscribers.....	47
Table B.1 Autocorrelation of the Series.....	48
Table B.2 partial autocorrelation of the series.....	49
Table B.3 Autocorrelation of the Series after Differencing.....	50
Table B.4 partial autocorrelation functions after differencing.....	51
Table B.5 residual autocorrelation function.....	52
Table B.6 residual partial autocorrelation function.....	53
Table B.7 observed and forecast values for subscribers.....	54



## **CHAPTER 1**

### **INTRODUCTION**

With more and more subscribers employing the various services of telecommunication and other financial institutions using microwave and fiber optics networks from telecommunication service providers as the backbone that connects their various branches in geographically dispersed areas, it becomes very important that these telecommunication companies have adequate forecast model that projects these user requirements to network resources.

The above mentioned problems seriously call for a study that can model and forecast future demands in the telecommunication industry thus mapping the rise in customers including services that will be subscribed to needed telecommunication infrastructure.

#### **1.1 BACKGROUND TO STUDY**

Demand is the willingness of consumers to buy a product at a certain price, or the aggregate demand of all consumers in an economy for all products and services. This Correlation is displayed in relation to the global market by the sold product quantity at one time-point. In the telecommunication industry, demand modeling may be defined as calculating the demand for system resources within a telecomm network .With our attention on the telecommunication services sector we focus on demand in telecommunication facilities (resources) by customers without compromising on the quality of service (QoS) by the telecommunication operator. The purpose of this project is to model the increase in the number of customers that will demand telecom services from a leading Telecommunications network in Ghana so that the appropriate

infrastructure could be put in place to meet such an increase while maintaining quality of service.

The telecommunication market is changing rapidly, with widespread predictions that demand will grow by orders of magnitudes over the decades [10-12], and almost weekly reports of new collaborations between established telecommunication companies, computer firms and consumer electronic giants, rapid advancement in technology. Telecommunication operators therefore face a complex environment experiencing rapid change resulting in large uncertainties and consequent risk associated with any decision that might be taken. It is in this light that the thesis seeks to develop a model to solve the above problem.

A model is a mathematical representation of a process, device, or concept by means of a number of variables which are defined to represent the inputs, outputs, and internal states of the device or process, and a set of equations and inequalities describing the interaction of these variables [1].

Modeling can also be defined as Mathematical representation of reality that attempts to explain the behavior of some aspect of it. The mathematical model serves the following purposes:

- to find an optimal solution to a planning or decision problem
- to answer a variety of what-if questions
- to establish understandings of the relationships among the input data items within a model and
- attempt to extrapolate past data to derive meaning.

Mathematical models include techniques such as Linear Programming, computer simulation, Decision Theory, Regression Analysis, Economic Order Quantity (EOQ) and Break-Even Analysis [2].

In this work we employ the mathematics of modeling to forecast the future rise of subscribers for a telecommunications company.

## **1.2 PROBLEM STATEMENT**

Telecommunication Demand Modeling has been a problem both in Ghana and the global world and has been under study for a long period of time. Developments in new communication technologies such as mobile devices and developments of new services to meet user demand such as video, multi-media and data services has led to an exponential growth in the demand for these services. In Ghana, the national regulator, the National Communications Authority puts cellular phone penetration in the country at about seventy percent (70%). The sharp increase in the number of customers over a short period and limited infrastructure has led users to raise concerns over the quality of service been offered. It is in this light that this thesis attempts to solve a case study of the telecommunication demand modeling for a leading cellular company in Ghana.

Worldwide, different approaches have been employed in solving this very sensitive demand problem. The thesis seeks to use the Box- Jenkins methodology to develop a model that can adequately forecast the future rise in the subscriber base so that this can be carefully factored in planning so proper decisions can be taken to minimize this demand problem.

### **1.3 GENERAL OBJECTIVES**

The general objective of the thesis is to develop a model that will forecast future rise in subscribers for a telecommunication system with the specific objectives being the following

1. To collect traffic data from one of the leading telecommunication companies in the country
2. To find the appropriate tools for analyzing the data
3. To propose an efficient model that forecast rise in demand of telecommunication resources.

### **1.4 RELEVANCE OF THE RESEARCH**

Ghana as in other parts of the world has seen a dramatic rise in the use of mobile devices and has even become part of everyday life. This rise over a short period of time without any corresponding development in infrastructure, has led to a lot of concerns over the quality of services been offered these cellular operators. Again, the introduction of many operators and competitive technologies has driven prices downwards, thereby, compelling more users across the social divide and ages to subscribers the services for different uses such as e-trading and other financial transactions.

It is thus imperative to develop forecast models to predict the rise in the number of customers so as to enable telecom operators in the country provide adequate infrastructure to meet the corresponding demand so as to minimize the problem of poor quality of service.

### **1.5 BRIEF HISTORY OF THE TELECOMMUNICATION SERVICES IN GHANA**

The history of telecommunications in Ghana has been poorly documented as with many other African countries however it can be grouped under two main headings; the colonial and post-

independence era. Elements can be found spanning from 1881 by the installation of the Telegraph Embryo and later a telephone system by the British colonial administration with the primary goal of satisfying its own needs in coordinating its administrative colony [5].

In the post-independence era between 1963 and 1968, Ghana introduced the direct access telephone networks to link the regional capitals. Demand was so high that the government at that time launched the Telecommunication Modernization and Expansion Program with assistance from Japan, the World Bank and the African development Bank. In 1974, the department of Posts and Telecommunications was transformed from a ministerial department, the Civil Service Department of Posts and Telecommunications, to an incorporated public enterprise, Ghana Post and Telecommunications Corporations [5]. The telecommunications wing provided telephone services, telex and fax services as well as the transmission of television and radio programs of the government-owned Broadcasting Corporation (GBC).

1975 was a major boost for the telecommunication sector when Ghana launched the First Telecommunications Project FTP with an amount of seventy six million dollars from her developing partners that saw the installation of twelve (12) electronic exchanges throughout the country, a telex switch, a satellite gateway and an extensive network of radio links [6]. FTP made possible the transmission of TV programs throughout the country and also provided international communications to overseas and neighboring African countries.

FTP was followed by Second Telecommunications Project STP with an estimated cost of one hundred and seventy three million dollars over the period of 1987 to 1994 with two main objectives; to continue with the rehabilitation, expansion and modernization of the network that began with FTP and to finance the institutional reforms of the sector [5-6].

Ghana deregulated its telecommunications sector in 1994 when the Government announced a five-year comprehensive restructuring of the industry known as the Accelerated Development Program 1994-2000(ADP 2000). The main policy objectives of the program were formulated with the assistance of the World Bank, consultants and other stakeholders,

The ADP adopted the following strategies to achieve the above-stated policy objectives [5][7].

- The authorization of two national network operators: Ghana Telecom and a new independent operator;
- Support of new financing: arrangements which promote investment in new telecommunications infrastructure throughout the country; and
- Privatization of Ghana Telecom through the sale of a strategic stake to an international operating company combined with measures to broaden share ownership in Ghana.

In June 1995, as part of the ADP reform program the telecommunication operator, Ghana Telecom was incorporated as a successor to the telecommunications division of Ghana Posts and Telecommunications Corporation (GPTC).

In December 1996, Ghana Telecom privatized its main line operations by awarding a Malaysian-led Consortium (Telecom Malaysia) a thirty percent (30%) stake in the state company with full management control for thirty eight million.

The National Communication Authority (NCA) was established by Parliamentary Act 1996 as a central regulatory body to regulate the telecommunications sector and to promote a stable operating environment for all participants, while also promoting fair competition and efficiency.

The main task of the NCA includes the licensing and regulation of telecommunications system operators and assigning or allocating systems frequencies.

The Ministry of Communications was created in 1997 to facilitate the strategic development and application of the use of the various communications resources; human, material and technological for effective communications throughout the country. The Ministry (acting on behalf of the Government of Ghana) has since then embarked on a process aimed at formulating a workable national communications policy for Ghana. Ghana may have paved the way in telecoms liberalization, but still faces many challenges getting its telecom infrastructures and facilities up to international standards [8].

A consortium of African Communications Group, led by Western Wireless Company based in Cambridge, USA and Ghana National Petroleum Company, won the bid as the second network operator with an offer of USD 10.1 million. The consortium, which trades under the name WESTEL, planned to invest between USD forty (40) million and USD seventy (70) million over a period of five years from its incorporation. The licensing of a second national operator was expected to meet the current backlog of three hundred thousand (300 000) telephones lines being demanded by consumers and to ensure efficiency and improved services. WESTEL was required to have hundred (100) payphones in service by the end of 1998, and the company also planned to have fifty thousand (50 000) customers within three years of its incorporation [6].

Millicom Ghana, a subsidiary of Millicom International, UK/Luxembourg, operating now under the name of Tigo started its operations in 1991 and was the first cellular network operator. Millicom Ghana had over twenty two thousand (22 000) subscribers in 1998 with a market share of above seventy (70) per cent of the mobile market. The company CellTell, owned by Kludjeson International, also started its operations in Ghana in 1993 using the AMPS technology, with coverage in Accra and Tema. OneTouch now Vodafone owned by Vodafone was the cellular arm of Ghana Telecom. It started its operations in 2000 providing nation-wide

cellular services using GSM technology. The second national operator, was WESTEL now Espresso and privatized. Scancom now MTN and the biggest telecommunication company in the country started operating in October 1996 using GSM 900 technology, with fifteen (15) sites and equipment from Ericsson [9].

Zain Ghana now Airtel also began operations as the fifth mobile telecom operator in the country in 2008. Globacom popularly known as Glo the sixth mobile telecom operator is set to start operations in the country before the end of the year ending 2010 after it successfully landed its GLO1 fiber optic cable from Europe to the shores of Ghana.

## **1.6 THESIS ORGANISATION**

The thesis is divided into four (4) main chapters.

Chapter 1 gives a preview of the thesis topic under consideration. It also gives a brief history of telecommunication services in Ghana.

Chapter 2 focuses on related literature to the scope of the study in the thesis.

Chapter 3 looks at the solution that has been employed in solving the defined problem thus the Box-Jenkins methodology that implements our model to the solution of the demand problem.

Chapter 4 which is Conclusions and Recommendations for Future Work summarizes what has been developed and also discuss works that needs to be done in future to further enhance the subject under study.

## CHAPTER 2

### LITERATURE REVIEW

#### 2.1 REVIEW OF RELATED WORKS

The telecommunication market is changing rapidly, with widespread predictions that demand will grow by orders of magnitudes over the decades [10-12]. It is thus imperative to develop a model that will forecast this future rise so as to maintain/improve upon the quality of service (QoS) offered.

The model can be constructed using standard statistical methods such as the Box-Jenkins methodology or by applying artificial intelligence methods like neural networks [39].

Research has shown that Neural Network Models produce forecasts that are more accurate than those from other methods [10], however, current advice is to avoid them due to:

- the method ignores prior knowledge
- the result are difficult to understand [10-12].
- much data is needed in order to estimate neural network models and to reduce the risk of over-fitting the data [10].

The Box-Jenkins methodology (ARIMA) was chosen as the model for this work owing to the fact that that it is suitable for any time series with any pattern of change and it does not require the developer to choose a priori value of any parameter like other models for predicting the magnitude of any variable, it was chosen as the model for the work. However its limitations include its requirement of a long time series.

