TIME SERIES MODELLING FOR TOTAL NUMBER OF DEFECTIVE PARTS OF PRINTED CIRCUIT BOARDS IN THE MANUFACTURING INDUSTRY IN GHANA

BY

MAHA YEN SAMUEL

(10442023)

THIS THESIS IS SUBMITTED TO THE UNIVERSITY OF GHANA, LEGON IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE AWARD OF MPHIL STATISTICS DEGREE

JUNE, 2015

DECLARATION

Candidate's Declaration

This is to certify that this thesis is the result of research work undertaken by me, Samuel Maha Yen towards the award of Master of Philosophy in Statistics in the Department of Statistics, University of Ghana. However, all references cited in the text have been duly acknowledged.

…………………………… ……………………….

SAMUEL MAHA YEN DATE

(10442023)

Supervisors' Declaration

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

…………………………… ……………………….

DR. ANANI LOTSI DATE

(Principal Supervisor)

…………………………… ……………………….

DR. F. O. METTLE DATE

(Co- Supervisor)

ABSTRACT

For stochastic time series modelling, an essential property is the underlying statistical model that is assumed to govern the number of defective parts of printed circuit boards in a production process in the manufacturing industries in Ghana. The data was for a period spanning from 2009 to 2014. Considering time series methodology, we specify differences in the data points as a stochastic process assumed to have Markov dependency with respective state transition probabilities matrices following the identified state space (i.e. increase, decrease or remain the same). We observed that the identified states communicate, hence the chains are aperiodic and ergodic signifying the possessing of limiting distributions. We established a methodology for determining wthether the daily number of defective parts increase, decrease or remained the same. A criteria for identifying the state(s) in which production of printed circuit boards is cost effective based on least transition probabilities was also applied. The established methodology is applied to daily number of defective parts during printed circuit boards production in the manufacturing industries in Ghana.

Chapman-Kolmogorov Equations and time series models were used in this study.The results showed that it was cost effective when productions are done in the first and second states since at these states least number of defective parts of printed circuit boards is recorded.The n-step probabilities were also determined by subjecting the transition matrix to powers.Autoregressive and moving average method was also applied. The data was differenced once for stationarity which was confirmed by the Augmented Dickey-Fuller Test. The model that was adjudged the best was the model with least AIC and BIC values. ARMA (1,1) model was adjudged the most ideal model for forecasting in this study since it met all the requirements of an ideal model.

DEDICATION

I dedicate this thesis to God Almighty for HIS support in my life and to my lovely wife: Mrs.Theresah Sampana Bugrepoka for her sacrifice , dedication and understanding.

ACKNOWLEDGEMENT

I am grateful to the Almighty God for the love, mercy, protection, knowledge and the opportunity given me to pursue this programme to the end. Various people contributed immensely to the success of this thesis and are acknowledged for their invaluable contributions. First, my gratitude goes to my Supervisors; Dr. Anani Lotsi and Dr. F. O. Mettle who provided academic guidance, comments and suggestions throughout this thesis. I expressed my profound appreciation to all the lecturers of Department of Statistics, particularly Mr. Abeku Kumi for the central position he played throughout my study period. To my family, I say thank you all for your prayer, support and encouragement throughout my studies and may the Almighty God bless you all.

Finally, I express my sincerest appreciation to the 2013 and 2014 batches of Master of Philosophy students in Statistics, University of Ghana, Legon for their support in diverse ways and my sisters Francisca Ayeefo, Florence Ayeefo and Lizzy Ayeefo.

Thank you and God bless you.

TABLE OF CONTENTS

Content **Page**

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

ACF:	Autocorrelation Function
ADF:	Augmented Dickey- Fuller
AIC:	Akaike Information Criterion
AOI:	Automatic Optical Inspection
AR:	Autoregressive
ARCH:	Autoregressive Conditional Heteroscedastic
ARFIMA:	Autoregressive Fractionally Integrated Moving Average
ARIMA:	Autoregressive Integrated Moving Average
ARMA:	Autoregressive Moving Average
BIC:	Bayesian Information Criterion
CPU:	Central Processing Unit
DEA:	Data Envelopment Analysis
EGARCH:	Exponential Generalized Autoregressive Conditional Heteroscedastic
GARCH:	Generalized Autoregressive Conditional Heteroscedastic
GDP:	Gross Domestic Product
GRA:	Gray Relation Analysis
IGARCH:	ntegrated Generalized Autoregressive Conditional Heteroscedastic
LAN:	Local Area Network
MA:	Moving Average
MSE:	Mean Square Error
MTN:	Mobile Telecommunication Network
NIC:	Network Interface Card
PACF:	Partial Autocorrelation Function

PARCH:	Power Autoregressive Conditional Heteroscedastic
PCBs:	Printed Circuit Boards
ppm:	Parts Per Million
SARIMA:	Seasonal Autoregressive Integrated Moving Average
SMT:	Surface Mount Technology
TGARCH:	Threshold Generalized Autoregressive Conditional Heteroscedastic

CHAPTER ONE

INTRODUCTION

1.0 Stochastic Modelling and Time Series

A stochastic process is the mathematical structure describing the probability structure of a time series. Matrix with the property of the row sums equal to 1 by the law of total probabilities is called stochastic.

Stochastic modelling on the other hand is a mathematical analysis of interactive processes within a system that involve randomness. The system could range from the telecommunication through to manufacturing. However, this research will particularly focus on the manufacturing industries. Manufacturing systems can be described as a network of queues and servers.

A time series $\{X_t; t \in T\}$ can be defined as an ordered sequence of random variables over time, where T denotes an index time points set (Akgun, 2003). It is regarded as the record of some activity, with observations taken at equally spaced time interval.

1.1 Background of the study

The manufacturing industries have become vital areas for sustaining the economies of developing countries. Due to this, many developing countries have embarked on a number of manufacturing industries in order to revamp their economies.

Ghana is no exception to this new wave of development. The country has undergone a process of manufacturing sector restructuring and transformation for the past decades in order to achieve emerging manufacturing sector status. The restructuring of the sector resulted in the creation of industries in Ghana.

Manufacturing constitutes about 9% of Ghana's Gross Domestic Product (GDP) and provides employment for over 250,000 people (Business News, 2009). There are around 25,000 registered firms, though more than 80% of them are small size enterprises and around 55% of them are located within the Greater Accra/Tema Region.

Major industries include mining, light manufacturing, aluminium smelting, food processing, cement and small commercial ship building. Other industries include food and beverages production, textiles, chemicals and pharmaceuticals, and the processing of metals and wood products; a relatively small glass- making and recycling industries have also developed (Nexus Strategic Partnership Limited, 2014).

The sector is underdeveloped and is characterised by a narrow industrial base dominated by agro-industries. Subsidiaries of multinational companies have a strong presence in the country; including Unilever, Coca Cola, Toyota and Accra Brewery, but there are also many medium sized local companies. In the 1980s, manufacturing share of GDP was more than 10% but structural adjustment programmes and failed state-led industrialisation policies have seen the sector's share decline (Nexus Strategic Partnership Limited 2014).

Electrical, electronic and telecommunication industries such as Compu Ghana, MTN, VODAFONE, AIRTEL, GLO and TIGO have also emerged. The study would therefore concentrate on electronics printed circuit boards (PCBs).

A printed circuit board (PCB) is a strong, electrically non-conductive platform on which electronic and hardware components are mounted. PCBs could be single sided, double sided or multilayers and are the building blocks of most electronic systems such as computer systems, cell phones and electrical gadgets. A printed circuit board is said to be defective if a unit fails to meet acceptance criteria due to one or more defects. On the other hand, defect is simply the failure of a PCB to meet one part of an acceptance criterion.

The electronics industry is a major part of today's manufacturing sector. Many firms are competing for their share of the market and managing operations efficiently is critical for maintaining competitiveness. However, operating effectively is becoming more difficult as product variety and complexity increase.

Furthermore, capital equipment expenses have increased due to the high degree of automation and versatility. As a result, profit margins decrease when attention is not paid to operational issues pertaining to total number of defectives and non-defectives PCBs produced. Therefore, the task of continuously improving productivity is a crucial effort. Production lines in electronics manufacturing are rather standard in their design (Gebus, 2000). They consist in paste printing, component placement and soldering which are linear and sequential in nature.

Hence, controlling the boards between each sequence is vital. In the world of electronic manufacturing, the final product has two different states: "Working" (non-defective) or "Not working" (defective). The state "Working" depends on the components used to make PCBs. A defective in any of the components have a significant impact on the performance of a PCB. Based on this fact, the electronic manufacturing world is a world where zero defects is a necessity and cannot therefore be compromised. From a statistical point of view, increasing the amount of components on a PCB would drastically increase the defect rate. If for instance a PCB made out of 1000 components with a yield of 0.999 for each one of them, the yield for the PCB will be 0.999^{1000}=37%. It then means that about two thirds of the production would undergo some rework.

The electronics plant manufactures several major product lines which include PCB. Within the plant, each product line is treated as an independent business unit and manufacturing is done in a dedicated work cell with its own equipment, support staff and technicians. Product lines, however, share shipping, receiving, stockroom staff and area. In addition, they share the printed circuit board (PCB) work cell, which is responsible for manufacturing boards needed for all products. In stochastic process, the Markov chain specifies a system of transitions of an entity from one state to another. For the identification of the transition as a random process, the Markov dependency system stresses “memoryless property” thus, the future state of any process by law depends on its current state but not its past state.

1.1.1 Types of PCBs

Generally, a PCB is a thin plate on which chips and other electronic components are placed. PCBs could be one sided, double sided or multilayer.

PCBs fall under the following categories:

- **Motherboard:** It is the principal board that has connectors for attaching devices. Basically, the motherboard contains the central processing unit (CPU), memory and controllers for the system.
- **Expansion board:** It is any board that plugs into one of the computer’s expansion slots which include controller boards, local area network (LAN) and video adapters.
- **Daughter card:** It is any board that attaches directly to another board.
- **Controller board:** It is a special type of expansion board that contains a controller for a peripheral device such as disk drives and graphics monitor.
- **Network Interface Card (NIC):** It is an expansion board that enables a personal computer to be connected to a local area network (LAN).
- **Video Adapter:** It is an expansion board that contains a controller for graphics.

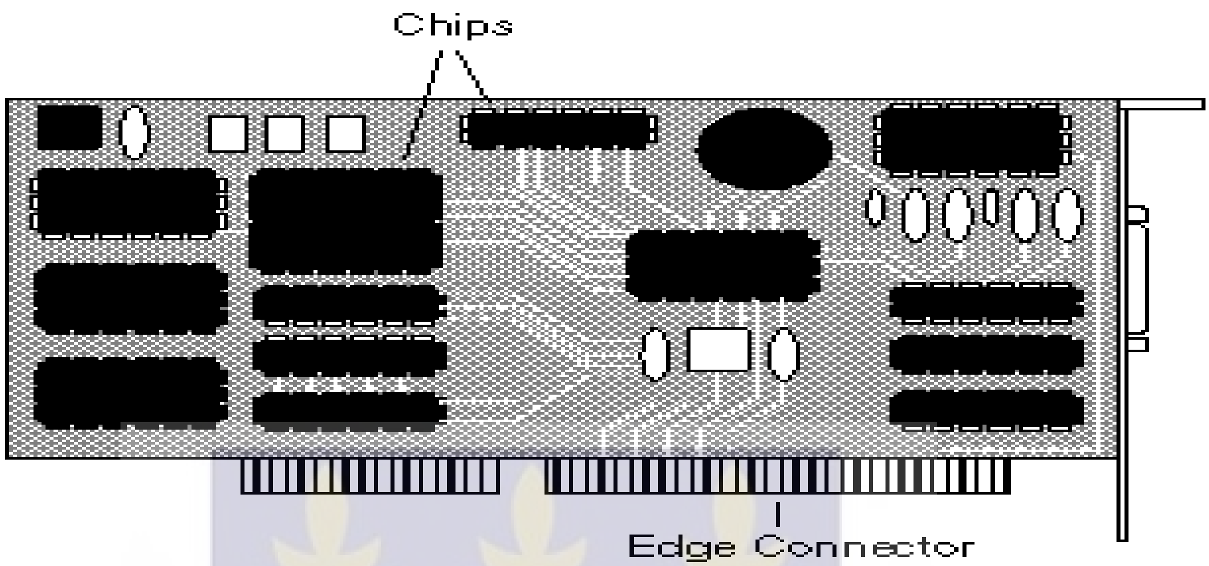

Figure 1: Sample of a PCB

1.1.2 Methods Used in PCBs Production

In the past, Through-Hole assembly was the most common technology used to produce PCBs. Currently, Surface Mount Technology (SMT) is more widely used because it is faster and more precise. With Through-Hole assembly, the parts are inserted through actual holes in the board, while with Surface Mount Technology (SMT) assembly; the parts are placed on a solder adhesive at pre-specified locations on the board.

The manufacturing process for producing PCBs can be generalised into three main steps: Paste application, pick-and-place and then reflow. For paste application, each board has a unique stencil that pushes soldering paste through this stencil onto the bare board where parts will be located. The essence is to place the right amount of solder in the right place on the board, since too much paste could cause the board to be defective. After the paste is applied, a pick-and-place machine that was previously loaded with parts, picks the needed part, orients it correctly and then places it at the specified location on the board and then finally the reflow process follows.

1.1.3 Role of PCBs

The function of a PCB is determined by what components are used, and how these components are connected to each other.

- The PCB provides a stable mechanical platform so that the electrical connections between the components remain intact at all times.
- The PCB is the building block of the system components.

1.1.4 Areas Where PCBs Are Used

Basically, printed circuit boards are used in machines and gadgets such as; Television sets, Computers, Cell phones, Electrical parts and many others.

1.2 Motivation of the research

This research is motivated by the fact that the manufacturing industry contributes significantly to Ghana's GDP. This research therefore focuses on developing a model to predict the number of defectives and non-defectives during PCBs production since the importance of PCBs cannot be underestimated. Again, it is motivated by the need for a tool to assist the producers of electronic products in PCBs in detecting the number of defective parts in their production lines.

Manufacturing systems could be described as network of queues and servers and this could best be explained by stochastic time series linear models on the part of defectives and non-defectives PCBs, since stochastic time series linear models are relatively simple in understanding and implementation.

1.3 Problem statement

The manufacturing industry in Ghana is faced with a fate of fluctuations in interest rates; high rate of inflation and instability of the cedi, couple with the energy crisis which has made it difficult to predict the future of the sector. The researcher noticed that little or no work has been carried out on the part of stochastic time series modelling on the part of number of defective parts in printed circuit boards in various manufacturing industries. Again, how to detect the number of defectives and non-defectives and how much time is spent in detecting defectives from a batch of production of PCBs in the electronic industry is very important.

This study therefore wishes to develope a methodology that could be used as a model to predict the number of defectives in the production of printed circuit boards in the electronic industry.

1.4 Objectives of the study

The main objective of the study is to generate a stochastic-time series model (linear or non-linear) for detecting defectives and non-defectives in the production of Printed Circuit Boards in the electronic manufacturing industry.

Specifically, the study seeks to:

1. Generate a suitable model by the use of Chapman-Kolmogorov Equations method to determine whether the total number of defective parts of PCBs increase, decrease or remain the same during the production process.

2. Identify a stochastic-time series model for the prediction of the total number of defective parts in PCBs production and hence check the adequacy and reliability of the model.

1.5 Significance of the study

The results and findings from this study would be significant in the following forms: First, the results from the study will serve as a guide to the public about which stochastic-time series model to consider when it comes to detecting the number of defectives and non-defectives during the production of PCBs. The study would also explain the ideas of stochastic modelling especially its uses in manufacturing processes. Hence, effective use of stochastic modelling could give businesses a competitive advantage in the market place.