## 2.2 APPROACHES TO TIME SERIES FORECASTING.

Time series models help to predict the future values of time series based on its past behavior. The time series data should be random. A simple random model is one where the observation  $Y_t$  consists of two parts [13] the overall arithmetic mean  $\mu$  and a random error component  $e_t$ , given by the relation,

$$Y_t = \mu + e_t \quad (1)$$

### 2.2.1 INTRODUCTION TO BOX-JENKINS METHODOLOGY

Unlike classical decomposition methods, which deal with systematic time series components (trend, circular and seasonal), the Box-Jenkins methodology deals with a residual (unsystematic) component [3]. This method involves searching for relations of individual observations. By this method we are able to describe time series that are not manageable by standard methods.

The time series is perceived as realization of the stochastic process which is defined as a series of random quantities arranged in time  $\{Y(s, t), s \in S, t \in T\}$  where  $S$  is a selective space and  $T$  is an index series. For each  $s \in S$  the realization of the stochastic process is defined on the index series  $T$  [4].

In [3] the Box-Jenkins methodology makes use of the following

1. Autoregressive process AR
2. Moving average process MA and the combination of the two called
3. Autoregressive moving average, ARMA.

These processes are as a result from the linear process of resetting all parameters till the final number. The parameters are chosen so as to ensure the stationary and invertibility of the process.

A special non-stationary Autoregressive Integrated Moving Average (ARIMA) is used in the Box-Jenkins methodology when the observed time series is not stationary [3].

## 2.2.2 COMPONENTS OF TIME SERIES

A time series is a chronological sequence of observations on a particular variable  $Y$  ( $Y_1, Y_2, \dots, Y_n$ ) ordered according to a discrete index variable  $t$  ( $1, 2, \dots, n$ ). The index variable  $t$  refers to intervals of time measured in a suitable scale [14]. In general a given time series has four main components that can be recognized and separated [15],

1. A regular, long-term component of variability, termed trend, that represent the whole pattern of the series
2. A regular, short-term component whose shape occurs periodically at intervals of  $s$  lags of the index variable termed seasonality
3. An AR( $p$ ) autoregressive component of  $p$  order, which relates each value  $Z_t = Y_t - (\text{trend and seasonality})$  to previous  $p$  values according to the following linear relationship

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + \varepsilon_t \quad (2)$$

Where  $Z_t$  is the response variable at time  $t$ ,  $Z_{t-p}$  is observation at time  $t-p$ ,  $\phi_i$  is the regression coefficients to be estimated and  $\varepsilon_t$  is the error or residual term at time  $t$ .

4. A MA( $q$ ) moving average component of order  $q$ , which relates each  $Z_t$  value to the  $q$  residuals of the  $q$  previous  $Z$  estimates also having the linear relationship

$$Z_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (3)$$

Where  $Z_t$  is the response variable at time  $t$ ,  $\theta_t$  is the regression coefficients to be estimated,  $\varepsilon_t$  is the error or residual term at time  $t$  and  $\varepsilon_{t-q}$  is the error term at time  $t-q$  [15].

### 2.2.3 LAG OR BACKWARD LINEAR OPERATOR (B)

From [5] the lag or backward linear operator B can be defined mathematically as

$$BZ_t = Z_{t-1} \quad (4)$$

If the lag operator is applied twice on a series, then we have the relation

$$B(BZ_t) = B(Z_{t-1}) = Z_{t-2} \quad (5)$$

The double application of the linear backward operator is indicated by  $B^2$ , and in general, for any integer  $k$ , it can be written as

$$B^k Z_t = Z_{t-k} \quad (6)$$

By using the backshift operator, equation (2) can be rewritten as

$$Z_t = \phi_1 BZ_t + \phi_2 B^2 Z_t + \dots + \phi_p B^p Z_t + \varepsilon_t \quad (7)$$

$$Z_t - \phi_1 BZ_t - \phi_2 B^2 Z_t - \dots - \phi_p B^p Z_t = \varepsilon_t \quad (8)$$

$$\phi(B)Z_t = \varepsilon_t \quad (9)$$

Where  $\phi(B)$  is the autoregressive operator of  $p$  order defined by

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (10)$$

By using the same analogy, equation (3) can also be rewritten as

$$Z_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} = \theta(B)\varepsilon_t \quad (11)$$

Where  $\theta(B)$  indicates the moving average operator of order  $q$  defined by

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (12)$$

The AR autoregressive part and the MA moving average part can be combined to form autoregressive moving average, ARMA (p, q) models [16]

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (13)$$

Or in the lag operator form

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Z_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t \quad (14)$$

This can also be rewritten as [15]

$$\phi(B) Z_t = \theta(B) \varepsilon_t \quad (15)$$

## 2.2.4 THE DIFFERENCE OPERATOR

In the analysis of time series it is important to separate the trend and the seasonality. The seasonal component can be estimated by seasonal decomposition procedures, which calculates a seasonal index based on the ratio of the observed values of the moving average [17]. The trend component can be evaluated by fitting a regular function, a polynomial, or a more complicated general function.

In the final stage of the modeling process, however both the trend and the seasonal component can be integrated in the ARMA (p, q) process and this approach will be used in this work. For the trend the integration is obtained by using the linear difference operator  $\nabla$  defined by

$$\nabla Y_t = Y_t - Y_{t-1} = Y_t - B Y_t = Y_t (1 - B) \quad (16)$$

A single application of the  $\nabla$  operator corrects the data for a linear increasing trend, whereas its repeated use for  $d$  times corrects a trend that can be fitted by a  $d$  order polynomial [15].

The stationary series  $Z_t$  obtained as the  $d^{th}$  difference ( $\nabla^d$ ) of  $Y_t$  is written as

$$Z_t = \nabla^d Y_t = (1 - B)^d Y_t \quad (17)$$

The combined use of the  $\nabla$  operator and the ARMA (p, q) process results in ARIMA (p, d, q) model thus autoregressive integrated moving average. ARIMA can account for the seasonal component of  $s$  lag period, by using both correlations between  $Z_t$  and  $Z_{t-s}$  values and those between the corresponding residual  $\varepsilon_t$  and  $\varepsilon_{t-s}$ . In mathematical terms, therefore, a seasonal ARIMA model is an ARIMA (p, d, q) [18] model whose residuals  $\varepsilon_t$  can be further modeled by an ARIMA (P,D,Q)<sub>s</sub> structure with linear operators (P,D,Q) being functions of the  $B^s$  operator.

The operators of a seasonal ARIMA model, defined as  $(p, d, q)^{\times} (P, D, Q)_s$ , can be expressed as follows:

$$\text{AR (p) nonseasonal operator of p order, } \phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (18)$$

$$\text{AR (P) seasonal operator of P order, } \phi(B) = 1 - \phi_1 B^s - \dots - \phi_p B^{sp} \quad (19)$$

$$\text{MA (q) nonseasonal operator of q order } \theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (20)$$

$$\text{MA (Q) seasonal operator of Q order, } \theta(B) = 1 - \theta_1 B^s - \theta_2 B^{2s} - \dots - \theta_Q B^{Qs} \quad (21)$$

and difference operator of d order [15].

$$\nabla^d = (1 - B)^d \quad (22)$$

In this work, the Box-Jenkins model-building strategy was utilized. The box-Jenkins methodology consists of a four-step iterative procedure [19];

1. model identification,
2. parameter estimation,
3. model diagnostics and

#### 4. forecasting.

This procedure can be represented diagrammatically as in the flow chart in Figure 1.1.

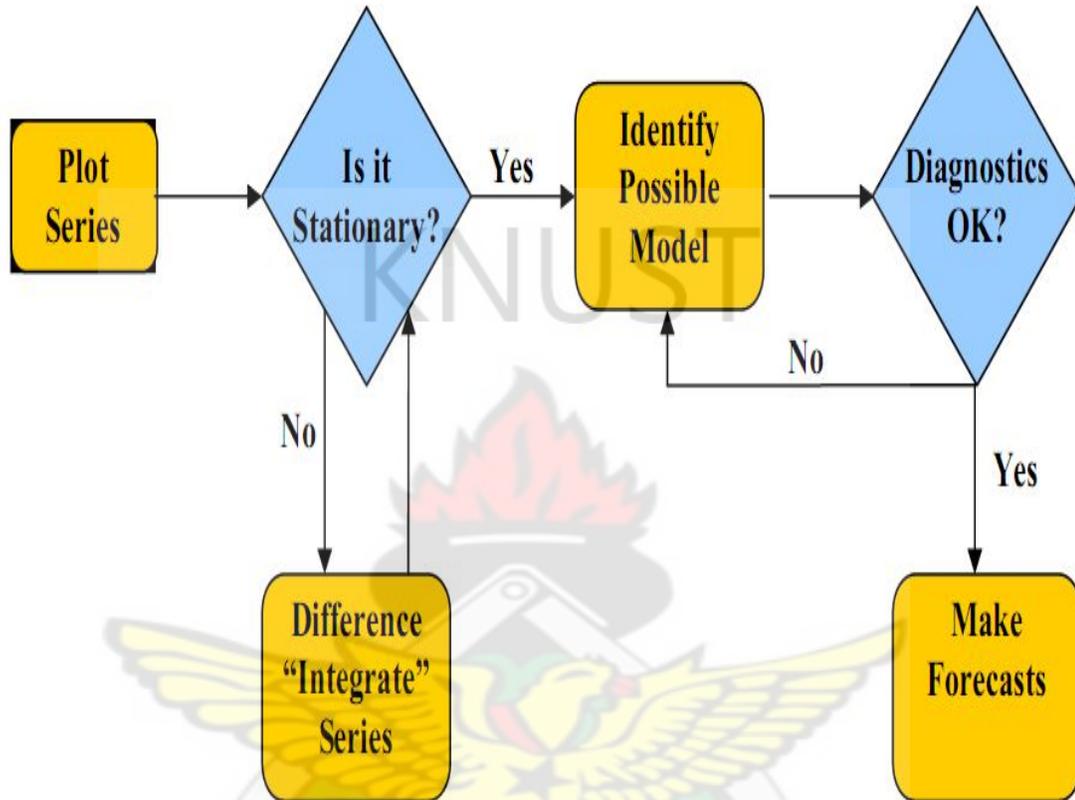


Figure1.1 The box-Jenkins procedure [19]

In model identification, the historical data are used to tentatively identify an appropriate Box-Jenkins model. We look at the time plot of the series and compute many important statistics of the data, such as the sample autocorrelation function and the sample partial autocorrelation function as the fundamental analytical tool to tentatively choose a model [3].

Both autocorrelation function (ACF) and partial autocorrelation function (PACF) are very important for the definition of the internal structure of the analyzed series [15]. The ACF  $\rho(k)$  at lag  $k$  of the  $Z_t$  series is the linear correlation coefficient between  $Z_t$  and  $Z_{t-k}$ , calculated for  $k = 0, 1, 2, \dots$  as in equation (23)

$$\rho_k = \frac{\text{cov}(z_t, z_{t-k})}{\sqrt{\text{var}(z_t) \text{var}(z_{t-k})}} \quad (23)$$

The PACF is defined as the linear correlation between  $Z_t$  and  $Z_{t-k}$ , controlling for possible effects of linear relationships among values at intermediate lags. Theoretically, both an AR (p) process and an MA (q) process should be associated with well-defined patterns of ACF and PACF, usually decreasing exponential; alternate in sign or decreasing sinusoidal patterns. A precise correspondence between ARMA (p, q) processes and defined ACF and PACF patterns is more difficult to recognize. When the order of at least one of the two components (AR or MA) is clearly detectable, however, the other can be identified by attempts in the following step of parameter estimation. Finally, the existence of a seasonal component of length s is underlined by the presence of a periodic pattern of period s in the ACF.