Second, the results from the study would be of great help to the academia and researchers by contributing to existing knowledge and literature.

Third, the results and findings will serve as a basis for future research work in a related area or field.

1.6 Scope and Methodology

The study was carried out in Ghana. Secondary data consisting of production day, part number, average number of defectives and average proportion number of defective parts of printed circuit board from January 2009 to 2014 December were collected from the various electronic companies in Ghana and used in this study.

A year-on-year defective parts in printed circuit boards (PCBs) in the electronic industry is where a unit fails to meet acceptance criteria due to one or more defects. A defect on its part is simply the failure of a printed circuit board to meet one part of an acceptance criterion.

The data was first simulated and smoothed by the linear and non-linear smoothing methods and then analysed.

1.7 Organization of the study

The study is divided into five chapters. Chapter one gives a brief introduction to the study, providing the background information of the study, statement of the problem, objectives of the study, the significance of the study, motivation of the research as well as the scope and brief methodology of the study. Chapter two presents a review of related literature that pertains to the study whilst chapter three presents the detailed methodology used for the study. Chapter four on the other hand presents the data analysis, presentation and discussion of the results. Finally, chapter five gives the summary, conclusions and recommendations of the study.

CHAPHER TWO

LITERATURE REVIEW

2.0 Introduction

This chapter presents the literature related to the manufacturing of PCBs and the number of defective parts. The research done in this area is divided into five main sections: Process modelling, Board design, testing and detecting the number of defective parts and the relevant concepts associated with moving average smoothing techniques.

Process modelling develops a model usually statistically based to mathematically predict the manufacturing yield of a process. Board design on its part, looks at the design parameters and special features of the PCBs and also predicts the manufacturing yield.

PCBs testing addresses the close relationship that exists between yield estimation and the percentage of faults and finally, detecting the number of defectives which are not in conformance with the purpose of production.

The five sections of this chapter summarize the most relevant literature within each of these sections.

2.1 Process Modelling

In the late 1970's and early 1980's, the development of more complex integrated circuits chips by the electronics industry made it clear that a method for quantifying the low productions that were being obtained needed to be developed. In particular, defect level and fault coverage became the key elements researchers examined to determine the production of a particular process.

Defect level is the percentage of chips delivered with undetected faults, and fault coverage is the percentage of fault that will be detected during board test Aas (1988).

Wadsack (1978) quantified the relationship between the two as linear. Wadsack (1981) redefined his model and introduced a two-parameter gamma distribution to model the differences in defect density across a chip. Later that year, Williams et al (1981) presented a paper in which a non- linear relation was derived for defect level versus production yield and fault coverage.

They assumed that a given chip had a given number of faults and that faults were independent of whether or not other faults had occurred. Based on these assumptions, they first derived an expression assuming uniformly distributed faults and using the binomial distribution as follows:

$n =$ Total number of process type i

$p =$ Probability of fault of type i occurring

$DL =$ Defect level

$Y =$ Yield

$T =$ Fault coverage

Probability of Board number faults $= (1-P)^n$

Probability of k faults $= \binom{n}{k}(1-P)^{n-k} P^k$

Probability of accepting a board with k faults when testing m of the n:

$$\binom{n}{k}(1-P)^{n-k} P^k \left(\frac{\begin{matrix} n-k \\ m \end{matrix}}{\begin{matrix} n \\ m \end{matrix}} \right) = \binom{n-m}{k}(1-P)^{n-k} P^k$$

Probability of accepting a board with one or more faults when testing m of the n :

$$P_a = (1-P)^m - (1-p)^n$$

The Williams and Brown model became extremely popular among researchers. Agrawal (1982) presented a variation of the Williams and Brown model incorporating the average number of faults on a chip and the yield of good chips. This model rather calculates the rejection rate as a function of the fault coverage. Williams (1985) again made a research in which the relationship between defect level and random patterns was used to determine accurate test procedures in a self-testing environment.

Parker (1989) also presented a paper on the way to statistically measure delay-fault coverage and then generalized his model by considering a testing strategy which is determined by the defect level for a given set of faults in Parker (1992). "Corsi (1993) on his part used conditional probability to simplify the assumption made in the original Williams and Brown model regarding equiprobable faults". This model incorporates a generalized weighted fault coverage parameter that accounts for the non-equiprobable faults. Sousa (1996) also extended his model to incorporate the concepts of non-equiprobable faults into defect-level calculations. His model computes the probability of occurrence of a given fault, provided that a critical area for the fault can be defined. This critical area can be determined by analysing typical integrated circuit process line defect statistics. In effect, the above authors concluded that the defect level depended on the critical areas associated with undetected faults and fault densities.

Stapper (1985) also generated a negative binomial distribution to represent the frequency distribution of the number of faults per chip also simulated clustered fault locations on a map using a radial Gaussian probability distribution.

Aas (1989) also developed various simulation experiments to study multiple fault coverage versus single fault coverage. They defined multiple fault coverage as the event where the combined effect of k faults leads to fault detection although each individual fault was not detected. They also defined fault masking as being the event in which one or more of the *k* faults will lead to single fault detection, but the combined effect of the *k* faults will lead to simultaneous fault masking. In conclusion, they found that fault or defective exposure and defect masking are high for single fault coverage and that it is less pronounced for multiple defect coverage.

Little or no information is available on the part of stochastic time series modelling in detecting the number of defective parts of PCBs production. Due to this, the obvious question is: how much defective coverage is enough? Henshaw (1989) used a cost model to answer this question. He developed an equation to calculate the defects in parts per production and fed this information into a cost model. He then concluded that for certain typical conditions, the economic level of defective coverage should be as close to 100% as possible.

2.2 Printed Circuit Boards

As printed circuit boards increased in complexity, production yield of PCBs became an issue and a topic of study for researchers Chevalier (1997). Electronic products found in today's market typically contain one or more PCBs; and therefore, the production of PCBs will directly affect the production of electronic products. Due to the stern competition that exists among some of these companies, much of this area had not been published. Joseph, Watt and Wigglesworth Joseph (1990) on his part applied the Fish-Born Analysis to analyse defective parts of PCBs.

Clark (1995) also developed a quality modelling system for predicting the yield of assembly and test processes in printed circuit boards industries. In order to correctly track the source of the fault, the quality modelling systems looked at three particular data requirements:

- Part number

- Process (or where the board is being assembled) and
- Fault type.

Once faults are identified, quality modelling system divides them into either assembly faults or component faults. Assembly faults are divided into those faults that cause:

- Short circuits,
- Open circuits,
- Non-functional faults and
- Visually unacceptable faults.

Based on these, defective rates were generated for each of the assembly faults.

Additionally, the Joseph (1990) assumed that component faults were related to the type of electronic functions that they perform. This assumption was made after analysing historical data and concluding that the defect rate was partially dependent on the components' complexity and technology.

Again, Joseph (1990) also categorized the boards into three different complexity levels: Simple, Complex and Very complex. This distinction between the three was based mainly on the component density of the board which was computed by taking the ratio of the area required to mount all the components over the total area available for mounting components on the board.

Given that production yield directly depends on the number of faults in the process, Linn (1996) on his part studied the placement errors for Surface Mount Technology. He determined then that the placement process errors depended on three features namely;

- the components used,
- the design nature of the PCBs and
- the placement machine.

He affirmed that defects of the PCBs level are caused by dimensional variations and locations offsets.

2.3 Board Design

The number of defective-free PCBs depends not only on the quality of the components used and the accuracy or capabilities of the process, but also on the design of the board. Li et al., (1993) also concluded that as the design complexity of the board increases, production decreases. Defective parts are more likely to take place in more complex designs because the probability of error increases as the number of components increases and the complexity of the assembly process also increases.

2.4 PCB Testing

Testing a PCB has both advantages and costs associated with it. The benefits of testing include improving quality control and feedback process and on the part of the cost include time and money spent in detecting the number of defective parts.

The quality of the testing process is directly proportional to the percentage of defectives that will be identified by the test. Hroundas (1990), in his work concluded that to remain competitive, the producers of PCBs must look at the quality of electrical testing. He added that to achieve parts per million (ppm) failure levels, the producer must increase not only the control over the processes, but also the quality of the testing tools since faults spectrum keep on changing. He concluded by saying that the high quality of the PCBs is directly dependent on the quality of the electrical test it receives and the quality of the electrical test rather depends on the testing tools' defective coverage.

Millman (1994) explained also rather that the quality level is the ratio of the number of defect free parts that pass the test over the total number of parts that pass the test. On the part of defective coverage, he pointed out that the key element is to precisely match the defect model applied with the defects actually occurring. Li and Hang., (2007) used a Markov Chain model to evaluate quality performance in flexible manufacturing systems and investigated the impact of different production scheduling policies on product quality.

According to Li et al., (2007) a defective part could be routed to three locations: rework, minor repair and component placement, according to the nature and severity of the defects. Parts with severe defects are sent back to the rework process where the whole part is worked on again. On the other hand, parts with minor defects are repaired either through component replacement or by minor repair. At component replacement, one or more defective parts in a job are replaced with new ones and inspected again.

As previously pointed out, little effort has been made to address the impact of production system design on product quality pertaining to the number of defective parts. Among the limited related references, Son and Park., (1987) presented a measure of productivity, quality and flexibility for production systems. Jacobs and Meerkov., (1991) studied the perturbation in the average steady-state production rate by quality inspection machines for an asymptotically reliable two-machine one-buffer line. This type of analysis was extended to a highly reliable multi-machine line by Han (1998).

Bulgak (1992) discussed the trade-offs between productivity and product quality, as well as their impact on optimal buffer designs. Khouja (1995) delineated the trade-off between throughput and quality for a robot whose repeatability deteriorates with speed.

Deliman and Feldman., (1996) on their part studied the imperfect inspection and rework of defective items in serial lines. Stochastic search techniques such as generic algorithms and simulated annealing are used by Viswanadham (1996) to investigate the impact of inspection allocation in manufacturing systems.

Urban (1998) analysed the competing effects of large or small batch sizes and develops a model for the interaction between batch size and quality. Cheng (2000) used quantitative measures to deduce that ∪-shaped lines of PCBs testing produce better quality products. Matanachai and Yano., (2001) also proposed a new line-balancing approach to improve quality by reducing work overload. Inman (2003) in a reviewed of the related literature presented evidence for the impact of production system design on quality. Finkelshtein (2005) also studied on off-line inspection following an unreliable production process by developing a model to minimize the expected inspection and disposition error cost.

Kim (2005) basically introduced an integrated model of a two-machine one-buffer line with inspection and information feedback to study both quality and quantity performance in terms of good job production rate. Li (2007) applied a Markov Chain model to evaluate quality performance in flexible manufacturing systems and investigate the impact of different production scheduling policies on product quality.

Repair and rework processes are important parts of production systems in many manufacturing industries, such as the semi-conductor, electronics, packaging and process industries, (Pendse1994). Again, Narahari (1994) also researched into the transient performance of manufacturing systems using unstable queuing networks with returning semi-finished part and re-entrant lines as examples. Narahari (1996) analysed a re-entrant manufacturing system with inspections, scrap and rework and presented an approximate technique based on mean value analysis to predict steady-state cycle times and throughput rate. Liu (1996) considered a single-stage imperfect production system processing both new and rework items. Moreover, a threshold control policy to study a similar system was introduced by Chern (1999).

2.5 Smoothing of time series data

Basically, in time series analysis, to smooth a data set is to make an approximating function that attempts to capture essential patterns that exist in the given set of data, while leaving out noise. Smoothing techniques is simply an approach to remove fluctuations from a time series with a minimum of preconceptions and assumptions as to what those patterns should be. The aim is to separate the data into a smooth component invariably a fitted curve or trend and a rough component termed as residuals or noise.

Smoothing helps in data analysis due to the fact that it is able to extract more information from the data as long as assumption of smoothing is meaningful. Additionally, it also provides analyses that are robust and flexible.

2.6 Moving Average Smoothing Techniques

Under the concept of statistics, a moving average also termed as rolling average is a type of finite impulse response filter applied to analyse a set of data points by creating a series of different subsets of the full data set. It is a linear smoothing technique commonly used with time series data to smooth out short-term fluctuations and highlight long-term trends or cycles. According to Velleman (1981), the most common smoothing technique is the moving average. The concept behind the use of moving averages for smoothing time series data is that observations which are nearby in time are also likely to be close in value. An average of the points near an observation provides a reasonable estimate of the trend at that observation. The average eliminates some of the randomness in the data, leaving a smooth trend component. An n-period moving average is the average value over the previous *n* time periods. As you move forward in time, the oldest time period is dropped from the analysis. Basically, this could be illustrated as indicated below:

$$Moving\ Average = \sum \frac{values\ in\ the\ previous\ n - periods}{n}$$

2.7 Overview of the Related Literature

The production of printed circuit boards has generated a huge industrial activity over the last four decades. Printed circuit boards are consumed as inputs by three main industrial sectors: computers, telecommunications and consumer electronics representing 72.5% of the total consumption in 1998 (Nakahara, 1999).

Over the years, printed circuit boards' production has transformed from a labour-intensive activity to a highly automated one, characterised by steady innovations at the level of design and manufacturing processes. The increasing complexity of printed circuit boards inevitably leads to higher failure rates in the production process.

The PCB assembly consists in placing (inserting, mounting) a number of electronic components of specified types at specified locations on a raw board. Several hundred components of a few distinct types (resistors, capacitors, transistors, integrated circuits,) are being placed on each board. The PCBs production and assembly process are basically optimized to ensure high quality products.

The quality of a product is determined by the collection of features and characteristics of a product that contribute to its ability to meet given standards. Quality is measured by the degree of conformance to predetermined specifications and standards. The deviations from these standards can lead to poor quality and less reliability. Efforts for quality improvements are aimed at eliminating defective parts of printed circuit boards. Defective parts of PCBs could arise in most circumstances due to incorrect or incomplete designs and poorly executed production and assembly.

The defective part of a PCB is considered as undesirable output as it has to be reprocessed after identifying the defects which in turn results in increase in time and cost. Experience and technical know-how to screen the slightest possible defect is very vital. To access the efficiency of different types of PCBs in the presence of undesirable outputs is also essential. This is because it could provide a framework to assess the quality of individual printed circuit boards and hence work out appropriate interventions to prevent failures in the production process.

A number of studies have been carried out to deal with the micro-level problems of PCBs basically on production technologies, such as facility design (Chan 2005; Dengiz 2000; Grunow 2000; Ryan 2007), manufacturing technology (Altinkemer 2000; Doniavi 2000; Tsai 2007; Tsai (2005). There are very few studies which looked at the efficiency evaluation of PCBs manufacturing industries.

Wu (2007) applied the technique of data envelopment analysis (DEA) to measure the performance of PCBs producers by considering current assets, the number of employees, inventory investment and operating expenses as inputs and sales and gross margin as the outputs. Wang (2000) measured production and marketing efficiencies in the PCBs industry using Gray relation analysis (GRA) and data envelopment analysis (DEA) techniques for a sample of twenty three industries. He concluded that some of the industries priorities were on: production and marketing; production or marketing. However, all the studies focused on performance evaluation of PCBs industries. One of the issues that are still unexplored in the literature is the performance evaluation of different types of PCBs and the number of defective parts in a production. This research addresses the efficiency identification problem with regards to number of defective parts in PCBs industries in Ghana.

A special group of efficiency identification problem considered in this case relate to the errors occurring during production of PCBs. However, none of the studies explicitly attempt to access the efficiency in the presence of undesirable outputs.