Once a suitable ARIMA  $(p, d, q) \times (P, D, Q)_s$  structure is identified, subsequent steps of parameter estimation and model validation must be performed. Parameter estimates are usually obtained by maximum likelihood. Estimators are usually sufficient, efficient, and consistent for Gaussian distributions and are asymptotically normal and efficient for several non-Gaussian distribution families. Validation of the goodness of fit of an ARIMA model can be developed according to the following steps:

1. Evaluation of statistical significance of parameters by the usual comparison between the parameter value and the standard deviation of its estimate. For a test statistic that is valid only asymptotically, a parameter whose value exceeds twice its standard error can be considered significant.
2. Analysis of the ACF of residuals. In this step, residuals  $\varepsilon_t$  are considered as a new time series, and ACF and PACF are estimated to be sure that values at lag  $k > 0$  are not statistically

different from zero. For prediction purposes, ARIMA models are different from the analytical functions of time:  $Z_t = f(t)$ , because ARIMA forecasting uses previous values of the series and errors in the previous estimates. Actually, this peculiarity of ARIMA forecasting is valid in the short term because parameters of the model cannot account, in the long term, for changes in the dynamics of the series [20].

Parameter estimation consists of finding the best possible estimates for the parameters of the tentatively identified model. In this stage, methods of estimation such as the method of moments, least-squares estimators and maximum likelihood estimators are considered to estimate the parameters [21].

In model validation, various diagnostics such as the method of autocorrelation of the residuals and the Ljung-Box-Pierce statistics are used to check the adequacy of the tentatively identified model. If the model is found to be inappropriate, we would return back to model identification and cycle through the steps until, ideally, an acceptable model is found.

Once a final model is obtained, it can be used to forecast future time series values.

The box-Jenkins methodology requires that the model to be used in describing and forecasting a time series to be both stationary and invertible [21]. Thus, in order to tentatively identify a box-Jenkins model, we must first determine whether the time series we wish to forecast is stationary. If it is not, we must transform the time series into a series of stationary time series values through the process of differencing. A time series is said to be stationary if the statistical properties such as the mean (first moment) and the variance (second moment) of the time series are essentially constant through time. From the plot of the time series values, if the observed values of a time series seem to fluctuate with a constant variation around a constant mean, then it is reasonable to believe that the series is stationary, otherwise, it is said to be nonstationary. We will also look at

the behavior of the sample autocorrelation function, and the partial autocorrelation function to decide whether the time series is stationary.

Any Box-Jenkins model can be used to express  $Z_t$ , an observation made at time  $t$ , as a function of past  $z$  – observations. A Box-Jenkins model is considered not invertible, if the weights placed on the past  $z$ - observations when expressing as a function of these observations do not decline as we move further into the past. A model which is invertible on the other hand, implies that these weights do decline. Intuitively, this condition should hold, since it seems logical that a recent observation should count more heavily than a more distantly past observations. The condition of stationarity and invertibility implies that the parameters used in the model under consideration satisfy certain criteria. When we obtain the final least squares point estimates of the parameters in our model, we should verify that these point estimates satisfy the stationarity and invertibility conditions. The model will be considered inadequate if those conditions are not met [15].

### **2.2.5 PURELY SEASONAL MODELS**

A purely seasonal time series is one that has only seasonal autoregressive (AR) or moving average (MA) parameters. For purely seasonal models the autocorrelation die down and partial autocorrelation cut off after one season lag for an SAR (1) model and after two seasonal lag for an SAR (2) model.

### **2.2.6 SEASONAL AUTOREGRESSIVE MODEL (SAR)**

Seasonal autoregressive models are developed with parameters known as seasonal autoregressive (SAR) parameters. The SAR parameters represent autoregressive relationships that exist between

time series data separated by multiples of the number of periods per season. A general seasonal autoregressive model with  $p$  SAR parameters is of the form

$$Y_t = \phi_s Y_{t-s} + \phi_{2s} Y_{t-2s} + \phi_{3s} Y_{t-3s} + \dots + \phi_{ps} Y_{t-ps} + e_t \quad (24)$$

With  $p$  SAR parameters, the order of the model is  $ps$ .

### 2.2.7 SEASONAL MOVING AVERAGE (SMA)

Seasonal moving average models are constructed with seasonal moving average parameters (SMA parameters). SMA parameters represent moving average relationships that exist among the time series observations separated by a multiple of the number of period per season. The general seasonal moving average model with  $q$  parameters is of the form

$$Y_t = \varepsilon_t + \theta_s \varepsilon_{t-s} + \theta_{2s} \varepsilon_{t-2s} + \dots + \theta_{qs} \varepsilon_{t-qs} \quad (25)$$

### 2.2.8 THE YULE WALKER EQUATION

The Yule Walker equations are set of linear equations that relate the parameters of an AR model with the autocorrelation sequence. It is used to estimate the parameters  $\phi_1, \phi_2, \dots, \phi_p$  in terms of autocorrelations [18]. The general form of the Yule Walker equation is

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad (26)$$

$$\text{If } p=2, \rho_1 = \phi_1 \rho_0 + \phi_2 \rho_1 \quad (27)$$

$$\text{and } \rho_2 = \phi_1 \rho_1 + \phi_2 \rho_0 \quad (28)$$

Solving for  $\phi_1$  and  $\phi_2$  and taking  $\rho_0 = 1$

$$\phi_1 = \frac{\rho_1(1-\rho_2)}{1-\rho_1^2} \quad (29)$$

$$\phi_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \quad (30)$$

Using the estimates  $r_k$  for  $\rho_k$ , the following results are obtained

$$\phi_1 = \frac{r_1(1 - r_2)}{1 - r_1^2} \quad (31)$$

$$\phi_2 = \frac{r_2 - r_1^2}{1 - r_1^2} \quad (32)$$

For higher values  $\rho$  of it is expressed in matrix form as:

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \dots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_k \end{bmatrix} \quad (33)$$

$\rho_k$  is  $k \times 1$  and  $k \times p$ ,  $\phi_k$  is  $k \times 1$  matrices respectively. Thus the solution to the Yule Walker equations for the parameters  $\phi$  in terms of the autocorrelations may be written as

$$\phi = \rho_k^{-1} \rho_k \quad (34)$$

### 2.2.9 AKAIKE INFORMATION CRITERIA (AIC)/ BAYESIAN INFORMATION CRITERIA (BIC)/ SCHWARZ-BAYESIAN INFORMATION CRITERIA (SBC)

$$\text{AIC is given by } AIC = (-2 \log L + 2m) \quad (35)$$

Where  $m = p + q + Q$  and  $L$  is the likelihood function. Since  $-2 \log L$  is approximately equal to  $\{n(1 + \log 2\pi) + n \log \sigma^2\}$  where  $\sigma^2$  is the model MSE, AIC can be written as

$$AIC = \{n(1 + \log 2\pi) + n \log \sigma^2 + 2m\} \quad (36)$$

and because the first term in this equation is a constant, it is usually omitted while comparing between models [22, 23]. As alternatives to AIC, sometimes Bayesian Information Criterion

(BIC) and Schwarz-Bayesian Information Criteria (SBC) are also used [21, 24], which is given by;

$$BIC = n \ln(\delta_e^2) + k \ln(n) \quad (37)$$

Where  $\delta_e^2$  is the error variance and  $n$  the sample size.

$$SBC = \log \sigma^2 + (m \log n) / n \quad (38)$$

In this work however, both AIC and BIC were used in order to come out with a more precise model.

### 2.2.10 THE MEAN ABSOLUTE ERROR (MAE)

The Mean Error is the arithmetic average of all prediction errors. It is given by,

$$ME = \frac{1}{N} \sum e_t \quad (39)$$

Where  $N$  is the sample size and  $e_t$  is the error term.

## 2.3 METHODOLOGY

This thesis employed a model based on problem information and real data from a leading telecommunication company in Ghana subjected to the condition of anonymity and for educational purpose only since such data is seen as confidential and crucial for the survival of the organization and other reasons best known to them.

The methodology that shall be employed to solve the problem are but not limited to the following.

1. Advance lectures on demand modeling

2. Intensive literature review on demand modeling
3. Data collection and analysis
4. Data manipulation using the Box-Jenkins Methodology , Autoregressive Models (AR), Moving Average Models(MA), Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA).
5. Enough literature on demand modeling in telecommunications, statistics, digital signal processing and Telecommunications.

After the collection of relevant data, the Box-Jenkins methodology was used to manipulate the data and simulated with SPSS Version 17 and Minitab version 16 on an HP Pavilion TX2000 notebook with an AMD Turion™ X2 Dual-Core Mobile RM-72 processor 2.1GHz CPU speed to solve the proposed problem.



## **CHAPTER 3**

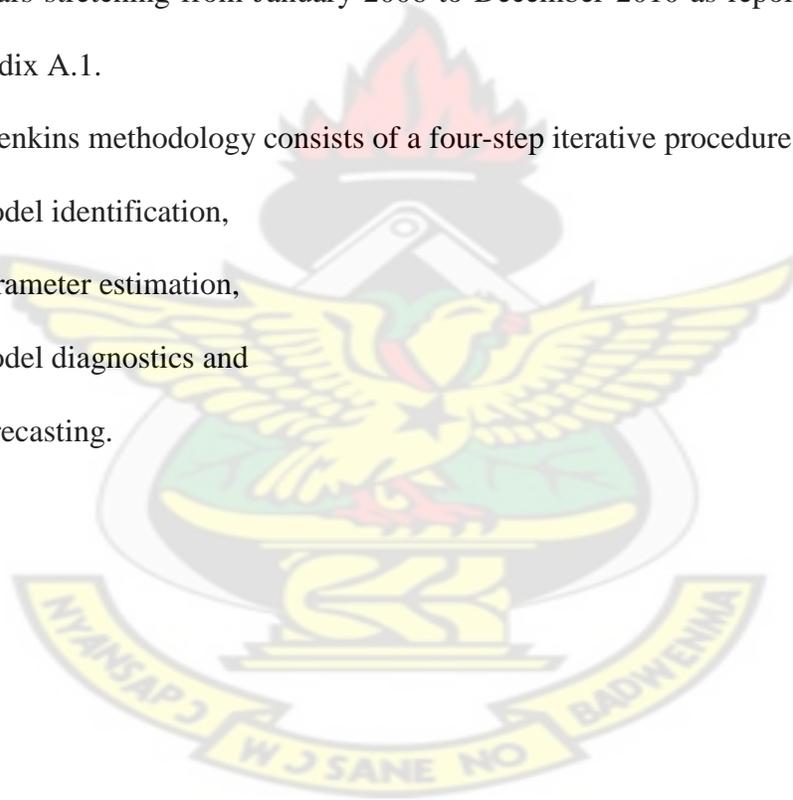
### **MODEL DEVELOPMENT AND IMPLEMENTATION**

#### **3.1 DATA FOR THE MODEL**

The thesis employed data from one of the telecommunication companies in the country who gave the data subjected to the condition of anonymity and for educational purpose only. The data is the monthly number of subscribers that joined the network for a period of five years stretching from January 2006 to December 2010 as reported in Figure 3.1 and appendix A.1.