Pitman (1983) offered a framework for assessing productivity when some outputs are undesirable and cannot be freely disposed. Basically, his focus however was on productivity and developing measures which penalized the performance of producers for generating undesirable outputs. Building upon Pitman's work, Färe (1989) introduced the idea of hyperbolic output efficiency measures, which provides an asymmetric treatment of defective and non- defective outputs.

The hyperbolic measure allows inputs to contract and outputs to expand. By different proportions, all non-defective outputs expand by a scalar and all defective outputs contract by the inverse of the scalar. A vital distinction of Färe (1989) however, was the introduction and imposition of weak disposability.

Since the work of Färe s (1989) there has been remarkable development of various models and their applications in the presence of undesirable outputs.

The state of the art in data envelopment analysis (DEA) models in the presence of undesirable outputs can be seen in works by Färe (1996, 2001, 2004,), Liu (1999), Lovell (1995), Färe (2004), Haynes (1994,1997), Tyteca (1997), Chung (1997), Scheel (2001), Seiford (2000). In the research papers of all the above mentioned researchers, little attempt was made on the part of detecting number of defective parts in PCBs using stochastic time series models. Rather, they focused on developing pollution prevention models using data envelopment analysis (DEA).

Due to the complex structure of this chemical operation based system and the batch work in most stages, classical flow line techniques are not applicable to analyse and hence detect the number of defective parts hence the need to apply stochastic time series models. This research demonstrates how stochastic time series model is used in practice to identify the number of defective parts in the production of PCBs. Dengiz s(2000) studied the assembly operations of the PCBs lines with different objectives using simulation as well as simulation based optimization techniques but did not consider the number of defective parts in the production of PCBs.

Additionally, since the process is frequently more complex, then distributions of defects are more appropriately modelled by the compound Poisson distribution.

In some situations, a defective item may have more than one defect that causes the item to be defective. Again, the production line defective items occur randomly over time which follows a Poisson distribution. On the other hand, if an item is defective, the number of defects per item will follow a different distribution.

In this research, it will be assumed that given an item is defective; the number of defects follows the geometric distribution. Thus, the distribution of defects over time is the Poisson compounded with the geometric. From the stand point of quality control, process quality can be improved by moving special causes. These two broad types of special causes are transient and persistent reported in the literature.

It is necessary that producers should identify and eliminate defects resulting from special causes in order to maintain and improve the product quality of PCBs.

According to Hawkins (1998) and Montgomery (1996) special causes can be classified into two broad types: sporadic (transient) and persistent special causes. Transient special cause affects a process for a short time and then disappears and the process would return to its original in-control level. However, the cause could occur again at some later time. The persistent special cause may shift the process to a new level and remain there until the problem detected and diagnosed.

Martiz (1970) is of the view that the empirical Bayes approach utilizes all relevant data in a statistical analysis. In a production process, observations are collected sequentially. We then use each observation one by one to monitor the production process. The first observation can provide the prior information for the second observation. Then based on the first observation, we could obtain posterior information about the second observation. Accordingly, the second observation would serve as prior information for the third observation.

He concluded that by doing this sequentially, we can update our knowledge about the mean defects in a production process. Manufacturing yield can be defined as the ratio of the number of good items produced to the total number of items produced and the number of good items produced accounts for the number of faulty items. It should be noted that more than one fault might be present in a faulty item.

Faults as defined by Hewlett Packard are “unacceptable deviations from the norm” (Orsejo, 1998). Testing and inspection is the means through which faults are observed. The objectives of testing and inspection are to: detect the faults, diagnose the faults, provide process control information and ensure confidence in the product.

In the printed circuit board industry, there are three major fault classes which included manufacturing faults, performance faults and specification faults , (Orsejo 1998) according to the author.

Defect detection by Automatic Optical Inspection (AOI) is a crucial step during the manufacturing process of printed circuit boards. Defect recognition is a strongly inter-disciplinary task, as it includes problems raised by industrial technology (estimating the effects of the artefacts), optics (exploiting up-to-date imaging devices, and dealing with their limitations) and advanced vision based quality assessment, (Kumar 2008).

By the method of (AOI), the paper proposed a hierarchical visual inspection framework, which implements a multi-level entity extraction approach. The researcher on this note introduced a model suited to a selected AOI task, called scooping detection. Scooping is a significant practical problem influencing the strength of solder joints in stencil prints. To the degree of scooping defect, the detection of these defects has a major role to improve the quality and reliability of electronic circuit assemblies. Times before, the quality of solder joints had been mostly verified by manual visual inspection, Takagi (1990). Holzmann (2004) concluded that as the number of components exceeds the possibilities of manual testing, reliable automation becomes a crucial need.

Therefore, to improve the quality and reliability of circuit board assemblies, the analysis of the manufacturing processes described above has a great importance. According to PCBs producers, the quality of the printed solder paste heavily influences the quality of solder joints. It has been reported in several studies that 52% - 71% of Surface Mount Technology (SMT) defects are related to the printing process, Richard (1999). According to Kripper & Beer., (2004), they concluded that although other opinions keep this phase less crucial, it was clear that detecting earlier the printing failures might result in notable cost savings. Based on the above assertion, they recommended that if in the inspection of PCBs, the number of defective parts and summarised volume of such defectives surpass a given thresholds, the board should be withdrawn.

While considerable research has been done on production planning in serial production systems, the concept of rework in stochastic time series modelling for total number of defective parts of PCBs has not been adequately addressed. Recognizing this need, this study is to address that problem. They observed that in each month, on average, 1.9% of the finished products needed major rework. They further showed that the percentage of products requiring rework had a significant impact on the overhead labour hours implying an increase of average cost per unit. SThe aim of this research is to provide a stochastic time series modelling framework that addresses the following questions: In a given production, how many defectives are obtained and how can one predict future number of defectives in a given production.

Shewhart on his part used control charts probability distributions to model the behaviour of product parameters in the production process. For control of defects, the assumption is made that the defects follow a Poisson distribution. In some situations, a defective item may have more than one defect that causes the item to be defective. Since in the production line, defective items occur randomly over time, the occurrence of defective items follow a Poisson distribution. On the other hand, if an item is defective, the number of defects per item will follow another distribution. Stapper (1995) also uses the compound Poisson model for defects in chips production. The defective chips in a wafer at time *t* follow a Poisson process. However, defects within chips at time *t* are assumed to follow a compound Poisson distribution. Jackson (1972) on his part points out that the compound Poisson results from many situations. One of these situations is one in which defects could occur in clusters. Thus, this situation is normally found in manufacturing industries. The clusters may follow a Poisson, but the defects within clusters show their own underlying phenomenon termed as branching process. However, the distribution of defects is also termed according to Neyman (1939) as contagious distribution. Jackson again points out that the compound Poisson process could occur when defects result from two underlying causes known as mixture of effects. Basically, three types of compound Poisson distribution are briefly reviewed.

2.8 Neyman's Compound Poisson Distribution

Neyman's Type *A* contagious distribution introduced by Neyman (1939) is a Poisson distribution compounded by another common Poisson distribution. Using the symbolic form proposed by Johnson (1969) we can represent Neyman's Type *A* distribution by:

$$\text{Poisson} \wedge \text{Poisson}$$

$$\frac{\lambda_1}{\lambda_2}$$

Thus, it is denoted by

$$P_k(\lambda_1, \lambda_2) = P_r[x = k] = \sum_{j-1}^{x} e^{-j\lambda_2}\left[(j\lambda_2)^k / k!\right], (k > 0)$$

So if $k = 0$, then

$$P_\circ(\lambda_1, \lambda_2) = P_r[x = 0] = \exp\left[-\lambda\left(1 - e^{-\lambda_2}\right)\right]$$

The expected value of defects in printed circuit boards is given as:

$$E(X) = \lambda_1 \lambda_2$$

and the variance as: $Var(X) = \lambda_1 \lambda_2 (1 + \lambda_2)$

It is clear that $Var(X)$ is always greater than $E(X)$, unless λ_2 is equal to zero. However, one of the features of the Poisson distribution is that the mean and variance are equal which is not suitable in describing manufacturing process. Due to this fact, Stapper (1985) and Gardiner (1987) discovered that a Poisson compounded by another Poisson would be better than a common Poisson describing the defect production process for chip manufacturing in the electronic industries. According to them, the relationship between production of chips per wafer and defect distribution can be illustrated as:

$Y = P_r(X_d = 0)$, where Y is the yield (PCBs) and X_d is the number of defects. It then follows that if the probability of zero defects is high; the production of PCBs is also high. Stapper in effect found that the variance of X_d, $Var(X_d)$ is much larger than the mean number of defects, $E(X_d)$.

2.9 Thomas Distribution

With regards to Johnson (1969) the Thomas distribution is illustrated as:

$$\frac{\lambda_1}{\lambda_2} + Poisson(\lambda_1)^{\wedge} Poisson(\lambda_2).$$

It can be observed that if defects occur in clusters, then the distribution of the number of defects follow a compound Poisson distribution. Hence, the number of defective items follows a Poisson, and the number of defects within a defective item follows another Poisson distribution. However, there must be at least one defect with a defective PCB. Obviously therefore, it is impossible to have a defective item with no defects at all since within defective PCB distribution is zero-truncated Poisson. Hence Thomas distribution is illustrated as:

$$Poisson(\lambda_1)^{\wedge} Poisson(1+\lambda_2).$$

John and Kotz also give the mean and the variance of the Thomas distribution as:

$$E(X) = \lambda_1(1+\lambda_2) \text{ and}$$

$$Var(X) = \lambda_1(1+3\lambda_2+\lambda_2^2).$$

Comparatively, the mean defects of the Thomas distribution are always λ_1 larger than the mean defects of the $Poisson(\lambda_1)^{\wedge} Poisson(\lambda_2)$.

2.10 Negative Binomial Distribution

According to Sherbrooke.,(1966) and Sheaffer (1976) a logarithmic Poisson distribution is a compound Poisson distribution in which the number of defectives has a Poisson distribution and the number of defects within a group follows a logarithmic series distribution. This is denoted as follows: $Poisson^{\wedge} \log arimic\, series$.

Sherbrooke on his part points out that the characteristics functions for both the logarithmic Poisson distribution and the negative binomial distribution are the same. Invariably, a negative binomial distribution is illustrated as:

$$P(x) = \frac{(k+x-1)!}{(k-x)! \times !} \times \frac{P^x}{(P+1)^{k+x}}$$, where p > 0 and k > 0.

Generating and expanding the above expression gives;

$$\left(\frac{P}{1+P} + \frac{1}{1+P}\right)^{-k}$$

The variance and mean of the negative binomial distribution or the logarithmic Poisson distribution according to Sheaffer and Leavenworth are:

$E(X) = kp$ and $Var(X) = kp(1+p)$

Clearly, the mean and the variance are not equal from the logarithmic Poisson distribution.

2.11 Researcher Interest and Model Justification

From the various literature observed by the researcher, it was realized that little or no information is available on the part of daily number of defective parts of printed circuit boards using stochastic time series methods. Therefore, since manufacturing systems can be described as a network of queues and servers, it is better represented and explained by the stochastic time series modelling and hence the Markov Chain system.

In most of the literature, binomial distribution, Poisson distribution, Gaussian probability distribution and logarithmic Poisson distribution models are used. The model that performed better and suitable to the researcher was the binomial distribution model.

The researcher therefore settled on the stochastic time series model with the reason that it can aid in adequately predicting, detecting the daily number of defective parts of printed circuit boards and hence determine whether the daily number of defective parts increase, decrease or remain the same and at which state would the process attain a steady state and at which state is production cost effective.

CHAPTER THREE

METHODOLOGY

3.0 Introduction

This chapter presents the methodology for the study. It has been sub divided into five sections. Section one would briefly discuss Markov Chains and its basic concepts. Section two would be dealing with transition probabilities, which will include defective items as a three state system and the formation of the transition probabilities. Section three also would discuss the Chapman-Kolmogorov equations. Additionally, section four would discuss the determination of the steady state distribution and then how the prediction of the number of defective parts in the production of the printed circuit boards could be calculated. Finally, the fifth section gives a brief discussion of time series and its basic concepts like stationarity and differencing processes.

3.1 Data Collection and Description

Secondary data is used in this research. The sample data consist of daily productions of printed circuit boards. Based on this, seventy-one (71) observations of the daily number of defective parts of printed circuit boards in the manufacturings industries in Ghana were obtained. It covers a six (6) year period spanning from January 2009 to December 2014.

The analysis would be carried out using STATA 12.1 and MINITAB 16 statistical software. The MINITAB 16 would be used to plot the graphs due to its pictorial clarity while STATA12.1 would be used to perform the model fitting. The data would be made stationary by differencing before analysing. Time series and Markov Chain models would be applied.

3.2 Stochastic Process

A stochastic process is a set of random variables $\{X_t / t \in T\}$, where T is called the parameter space of the process. The process $X_{(t)}$ is termed a stochastic process and the values assumed by the process are termed the states. Moreover, the set of possible values is termed the state space and the set of possible values of the indexing parameter t is known as the parameter space. According to Hiller (1990) a stochastic is defined to be an index collection of random variables $\{X_t\}$ where the index t runs through a given set of non-negative integers T. If $X_t = i$ then the process is said to be in state i at time t.

This means therefore that whenever the process is in state i, there is a fixed probability that the process will next be in state j.

A stochastic process $\{X_t\}$ is said to have the Markovian property if

$$P\{X_{t+1} = j / X_t = i, X_{t-1} = i_{t-1}, \ldots X_1 = i_1, X_0 = i_0\} = P\{X_{t+1=} \; j / X_t = i\} \qquad (1)$$

Equation (1) could be interpreted as stating that the conditional distribution of X_{t+1}, for given values of X_0, $X_1 \ldots X_t$, depends only on X_t, the most recent known state.

The conditional probabilities $P\{X_{t+1} = j / X_t = i\}$, denoted by P_{ij}, are called one-step transition probabilities. The value P_{ij} represents the probability that the process will, when in state *i*, next make a transition into state *j*. Since P_{ij} are conditional probabilities and since the process must make a transition into some state, we have

$$P_{ij} \geq 0, \; i, j \geq 0 \qquad (2)$$

$$\sum_{j=1}^{m} P_{ij} = 1, i = 0,1 \qquad (3)$$

Since we have one-step transition probabilities P_{ij}, we may also have the n-step transition probabilities. The "n-step" refers to the time interval between observations.

P_{ij}^{n} is the probability that a process in state i will be in state j after n additional transitions and is represented as:

$$P_{ij}^{n} = P\left\{X_{t+n} = j / X_{t} = i\right\}, n \geq 0, i, j \geq 0 \tag{4}$$

Basically, there are four types of stochastic processes which include:

- Discrete parameter space and continuous state space
- Continuous parameter space and discrete state space
- Continuous parameter space and continuous state space and
- Discrete parameter space and discrete state space.

The study would focus only on discrete parameter space and discrete state space since the production of printed circuit boards is discrete and the number of defective parts which is of three state spaces is also discrete.

For a stochastic process with discrete parameter and state spaces, we have

$$P_{ij}^{(m,n)} = P\left(X_{n} = j / X_{m} = i\right) \tag{5}$$

where $n \geq m$ and $i, j \in S$ (the state space). These probabilities are termed as transition probabilities which would be discussed in section two of this study.

3.3 Markov Chain and basic concepts

Within this context, the major purpose of this study is to discuss the concept of Markov Chain process and to indicate its potential usefulness in analysing the number of defective parts in printed circuit boards since it is a time – ordered data with some time span.

To provide the base for the analysis to follow, we sketch in this section the basic concepts of Markov Chain process and state the assumptions, definitions and theorems underlying this method that are necessary for this study.