The box-Jenkins methodology consists of a four-step iterative procedure [19];

1. model identification,
2. parameter estimation,
3. model diagnostics and
4. forecasting.



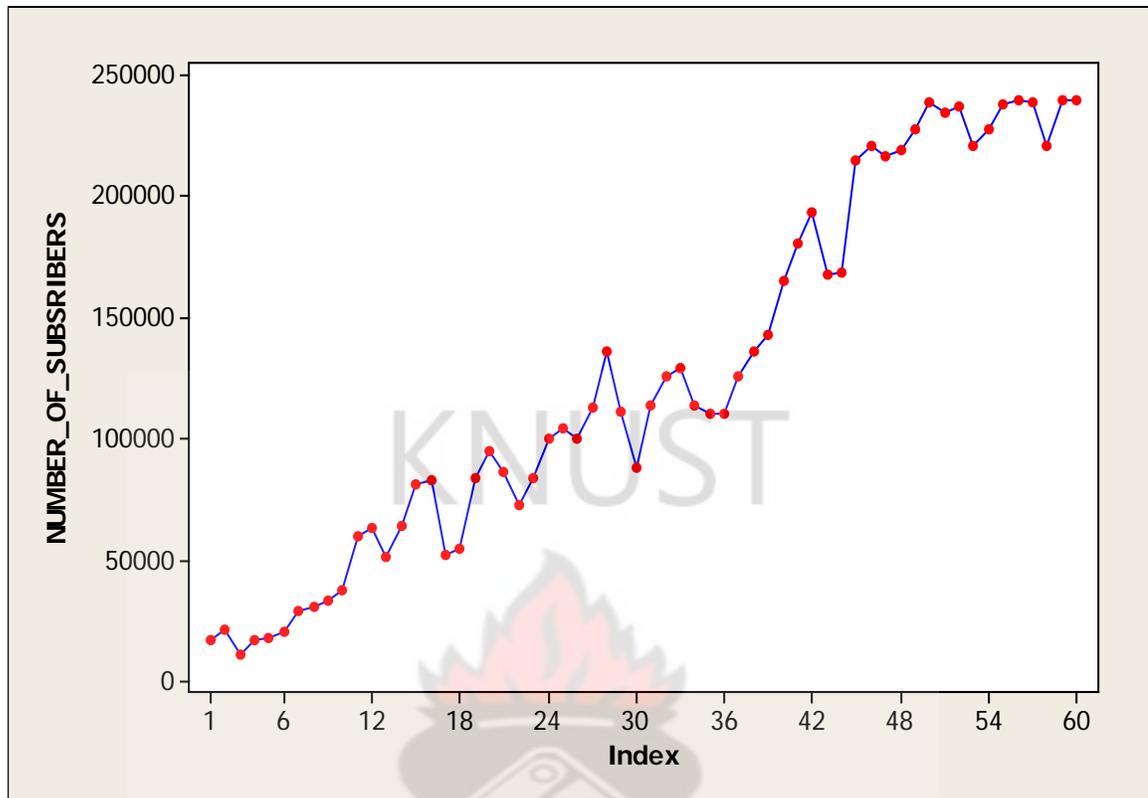


Figure 3.1 Time plot of the number of subscribers

### 3.2 MODEL IDENTIFICATION

ARIMA models can only be developed after transforming the variable under consideration into a stationary one. A time series is said to be stationary if its underlying generating process is based on a constant mean and constant variance with its autocorrelation function (ACF) essentially constant through time [32]. Thus, if we consider different subsets of a realization (time series sample) the different subsets will typically have means, variances and autocorrelation functions that do not differ significantly. There are several ways to ascertain stationarity. One is by observing the graph or the plot of the series and another is by examining the plot of the autocorrelation and partial autocorrelations of the data [18].

Theoretical ACFs and PACFs (Autocorrelations versus lags) are available for the various models chosen [18] for various values of orders of autoregressive and moving average components i.e.  $p$  and  $q$ . where the correlograms (plot of sample ACFs versus lags) obtained from the given time series data are compared with these theoretical ACF or PACFs, to find a reasonably good match and tentatively select one or more ARIMA models. The general characteristics of theoretical ACFs and PACFs are as in Table 3.1.

Table 3.1 Theoretical ACFs and partial ACFs [33]

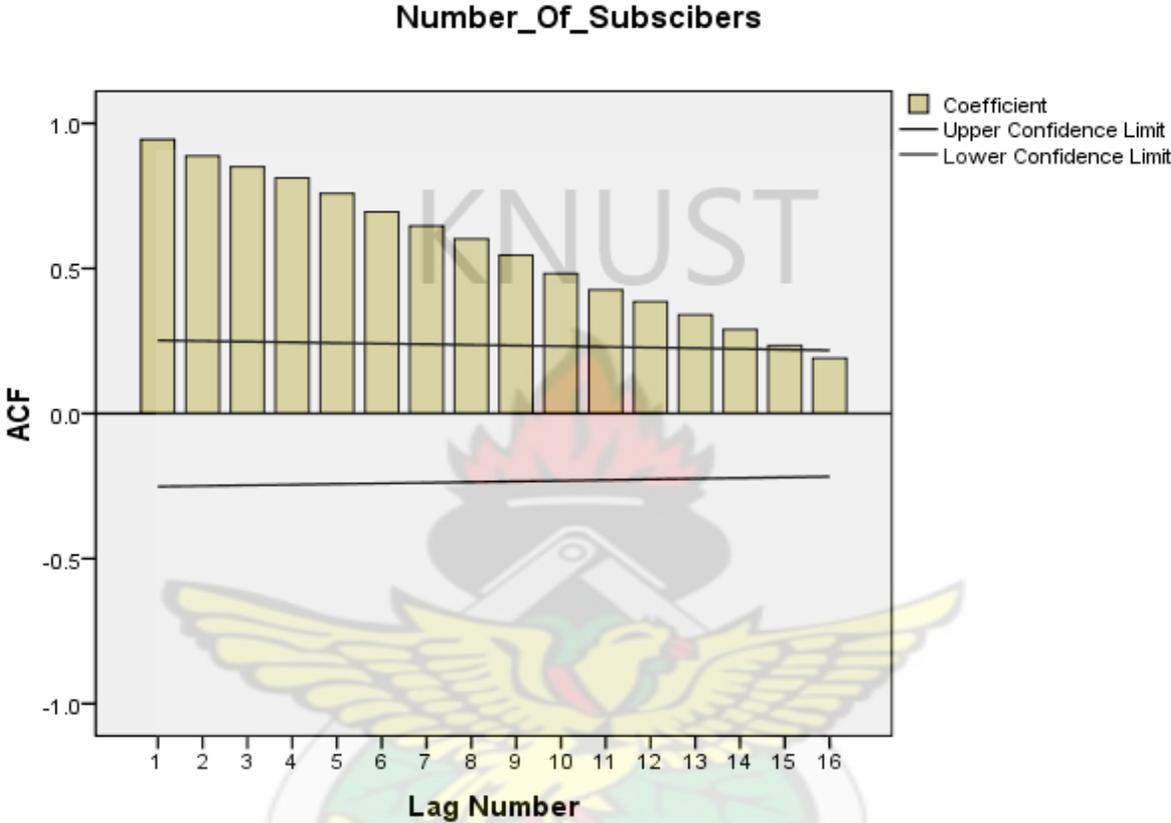
Model	ACF	PACF
AR(p)	Die out	Cut off after the order $p$ of the process
MA(q)	Cut off after the order $p$ of the process	Die out
ARMA(p, q)	Die out	Die out

In this work an observation of the plots of the series and that of the autocorrelation were used in determining the stationarity of the series.

From the Figure 3.1, the series show an upward trend which is an indication that the series is not stationary. The plot of the autocorrelation function can be observed from Figure 3.2 that the first several autocorrelations are persistently large and trails off to zero rather slowly which is also an indication that a trend exists and that this time series is nonstationary.

Since the observed series is not stationary we need to difference the data to achieve stationarity.

Figure 3.2 Plot of the autocorrelation of the series



The plot of the series after differencing is reported in Figure 3.3

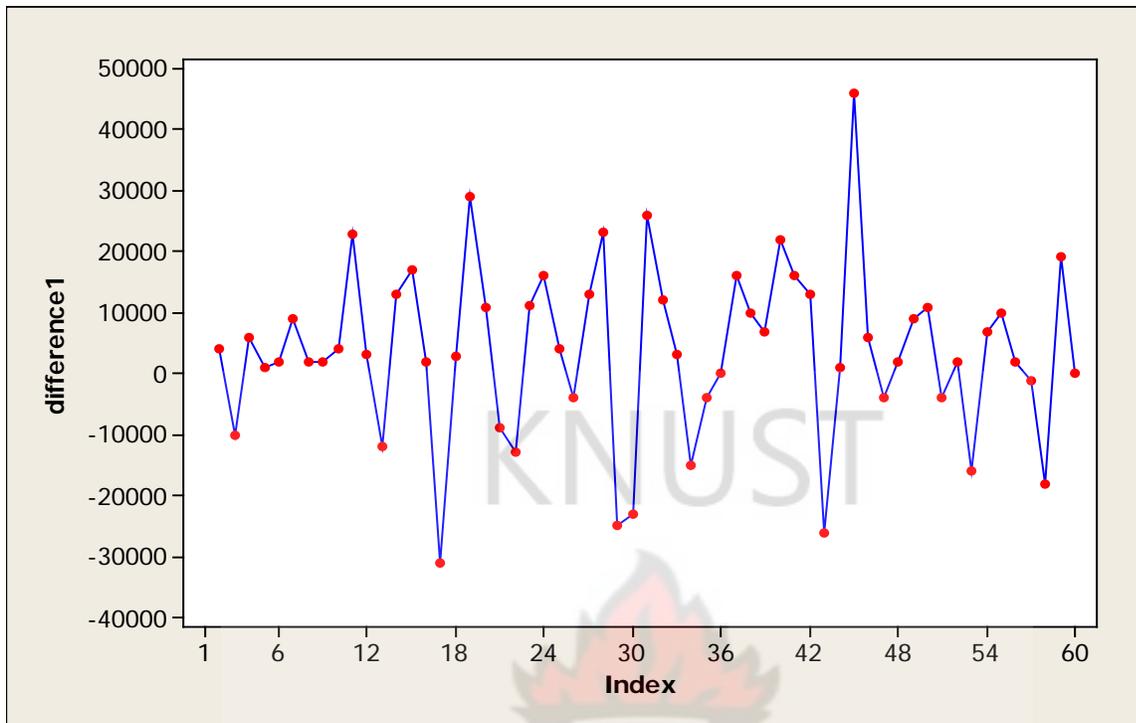


Figure 3.3 Plot of the series after differencing

The plot of the differenced data appears to vary about a fixed level. From Figure 3.4 which is the plot of the autocorrelation function after differencing, it can be seen that it trails to zero rapidly and also the ACF appears to cut off after lag 2, suggesting an MA (2) behavior. Again only lag 2 from the partial autocorrelation plot in Figure 3.5 is significant, the PACF appears to cut off after lag 2, indicating AR (2) behavior.

Since stationarity was obtained after the first difference it suggests an ARIMA (2, 1, 2) model.

In practice, though, the value of the moving average component,  $q$  rarely exceeds 2 [33]. In this work five tentative models shall be considered and they are;

ARIMA (0, 1, 2)

ARIMA (1, 1, 2)

ARIMA (2, 1, 0)

ARIMA (2, 1, 1) and

ARIMA (2, 1, 2)

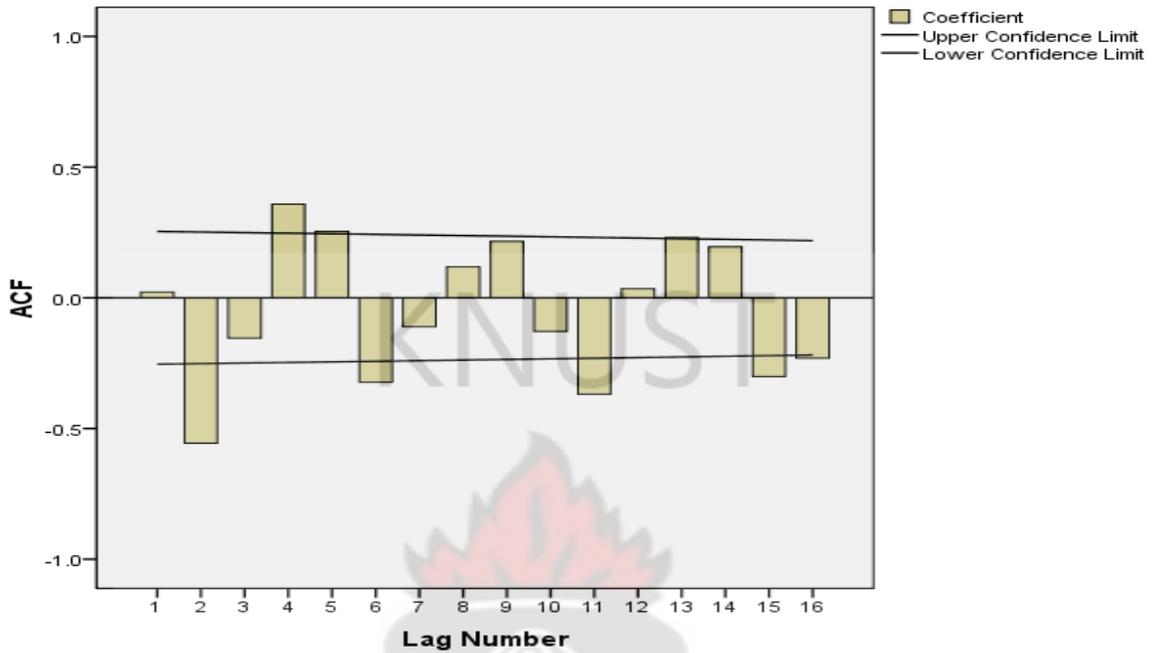


Figure 3.4 plot of autocorrelation of the series after differencing

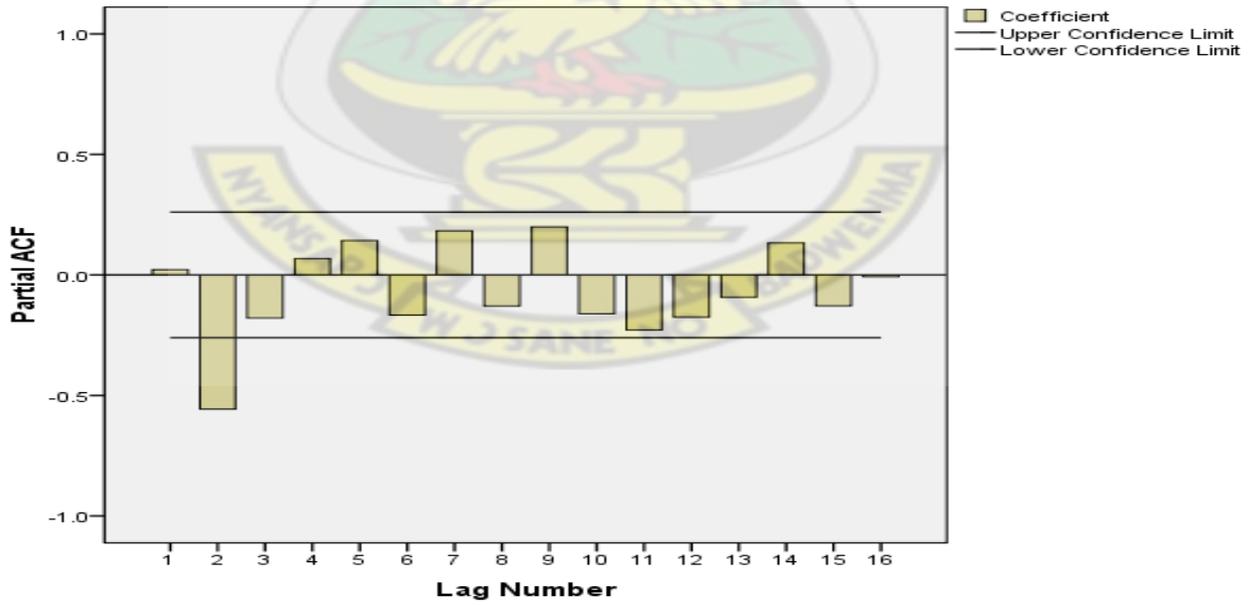


Figure 3.5 plot of partial autocorrelation of the series after differencing

From [3] the principle of parsimony states that “all things been equal, simple models are preferred to complex models” the model with the minimum Akiake Information Criterion (AIC) and Bayesian Information Criterion (BIC) values will be considered.

### 3.3 SELECTING FROM COMPETING MODELS

The five models and their corresponding AIC, BIC and Mean Absolute Errors (MAEs) are shown in Table 3.2. From Table 3.2 the most suitable model is ARIMA (2, 1, 0) since it has relatively the lowest AIC and BIC values.

Table 3.2 ARIMA Models and their corresponding AIC and BIC

Model	AIC	BIC	Mean Absolute Error
ARIMA(0,1,2)	1277.21444624808	1283.44705857979	8845.5556020992
ARIMA(1,1,2)	1281.26038195211	1289.57053172774	9076.45572491041
ARIMA(2,1,0)	1276.22610482151	1282.45871715323	8405.29948635534
ARIMA(2,1,1)	1277.24271364289	1285.55286341851	8306.26208973975
ARIMA(2,1,2)	1277.96015107261	1288.34783829213	8203.3120975769

### 3.4 PARAMETER ESTIMATION

Once an appropriate model has been chosen the next step is estimating the regression coefficients that minimize the mean square error. An iterative non-linear least square procedure is applied to the parameter estimates of an ARMA (p, q) model. The method minimizes the sum of squares of

error  $\sum \varepsilon_t^2$  given the form of the model and data. The estimates usually converge on an optimal value with a small number of iterations. There is however software packages available for estimating these values such as Minitab, SPSS, SAS, X12 and MATLAB. In this work the model parameters were estimated using Statistical Package for Social Sciences (SPSS) Analyze version 17 and Minitab 16. The results are shown in Table 3.3

Table 3.3 ARIMA Model Parameters.

	<b>Estimates</b>		<b>Standard Error</b>	<b>Significance</b>
<b>Non-seasonal lags</b>	AR1	.042	.110	.699
	AR2	-.564	.111	0.000
	Constant	3835.022	1011.689	0.000

### 3.5 MODEL DIAGNOSTICS AND FORMULATION

This third stage of the ARIMA process is concerned with checking the residuals of the model to see if they contain any systematic pattern which can still be removed to improve on the chosen ARIMA. This is achieved by examining the autocorrelations and partial autocorrelations of the residual [34, 35]. In this work the various correlations up to sixteen (16) lags were computed and the same along with their significance which is tested by Box-Ljung test. The results are reported in Table B.5 and B.6 in Appendix B. As the results indicate, none of these correlations is significantly different from zero at a reasonable level. This is an indication that the selected ARIMA model is an appropriate model [15]. The ACF and PACF of the residuals as shown in Figures 3.6 and 3.7 also indicate a “good fit” of the model since most of the sample

autocorrelation coefficients of the residuals are within the limits  $\pm 1.96/\sqrt{60}$  which computes to  $\pm 0.253$ . Where 60 is the number of observations upon which the model is based. In other words the residuals are white noise, thus showing no significant pattern. Hence the fitted model of the subscriber data is given by

$$Z_t = Z_{t-1} + \phi_1(Z_{t-1} - Z_{t-2}) + \phi_2(Z_{t-2} - Z_{t-3}) + \varepsilon_t \quad (44)$$

And by substituting the estimated values, our final model equation is given by

$$Z_t = Z_{t-1} + .042(Z_{t-1} - Z_{t-2}) - .564(Z_{t-2} - Z_{t-3}) + 3835.022 \quad (45)$$