The stochastic process $\left\{X_{(t)}, t \in T\right\}$ displays a Markov dependence if for a finite group of data points $\left(t_0, t_1, \ldots, t_n, t\right), t_0 < t_1 < t_2 < \ldots < t_n < t$ where $t,\ t_r \in T\left(r = 0, 1, 2, \ldots, n\right)$.

$$P\left(X_{(t)} \le x \mid X_{(tn)}\right) = x_n, X_{(tn-1)} = X_{n-1}, \ldots, X_{(tn)} = x_0)$$

$$= P\left[X_{(t)} \le x \mid X_{(tn)} = x_n\right] = F\left[X_n, x; t_n, t\right] \qquad (6)$$

From equation (6), the relation is as follows;

$$F\left(X_n, x; t_n, t\right) = \int_{y \in s} F\left(y, x; \tau, t\right) \partial F\left(X_n, y; t_n, \tau\right) \qquad (7)$$

where $t_n < \tau < t$ and S is the state space of the process $\left\{X_{(t)}\right\}$. When the stochastic process has discrete state and parameter space in equation (7) becomes: for $n > n_1 > n_2 > \ldots n_k$ and $n, n_r \in T\left(r = 1, 2, \ldots k\right)$

$$P\left(X_n = j \mid X_{n_1} = i, X_{n_2} = i_2, \ldots X_{nk} = ik\right) \qquad (8)$$

$$= P\left(X_n = j \mid X_{n_1} = i_1\right) = P_{ij}^{(nk,n)}$$

A stochastic process with discrete space and parameter space which displays Markov dependency as in equation (8) is known as a Markov process.

Markov Chain is a mathematical system that undergoes transitions from one state to another on a state space. It is basically a random process normally termed as memory less since the next state depends only on the current state and not on the sequence of events that preceded it.

Furthermore, a Markov Chain is a sequence of random variables $X_1, X_2, ... X_n$ with the Markov property which states that given the present state, the future and the past state are independent.

Mathematically, a Markov Chain is presented as follows:

$$P\{X_{n+1} = x / X_1 = x_1, X_2 = x_2 ... X_n = x_n\} \quad (9)$$

$$P\{X_{n+1} = x / X_n = x_n\} \quad (10)$$

$$P\{X_1 = x_1 ... X_n = x_n\} > 0 \quad (11)$$

where the possible values of X_i from a countable set S is termed as the state space of the chain. In this study, the state space is of three states.

If in any given sequence of experiments the outcome of each particular experiment depends on some chance event, then any such sequence is termed a stochastic process. The process is finite since the set of possible outcomes is finite.

3.4 Finite Stochastic Process in relation to Markov Chain

A finite stochastic process with outcome functions such as $X_0, X_1 ... X_n$ is a stationary Markov Chain process if the starting state, given by X_0, is fixed and

$$P[\{X_0 = t / X_{n-1} = s\}] \quad (12)$$

$$P[\{X_n = t / X_{n-1} = s\}] = P\{X_m = t / X_{m-1} = s\} \quad (13)$$

For all $m \geq 1, n \geq 2$ and any possible sequence of outcomes s, t

This prior statement could be read as the probability of $X_n = t$ given

$X_{n-1} = s$.

The interpretation of the above definition is that the outcome of a given process depends only on the outcome of the immediately preceding process and that this dependence is the same at all stages.

3.5 Types of States

- Stationary State: A probability distribution S on X is stationary if $S = S.P$.
- Accessibility State: A state j is said to be accessible from state i if there is a positive-probability path from i to j.
- Communicating State: States i and j are said to communicate if each is accessible from the other. This relation is denoted by $i \leftrightarrow j$.
- Periodic State: The period of a state i is the greatest common divisor of the set $\{n \in N : P^n(i,i) > 0\}$. If every state has period 1 then the Markov Chain is termed as aperiodic.
- Transient State: A state i is said to be transient if, given that we start in state i, there is a non-zero probability that we will never return to i.
- Ergodic State: A state i is said to be ergodic if it is aperiodic and positive recurrent.

3.6 Markov Property

The Markov property is of the view that if a state is known for any specific value of the time parameter t, that information is enough to predict the behaviour of the process beyond t.

From the above definition and for $n_k < r < n$ we obtain

$$P_{ij}^{(nk,n)} = P\left(X_n = j \mid X_{nk} = i\right)$$

$$= \sum_{m \in s} P\left(X_n = j \mid X_r = m\right) P\left(X_r = m \mid X_{nk} = i\right)$$

$$= \sum_{m \in s} P_{ij}^{(nk,r)} P_{mj}^{(r,n)} \qquad (14)$$

Equations (6) and (14) are termed as the Chapman-Kolmogorov equations for the process.

3.7 Assumptions of Markov Chains

- Only the last state influences the next state.
- On the part of time stationary property, one-step transition probabilities do not depend on when the transition occurs. Thus, $P\{X_{n+1} = j / X_n = i\}$ is the same for all $n = 0,1,2...$

3.8 Transition Probabilities

The one-step transition probability is the probability of moving from one state to another in a single step. It is worth mentioning that the Markov Chain is said to be time homogeneous if the transition probabilities from one state to another are independent of time index. In a mathematical representation of transition probabilities, we have

$$P_{ij} = P\{X_n = j / X_{n-1} = i\} \qquad (15)$$

Under the production of printed circuit boards, it is observed that the number of defective items would experience three state spaces which include the following: Thus, the number of defective parts could either;

- Increase
- Decrease or
- Remain the same.

For the purpose of this study, the information is summarized in a table shown below.

Table 1: Values, States and their Descriptions

Value	State	Description
1	First State	Defective Increase
2	Second State	Defective Decrease
3	Third State	Defective Remain the same

It is worthwhile indicating that if a number of defective parts are observed to increase from one state to another it automatically falls under state one. On the other hand, a decrease from one state to another means it falls under state two and if a number of defective parts remain the same from one state to another falls under state three.

3.9 Formation of the Transition Matrix

The transition matrix P is the matrix consisting of the one-step transition probabilities P_{ij} .With the information specified above, it would then be possible to construct a possibility tree and attach branch weights that would describe the process as it moves through any finite number of steps. Alternatively, the transition probabilities P_{ij} can be represented in the form of a transition matrix P such as:

$$P = \begin{pmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{pmatrix} \qquad (16)$$

where $\sum_j P_{ij} = 1$ and $P_{ij} \geq 0$ for all i and j.

In the elements of equation (16), the P_{ij} denotes the probability of moving from state S_i to S_j in the next step. Since the elements of this matrix are non-negative and the sum of the elements in any row is 1, each row of the matrix is termed a probability vector and the matrix $[P]$ is a stochastic matrix. This matrix together with an initial starting state, completely defines a Markov Chain Process.

3.10 The n-Step Transition Probability Matrix

Let P be the transition probability matrix of a Markov Chain $\{X_n, n = 0,1,2....\}$ is defined as the time interval between observations. When n=1, we shall write $P_{ij}^{(1)} = P_{ij}$. Due to the dual subscripts, it is convenient to arrange these transition probabilities in a matrix form. Thus, we shall write $P = \left\| P_{ij} \right\| = \begin{bmatrix} P_{00} & P_{01} & P_{02}\cdots \\ P_{10} & P_{11} & P_{12}\cdots \\ . & . & . \end{bmatrix}$ and

$$P^n = \left\| P_{ij}^{(n)} \right\| = \begin{bmatrix} P_{00}^{(n)} & P_{01}^{(n)} & P_{02}^{(n)}\cdots \\ P_{10}^{(n)} & P_{11}^{n} & P_{12}^{(n)}\cdots \\ . & . & . \end{bmatrix} \qquad (17)$$

Clearly, we have $\sum_{j \in s} P_{ij}^{(n)} = 1, \; n = 0,1,2...$

Thus, if $P^2 = \begin{bmatrix} P_{00}^{(2)} & P_{01}^{(2)} & P_{02}^{(2)} \\ P_{10}^{(2)} & P_{11}^{(2)} & P_{12}^{(2)} \\ P_{20}^{(2)} & P_{21}^{(2)} & P_{22}^{(2)} \end{bmatrix} = P_{ij}^2$, then $P_{ij}^{(2)}$ is the probability of being the j^{th} state from state i in two steps. Again, if $P^3 = \begin{bmatrix} P_{00}^{(3)} & P_{01}^{(3)} & P_{02}^{(3)} \\ P_{10}^{(3)} & P_{11}^{(3)} & P_{12}^{(3)} \\ P_{20}^{(3)} & P_{21}^{(3)} & P_{22}^{(3)} \end{bmatrix} = P_{ij}^{(3)}$, then $P_{ij}^{(3)}$ is the probability of being the j^{th} state from i in three steps.