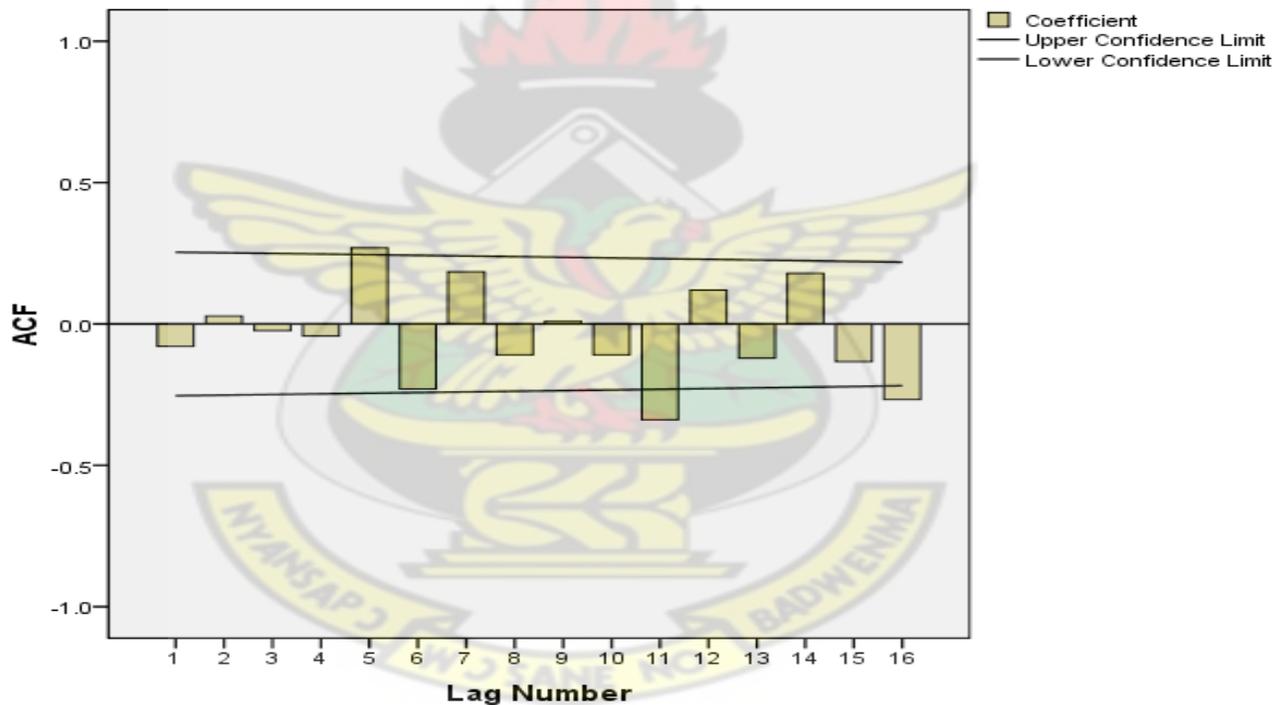


Figure 3.6 Plot of residual autocorrelation function

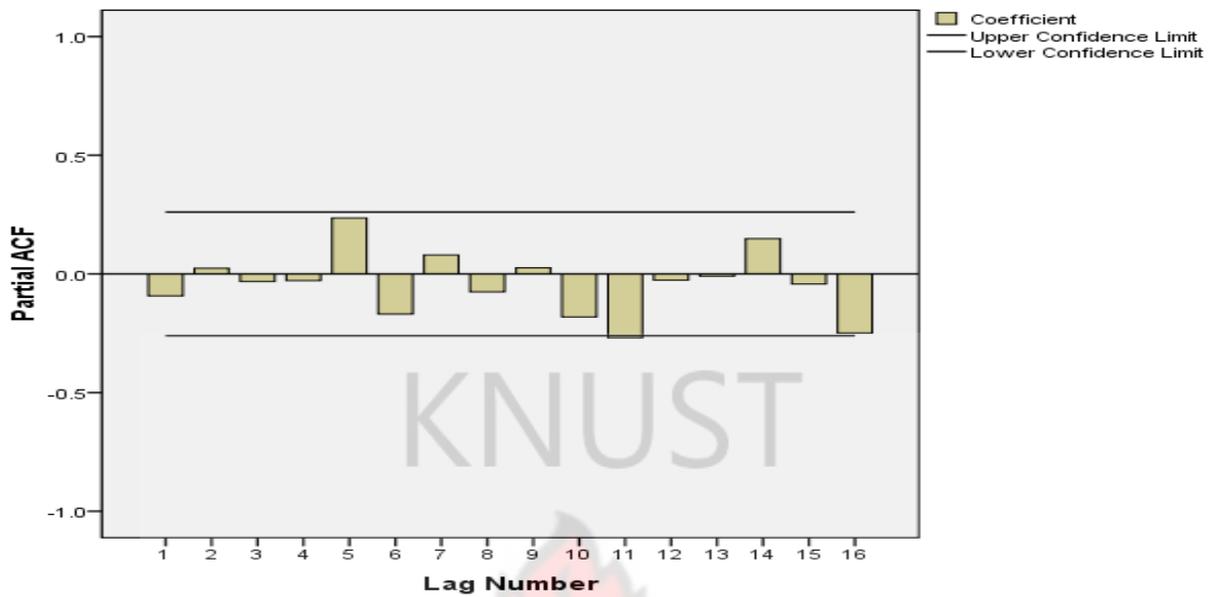


Figure 3.7 Plot of residual partial autocorrelation function

### 3.6 FORECASTING

Once the model is considered acceptable and adequate, the final step of the ARIMA modeling process is to forecast future time series values. There are however two kinds of forecasts: sample period forecasts and post-sample period forecasts. The former are used to develop confidence in the model and the latter to generate genuine forecasts for use in planning and other purposes. The ARIMA model developed can be used to yield the two kinds of forecasts.

#### 3.6.1 UPPER AND LOWER CONFIDENCE LIMIT

Confidence limits for the forecasts are interval estimates for the forecast [36, 37]. Interval estimates are often desirable because the estimate of the forecast varies from sample to sample. Instead of a single estimate for the forecast, a confidence interval generates a lower and upper

limit for the forecast. The interval estimate gives an indication of how much uncertainty there is in our estimate of the true forecast. The narrower the interval, the more precise is our estimate.

Confidence limits are expressed in terms of a confidence coefficient. Although the choice of confidence coefficient is somewhat arbitrary, in practice 90%, 95%, and 99% intervals are often used, with 95% being the most commonly used.

The confidence coefficient is simply the proportion of samples of a given size that may be expected to contain the true forecast. That is, for a 95% confidence interval, if many samples are collected and the confidence interval computed, in the long run about 95% of these intervals would contain the true forecast. It is given by

$$CI = F_t \pm z\sqrt{MSE} \quad (46)$$

Where  $CI$  is the confidence interval,  $F_t$  the forecast of demand at time  $t$ , the value of  $z$  to be used depends on the level of confidence interval required. The  $z$  values are standard and are reported in Table 3.4 [38] and  $MSE$  is the mean square error.

Table 3.4 Confidence limits and their z values

Probability	Z
0.5	0.674
0.68	1.00
0.75	1.150
0.80	1.282
0.85	1.440
0.90	1.645
0.95	1.960

0.98	2.326
0.99	2.576

Suppose we have made a forecast for next month of 195, and we have a MSE of 84.46. If we wanted to create a 75% prediction interval, we would have:

$$195 \pm 1.150\sqrt{84.46} = 195 \pm 1.15 * 9.19 = 195 \pm 10.57 = [185.81, 205.57].$$

A 90% prediction interval would be:

$$195 \pm 1.645 * 9.19 = [179.88, 210.12].$$

As these two examples show, increasing the amount of confidence you have in a forecast results in a wider range of possible values.

### 3.6.2 SAMPLE PERIOD FORECASTS

The sample period forecasts are obtained by simply plugging the actual values of the observed variables in the estimated model equation (45) thus the explanatory variables here are the lagged values of  $Z_t$  and the estimated lagged errors. The obtained values for  $Z_t$  together with the actual values are reported in appendix B Table B.7 and also in Figure 3.8

To judge the forecasting ability of the fitted ARIMA model, important measures of the sample period forecasts accuracy were computed. From [27] in general a mean absolute percentile error (MAPE) of 10% is considered very good, a MAPE in the range 20% - 30% or even higher is quite common. In our study the Mean Absolute Error (MAE) from equation (41) for the number of subscribers turns out to be 8405 and the MAPE from equation (42) is of 9.146. These values are indication that the forecasting inaccuracy is low.

### **3.6.3 POST SAMPLE FORECASTS**

The principal objective of developing an ARIMA model for a variable is to generate post sample period forecasts for that variable. In this work this was achieved using our estimated equation (45) which predicted the number of subscribers that will hook onto our telecommunication company under study.

The forecasts for the number of subscribers for January through December 2011 are given in Figure 3.8

### **3.7 SIMULATED RESULTS DISCUSSION**

The equation that implements our model is equation (45) which solves the proposed model. The parameters used in the simulation are reported in Table 3.3.

From figure 3.8 which is the simulated output, it can be seen that the demand for telecommunication service although maintains an upward trend, it somehow oscillates resulting in rise and fall in demand. The simulated output from figure 3.8 shows a linear trend from index 1 with shocks occurring from the eleventh (11<sup>th</sup>.) month of 2006 to about the twentieth (20<sup>th</sup>.) month of 2007. The almost linear trend is again maintained, then experiences shocks from the twenty eight (28<sup>th</sup>.) to the thirty fourth (34<sup>th</sup>.) month. The cycle is maintained until it experiences drops sharply from the forty third (43<sup>rd</sup>.) to the fiftieth month and then again in the sixtieth (60<sup>th</sup>.) month which is December 2010 from about 260000 subscribers to less than 2240000 sixty first month which is also January 2011. The subscriber base however recovers from the fall in the subsequent month and maintains a study growth.

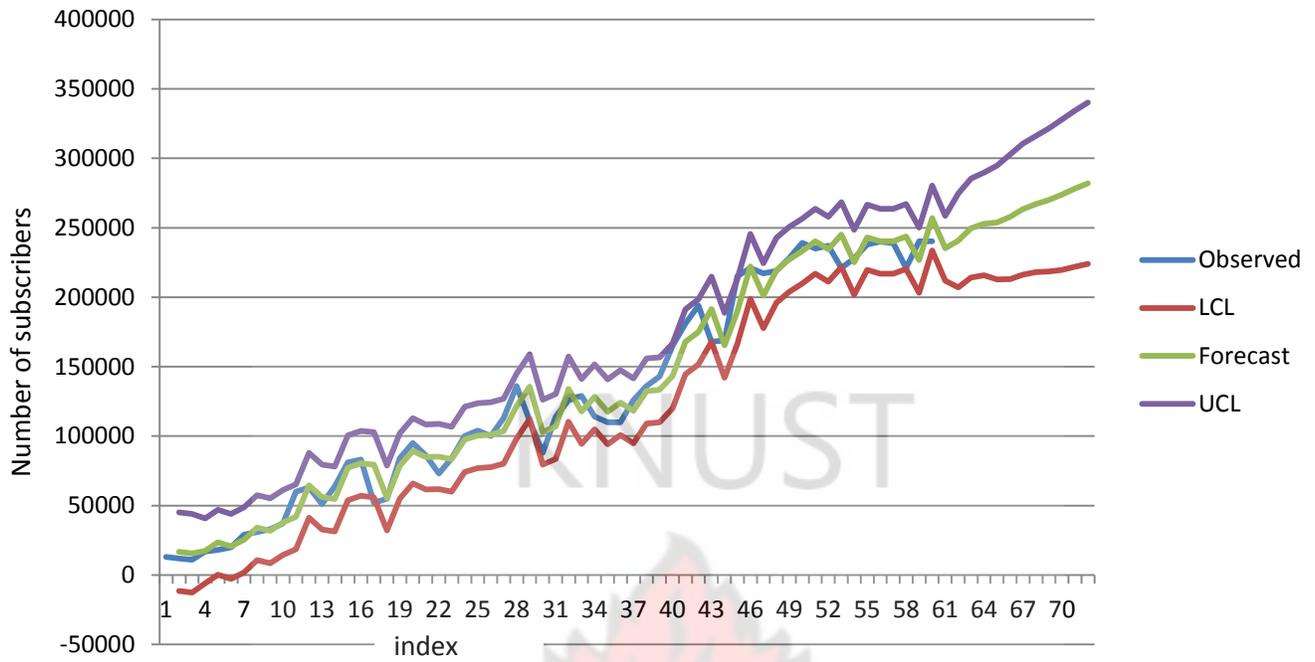


Figure 3.8 Plot of the observed and forecast values with their confidence limits.



## CHAPTER 4

### CONCLUSIONS AND RECOMMENDATIONS

#### 4.1 CONCLUSION

In this work we have been able to use the Box-Jenkins methodology to deduce ARIMA (2, 1, 0) and a corresponding equation of  $Z_t = Z_{t-1} + .042(Z_{t-1} - Z_{t-2}) - .564(Z_{t-2} - Z_{t-3}) + 3835.022$

which can be used to solve a demand problem in the telecommunication sector. The main aim of the researchers was to come out with a model that will accurately forecast future rise in the subscriber base of a telecommunications company so as to make them adequately prepared for such rise in order to put in the necessary infrastructure to meet the demands.

From our simulated output in Figure 3.8 it can be reported in Table 4.1 which shows the sample period forecasts. It can be seen clearly from table 4.1 that the sample period forecast shows a high level of confidence since majority of the forecasts, produce residues that are less than the mean average error which is 8405.

ARIMA models offer a good technique for predicting the magnitude of any variable. Its strength lies in the fact that the method is suitable for any time series with any pattern of change and it does not require the forecaster to choose a priori value of any parameter. Its limitations include its requirement of a long series of data. It is often called a “Black Box” model. Like any other method, ARIMA model does not guarantee perfect forecasts. Nevertheless, it can be successfully used for forecasting long time series data.

Table 4.1 sample period forecasts.

<b>Month_And_Year</b>	<b>Observed</b>	<b>Forecast</b>	<b>Difference (absolute)</b>
Jan-06	13005		n/a
Feb-06	12000	16840	4840
Mar-06	11000	15704	4704
Apr-06	17000	17358	358
May-06	18074	23652	5578
Jun-06	20054	20571	517
Jul-06	29005	25366	3639
Aug-06	31016	34103	3087
Sep-06	33036	31890	1146
Oct-06	37082	37822	740
Nov-06	60045	41949	18096
Dec-06	63097	64574	1477
Jan-07	51091	56117	5026
Feb-07	64023	54694	9329
Mar-07	81091	77174	3917
Apr-07	83099	80361	2738
May-07	52083	79397	27314
Jun-07	55050	55467	417
Jul-07	84068	78492	5576
Aug-07	95002	89462	5540
Sep-07	86037	84944	1093
Oct-07	73057	85327	12270
Nov-07	84087	83392	695
Dec-07	100081	97706	2375
Jan-08	104029	100377	3652
Feb-08	100060	101015	955
Mar-08	113001	103500	9501
Apr-08	136097	121622	14475
May-08	111066	135618	24552
Jun-08	88083	102818	14735
Jul-08	114025	107049	6976
Aug-08	126036	133916	7880
Sep-08	129051	117758	11293

Oct-08	114008	128243	14235
Nov-08	110032	117503	7471
Dec-08	110065	124176	14111
Jan-09	126009	118141	7868
Feb-09	136074	132502	3572
Mar-09	143004	133348	9656
Apr-09	165057	143459	21598
May-09	181005	167922	13083
Jun-09	194000	175086	18914
Jul-09	168003	191397	23394
Aug-09	169066	165407	3659
Sep-09	215054	189598	25456
Oct-09	221010	222243	1233
Nov-09	217056	201175	15881
Dec-09	219054	219365	311
Jan-10	228072	227201	871
Feb-10	239032	233163	5869
Mar-10	235096	240248	5152
Apr-10	237021	234585	2436
May-10	221058	245155	24097
Jun-10	228002	225128	2874
Jul-10	238001	243129	5128
Aug-10	240054	240346	292
Sep-10	239000	240339	1339
Oct-10	221004	243632	22628
Nov-10	240186	226667	13519
Dec-10	240200	256978	16778

Also none of the simulated data or forecast fell outside of the confidence limits which is also an indication that our developed model is a near perfect one. Our model which is ARIMA (2, 1, 0) resulted in a mean absolute percentile error (MAPE) of 9.146% which from [27] is considered very good.

Over the past years there have been enormous agitations from consumers in respect of the poor quality of service provided by the various telecommunications operators in the country which prompted the National Communications Authority to impose a temporal ban on the sale of new subscriber lines and the imposition of fines on many of the telecommunication operators. This problem does not hold sway in only one direction, but one of the obvious reasons is the reluctance of these operators to invest huge capital in terms of infrastructure without any “foreseeable” corresponding demand. It is thus imperative that this study was undertaken to adequately forecast such future demand.

Based on our model we can predict the number of subscribers for twelve (12) months ahead for the period spanning January 2011 to December 2011 which is reported in table 4.2

Table 4.2 post sample period forecast

Month and year	Forecasts
January 2011	235222
February 2011	240836
March 2011	249715
April 2011	252761
May 2011	253720
June 2011	257877
July 2011	263347
August 2011	267070
September 2011	269979
October 2011	273838
November 2011	278196

December 2011	282039
---------------	--------

Further table 4.2 also indicates that for the period the number of subscribers will be about three million one hundred and twenty four thousand six hundred (3124600). This will give operators adequate information on the subscriber base to make informed decisions during infrastructural planning.

This can be used as a tool for forecasting future subscriber base taking into consideration the mean absolute error MAE of 8405 which indicates that for every forecasted value there is a probability of a plus or minus 8405.

The larger the number of period ahead the less efficient the model is. However, the reliability of our forecast is assured when the number of period ahead is not too large since past observations have minor effects on the forecast.

The validity of the forecasted values can be checked when the data for the lead periods become available. The model can be used by telecommunication operators for forecasting future subscriber base in Ghana. However it should be updated from time to time with the incorporation of current data.

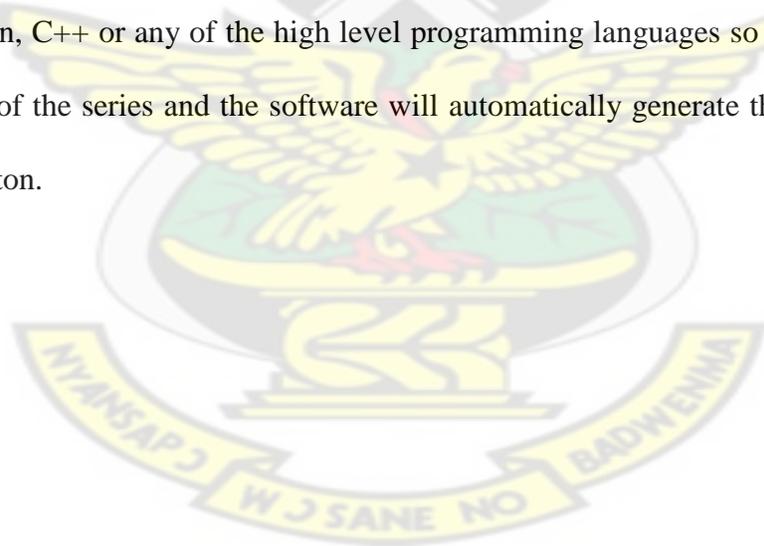
#### **4.2 RECOMMENDATIONS FOR FUTURE WORKS**

Since the main aim of the work was done with the telecommunication sector in mind, it is recommended that telecommunications operators in the country will adopt the proposed ARIMA (2,1,0) model to help them in planning since the model generate near perfect forecast.

Developing ARIMA models require a long series of data. The longer the series the more accurate the forecasts are. Box and Jenkins [3] recommended the length of the series to be at least hundred (100). In our work the length used was however sixty (60) since no company was willing to release such data. It is hoped that in future operators release such data so as to improve on the adequacy of the mode.

The model developed in this work was done with only one company in mind. We recommend that in future data from different operators be combined so as to make them general for all telecommunication operators in the country. We would also develop ARIMA models for institutions and organizations like the Volta River Authority and the Bui Dam Authority who rely heavily on forecast data to predict the amount of water in their reservoirs.

We would also recommend designing a software program with a graphical user interface using MATLAB, Python, C++ or any of the high level programming languages so that users will only input the figures of the series and the software will automatically generate the forecast with the click of just a button.



## REFERENCES

- [1] <http://www.answers.com/topic/mathematical-model#ixzz1IxgZI7Cc> accessed 5<sup>th</sup>. July, 2011
- [2] <http://www.answers.com/topic/mathematical-model#ixzz1IxgnvhnV> accessed 5<sup>th</sup>. July, 2011
- [3] Box, G. E. P. and Jenkins, G. M (1970). Time Series Analysis, Forecasting and Control, Holden Day, San Francisco.
- [4] Pandit, S.M. and Wu, 1983. Time series analysis with applications. John Wiley and sons. New York
- [5] Cheikh Tidiane Gadio, Institutional Reforms of Telecommunications in Senegal, Mali and Ghana: the interplay of structural adjustment and international policy diffusion.
- [6] Edward K Salia, Telecommunications in Ghana. Africa communications (November/December 1993), page 20-21.
- [7] Gyimah, J. k. Paper on Development of Telecommunications and Broadcasting by the Year 2000, Policy Guidelines. Presented at Seminar on Transport and Communications Policy July 6-10, 1992, Accra.
- [8] The Republic of Ghana Telecommunications Master Plan: 1987-2006 (Accra: GPTC, October 1988)
- [9] [www.nca.org.gh](http://www.nca.org.gh)
- [10] Gilder, G. (1991). Into the Telecommunications. Harvard Business Review. March-April 1988, pp.70

- [11] Lyons, M.H., M Gell and P Cochrane (1993a). Telecommunications in the third Millennium, CBTA Conf. Page 141-146.
- [12] Lyons, M.H., K.O. Jensen and I Hawker (1993b). Traffic scenarios for the 21<sup>st</sup> century
- [13] Chatfield, C (1996). The Analysis of Time Series, 5th ed., Chapman & Hall, New York, NY
- [14] Mushin, A .A. 1995. Time Series Models for Treatment of Surface Waters: A Case Study of the Tigris River in Baghdad.
- [15] Mandal B. N. (2003). Forecasting Sugarcane Production In India With ARIMA Model
- [16] Montgomery, D.C., Johnson, L.A. and Gardiner, J.S.(1990) Forecasting and Time Series Analysis, 2nd edition. New York: McGraw-Hill.[1.5]
- [17] [53] A. A. Sveshnikov. Problems in Probability Theory, Mathematical Statistics, and Theory of Random Functions. Dover, New York, 1968.
- [18] Pankratz, A. (1983). Forecasting with univariate Box – Jenkins models: concepts and cases, John Wiley, New York.
- [19] Dobre I. and Alexandru A.A (2008). Modeling Unemployment Rate Using Box-Jenkins Procedure.
- [20] Parzen, E.(1982) ARARMA models for time series analysis and forecasting. J. Forecasting, 1, 67-82. [4.3, 6.4]
- [21] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994). Time series analysis: Forecasting and control, Pearson Education, Delhi.

- [22] Montgomery, A.L., Zarnowitz, V., Tsay, R.S. and Tiao, G.C.(1998) Forecasting the U.S. unemployment rate. *J. Amer. Statist. Assoc.*, 93,478-493. [2.1, 3.4]
- [23] Akaike, Hirotugu (1974). "A new look at the statistical model identification". *IEEE Transactions on Automatic Control* 19 (6): 716–723
- [24] Schwarz, G., 1978. "Estimating the dimension of a model". *Annals of Statistics* 6(2):461-464.
- [25] G. M. Ljung; G. E. P. Box (1978). "On a Measure of a Lack of Fit in Time Series Models". *Biometrika* **65**: 297–303
- [26] Peter Brockwell; Richard Davis (2002). *Introduction to Time Series and Forecasting* (2nd. ed.). Springer. p.36
- [27] Box, G. E. P. and Pierce, D. A. (1970), Distribution of residual correlations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65, 1509–1526.
- [28] Wikipedia
- [29] <http://math.hws.edu/javamath/ryan/ChiSquare.html> accessed 15<sup>th</sup>. July, 2011
- [30] <http://www2.lv.psu.edu/jxm57/irp/chisquar.html> accessed 15<sup>th</sup>. July, 2011
- [31] <http://www.uoguelph.ca/htm/MJResearch/ResearchProcess/StatisticalSignificance.htm>
- [32] McCullough, B.D.(1996) Consistent forecast intervals when the forecast-period exogenous variables are stochastic. *J. Forecasting*, 15, 293-304.[7.5]
- [33] Makridakis, S., Wheelwright, S.C. and Hyndman, R.J.(1998) *Forecasting: Methods and Applications*, 3rd edition. New York: Wiley.[1.5]

[34] On a Measure of a Lack of Fit in Time Series Models", G. M. Ljung and G. E. P. Box, Biometrika, 1978, 65, pp. 297-303

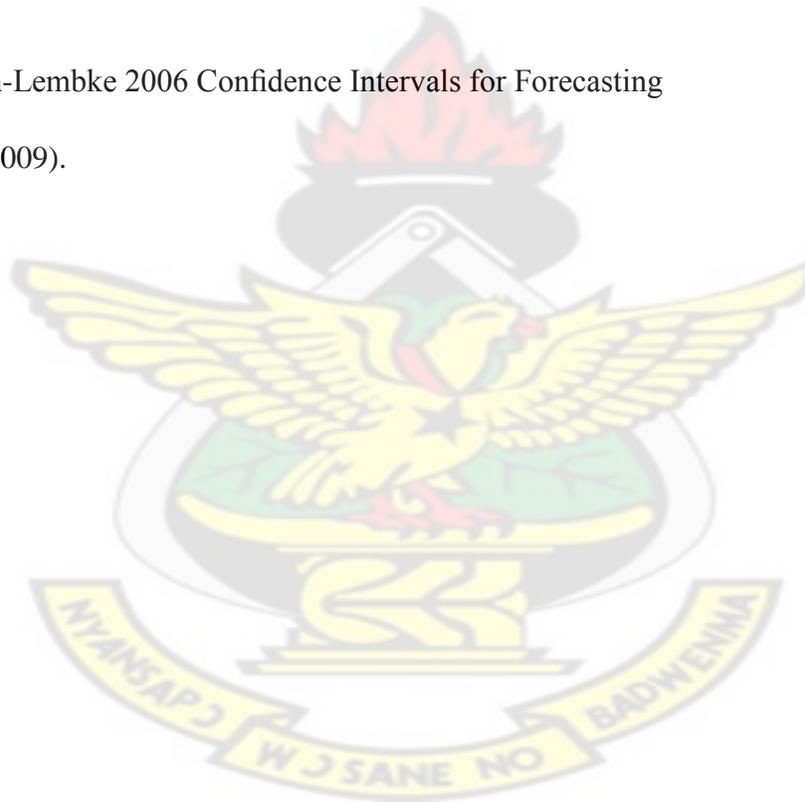
[35] Introduction to Time Series and Forecasting", 2nd. Ed., Peter Brockwell and Richard Davis, Springer, 2002, p. 36.

[36] Makridakis, S. and Hibon, M.(1997) ARMA models and the Box-Jenkins methodology. J. Forecasting, 16, 147-163.[4.2]

[37] Kendall, M.G. and Ord, J.K.(1990) Time Series, 3rd edn. Sevenoaks, U.K Arnold. [1.5, 4.3, 7.4, 7.5]

[38] Ron Tibben-Lembke 2006 Confidence Intervals for Forecasting

[39] Chvalina (2009).



## APPENDIX A

### DATA

Table A.1 Monthly number of subscribers

Month	2006	2007	2008	2009	2010
January	13005	51091	104029	126009	228072
February	12000	64023	100060	136074	239032
March	11000	81091	113001	143004	235096
April	17000	83099	136097	165057	237021
May	18074	52083	111066	181005	221058
June	20054	55050	88083	194000	228002
July	29005	84068	114025	168003	238001
August	31016	95002	126036	169066	240054
September	33036	86037	129051	215054	239000
October	37082	73057	114008	221010	221004
November	70000	84087	110032	217056	240186
December	63097	100081	110065	219054	240200

## APPENDIX B TABLES

Table B.1 Autocorrelation of the series

Lag	Autocorrelation	Std. Error <sup>a</sup>	Box-Ljung		
			Statistic	Value	df
1	.944	.126	56.217	1	.000
2	.885	.125	106.415	2	.000
3	.848	.124	153.308	3	.000
4	.809	.123	196.806	4	.000
5	.756	.122	235.443	5	.000
6	.693	.120	268.493	6	.000
7	.644	.119	297.622	7	.000
8	.600	.118	323.381	8	.000
9	.544	.117	344.972	9	.000
10	.480	.116	362.146	10	.000
11	.425	.115	375.886	11	.000
12	.385	.114	387.356	12	.000
13	.339	.112	396.465	13	.000
14	.289	.111	403.214	14	.000
15	.235	.110	407.779	15	.000
16	.191	.109	410.873	16	.000

Table B.2 partial autocorrelation of the series

Lag	Partial Autocorrelation	Std. Error
1	.944	.129
2	-.064	.129
3	.178	.129
4	-.051	.129
5	-.121	.129
6	-.122	.129
7	.069	.129
8	-.029	.129
9	-.085	.129
10	-.084	.129
11	.002	.129
12	.058	.129
13	-.040	.129
14	-.016	.129
15	-.106	.129
16	.019	.129

Table B.3 autocorrelation of the series after differencing

Lag	Autocorrelation	Std. Error <sup>a</sup>	Box-Ljung		Sig. <sup>b</sup>
			Statistic	Value	
1	.026	.127	.042	1	.837
2	-.566	.126	20.303	2	.000
3	-.156	.125	21.877	3	.000
4	.367	.124	30.710	4	.000
5	.253	.122	34.964	5	.000
6	-.328	.121	42.257	6	.000
7	-.111	.120	43.115	7	.000
8	.135	.119	44.403	8	.000
9	.209	.118	47.560	9	.000
10	-.142	.117	49.042	10	.000
11	-.359	.115	58.715	11	.000
12	.041	.114	58.846	12	.000
13	.226	.113	62.846	13	.000
14	.170	.112	65.167	14	.000
15	-.290	.111	72.064	15	.000
16	-.213	.109	75.874	16	.000

Table B.4 partial autocorrelation functions after differencing

Lag	Partial Autocorrelation	Std. Error
1	.026	.130
2	-.568	.130
3	-.175	.130
4	.069	.130
5	.137	.130
6	-.167	.130
7	.190	.130
8	-.116	.130
9	.194	.130
10	-.162	.130
11	-.221	.130
12	-.186	.130
13	-.102	.130
14	.103	.130
15	-.123	.130
16	-.013	.130

Table B.5 residual autocorrelation function

Lag	Autocorrelation	Std. Error <sup>a</sup>	Box-Ljung	df	Sig. <sup>b</sup>
			Statistic		
			Value		
1	-.079	.127	.389	1	.533
2	.028	.126	.437	2	.804
3	-.024	.125	.473	3	.925
4	-.043	.124	.591	4	.964
5	.269	.122	5.406	5	.368
6	-.230	.121	9.004	6	.173
7	.184	.120	11.357	7	.124
8	-.109	.119	12.203	8	.142
9	.009	.118	12.208	9	.202
10	-.109	.117	13.081	10	.219
11	-.339	.115	21.674	11	.027
12	.120	.114	22.771	12	.030
13	-.121	.113	23.916	13	.032
14	.179	.112	26.487	14	.022
15	-.133	.111	27.938	15	.022
16	-.267	.109	33.907	16	.006

Table B.6 residual partial autocorrelation function

Lag	Partial Autocorrelation	Std. Error
1	-.079	.130
2	.021	.130
3	-.020	.130
4	-.047	.130
5	.265	.130
6	-.207	.130
7	.171	.130
8	-.101	.130
9	.019	.130
10	-.218	.130
11	-.253	.130
12	-.038	.130
13	-.029	.130
14	.127	.130
15	-.048	.130
16	-.233	.130

Table B.7 **observed and forecast values for subscribers**

Month_And_Year	Observed	LCL	Forecast	UCL
Jan-2006	13005			
Feb-2006	12000	-11487	16840	45167
Mar-2006	11000	-12613	15704	44020
Apr-2006	17000	-6032	17358	40748
May-2006	18074	262	23652	47042
Jun-2006	20054	-2819	20571	43961
Jul-2006	29005	1977	25366	48756
Aug-2006	31016	10713	34103	57493
Sep-2006	33036	8500	31890	55280
Oct-2006	37082	14432	37822	61212
Nov-2006	60045	18559	41949	65339
Dec-2006	63097	41184	64574	87964
Jan-2007	51091	32727	56117	79507
Feb-2007	64023	31304	54694	78084
Mar-2007	81091	53784	77174	100564
Apr-2007	83099	56971	80361	103751
May-2007	52083	56007	79397	102787
Jun-2007	55050	32077	55467	78857
Jul-2007	84068	55102	78492	101882
Aug-2007	95002	66072	89462	112852
Sep-2007	86037	61554	84944	108334
Oct-2007	73057	61937	85327	108717

Nov-2007	84087	60002	83392	106782
Dec-2007	100081	74316	97706	121096
Jan-2008	104029	76987	100377	123767
Feb-2008	100060	77625	101015	124405
Mar-2008	113001	80110	103500	126890
Apr-2008	136097	98232	121622	145012
May-2008	111066	112228	135618	159008
Jun-2008	88083	79428	102818	126208
Jul-2008	114025	83659	107049	130439
Aug-2008	126036	110526	133916	157306
Sep-2008	129051	94368	117758	141148
Oct-2008	114008	104853	128243	151633
Nov-2008	110032	94113	117503	140893
Dec-2008	110065	100786	124176	147566
Jan-2009	126009	94751	118141	141531
Feb-2009	136074	109112	132502	155892
Mar-2009	143004	109958	133348	156738
Apr-2009	165057	120069	143459	166849
May-2009	181005	144532	167922	191312
Jun-2009	194000	151696	175086	198476
Jul-2009	168003	168007	191397	214787
Aug-2009	169066	142017	165407	188797
Sep-2009	215054	166208	189598	212988

Oct-2009	221010	198853	222243	245633
Nov-2009	217056	177785	201175	224565
Dec-2009	219054	195975	219365	242755
Jan-2010	228072	203811	227201	250591
Feb-2010	239032	209773	233163	256553
Mar-2010	235096	216858	240248	263638
Apr-2010	237021	211195	234585	257975
May-2010	221058	221765	245155	268545
Jun-2010	228002	201738	225128	248518
Jul-2010	238001	219739	243129	266518
Aug-2010	240054	216956	240346	263735
Sep-2010	239000	216949	240339	263729
Oct-2010	221004	220242	243632	267022
Nov-2010	240186	203277	226667	250057
Dec-2010	240200	233588	256978	280368
		211832	235222	258612
		207048	240836	274625
		214105	249715	285324
		215740	252761	289782
		212778	253720	294661
		212980	257877	302774
		216222	263347	310472
		218019	267070	316121

		218415	269979	321542
		219731	273838	327944
		222015	278196	334377
		223944	282039	340135

# KNUST