The above illustrations can obviously be indicated from the Chapman-Kolmogorov equation (14) as follows; for a given r and s ,

$P_{ij}^{(s+r)} = \sum_{k \in s} P_{ik}^{(r)} P_{kj}^{(s)}$ and setting r = 1 and s =1 we obtain,

$$P_{ij}^{(2)} = \sum_{k \in s} P_{ik} P_{kj}$$

Basically, $P_{ij}^{(2)}$ is the $(i, j)th$ element for the matrix product $P \times P = P^2$. Supposing $P_{ij}^{(r)} \left(r = 3, 4, 5, \ldots, n\right)$ is the $(i, j)th$ of P^r then by the Kolmogorov equation, the $P_{ij}^{(r+1)} = \sum_{k \in s} P_{ik}^{(r)} P_{kj}$ which in effect can be described as the $(i, j)th$ element of the matrix product $P^r P = P^{r+1}$. Therefore by the application of the method of induction, $P_{ij}^{(n)}$ is the $(i, j)th$ element of $P^n, n = 2, 3, 4, \ldots$

For the model specification, the basic assumption is established about the identified n-step transition probability. Hence, the probability matrix is accessible and communicates. Recurrence and transience states also exist in the process.

3.11 Limiting distribution of a Markov chain

Given that P is the transition probability matrix of an aperiodic, irreducible, finite state Markov chain, then

$$\lim_{t\to\infty} P^t = \pi = \begin{bmatrix} \sigma \\ \sigma \\ . \\ . \\ \sigma \end{bmatrix} \tag{18}$$

where $\sigma = [\sigma_1, \sigma_2, ...\sigma_m]$ with $0 < \sigma_j < 1$ and $\sum_{j=1}^{m} \sigma_j = 1$. Therefore, a Markov chain that exhibit this type of feature is said to be ergodic and has a limiting distribution π.

3.12 Chapman-Kolmogorov Equations

From the above discussed information, we could determine the outcome of, say, the n^{th} step. In matrix language, this could be developed in the following way: We let,

$S_0 =$ The initial vector or starting state matrix and

$\boldsymbol{P} =$ The transition probability matrix, then

$$S_1 = S_0 P \tag{19}$$

$$S_2 = S_1 P \tag{20}$$

.

.

$$S_n = S_{n-1} P \tag{21}$$

From the above expression, we assumed the initial state distribution matrix to be as follows:

$$S_0 = (1 \quad 0 \quad 0) \tag{22}$$

Therefore, if we start in state i then S_1 is the i^{th} row of P_1S_1 which is the i^{th} row of P_n, and thus S_n is the i^{th} row of P_n. Hence, the rows of P_n give us the outcome vectors for various starting states. The P_{ij} will be the probability that the process will be in state j after n steps if it started in state i .

In effect, we require all states to be accessible, meaning that there is a non-zero probability of moving from state i to state j in a finite number of time periods. The conditions are that some powers of the matrix should have only positive components and equation (19) must be satisfied.

Again, for equilibrium number of defective parts of printed circuit boards to exist, equations (23) and (24) must be satisfied.

3.13 Determining the Steady State Distribution

The steady state is the probabilities that various states will be repeated will remain constant. The steady state is not achieved until sometime after the system is initiated and this initial situation is termed as the transient state. Mathematically, the steady state is obtained if and only if

$$S.P = S \qquad (23)$$

$$S\left(P - I\right) = 0 \qquad (24)$$

where S is the stationary matrix, P is the transition probability matrix, I is the (3×3) unit or identity matrix and 0 is the (3×3) null matrix. Since S is a probability vector, we have

$$\sum_j S_j = 1 \qquad (25)$$

With this information, we could combine equations (23) and (24) to form a system of n unknowns from which we can solve for the unique values of S .

3.14 Model specification

Given definition to the problem (daily number of defective parts of PCBs as a three-state Markov chain process); we let F_t represent the daily number of defective parts of PCBs at time t where t = 0,1,2,….n (t is measured in daily production time intervals). Furthermore, we define $\Delta_t = Y_t - Y_{t-1}$ which measures the change in the daily number of defective parts of PCBs at time t. Taking note of the production of PCBs as a discrete time unit for which we define a random variable X_t to indicate the state of daily number of defective parts of PCBs at time t, a vector spanned by 1,2,3;

$$X_t = \begin{cases} 1 \; if \; d_t = 1 \, increase \; in \; daily \; number \; of \; defective \; parts \; of \; PCBs \; from \; t-1 \; to \; t \\ 2 \; if \; d_t = 2 \, decrease \; in \; daily \; number \; of \; defective \; parts \; of \; PCBs \; from \; t-1 \; to \; t \\ 3 \; if \; d_t = 3 \, no \; change \; in \; daily \; number \; of \; defective \; parts \; of \; PCBs \; from \; t-1 \; to \; t \end{cases}$$

Therefore, the estimates of the probability that the daily number of defective parts increase, decrease or remain the same can be calculated respectively by;

$$\hat{P}_0 = \frac{n_0}{n}, \hat{P}_1 = \frac{n_1}{n} \; and \; \hat{P}_2 = \frac{n_2}{n}$$

3.15 Time Series and its Basic Concepts

A time series $\{X_i; t \in T\}$ can be defined as an ordered sequence of random variables over time, where T denotes an index time points set (Akgun, 2003). It is regarded as the record of some activity, with observations taken at equally spaced time interval. Basically, there are two types of time series data namely discrete data, where we have observations at (usually regularly) space interval for instance weekly prices of shares, daily rainfall, monthly inflation, number of defective parts of printed circuit boards and many others and continuous data, where we have observations at every stage of time such as electrocardiograms. However, the discrete time series data is considered under this research since both the parameter space and the state space of the research are discrete.

Time series analysis consist of methods that attempt to comprehend the underlying generation process of the data points and construct a mathematical model to represent the process. The constructed model is then used to forecast future events based on known past events. Time series often make use of the natural one-way ordering of time so that values in a series for a given time will be expressed as being derived from past values rather than future values. Normally a time series model reflects the fact that observations close together in time domain are more correlated as compared to observations further apart. Usually time series data are made up of four patterns that are derived on casual factors identified by time series analysis methods. The four patterns that characterize economic and business series are the trend component, cyclical or periodic component, seasonal component and the error or residual component.

The trend component deals with the general and overall pattern of the time series; the cyclical component refers to the variation in the series which arise out of the phenomenon of business cycles, which usually spans within periods of more than one year. The seasonal variations refers to the periodic ups and downs in the series that occur within a year and finally the error term is the component that contains all moments which neither belong to the trend nor to the cycle nor to the seasonal component.

A lot of models are used for time series data analysis. However, these models are classified as the linear and the non-linear. Basically, the linear models are the autoregressive (AR) model of order (p), moving average (MA) of order (q), and a combination of the autoregressive (AR) model and the moving average (MA) model to give the autoregressive moving average (ARMA) model of order (p, q). Other linear models include the autoregressive integrated moving average (ARIMA) model, the seasonal autoregressive integrated moving average (SARIMA) model, and the autoregressive fractionally integrated moving average (ARFIMA) model and many others.

The other types of models which often represent or reflect the changes of variance along with time known as heteroscedasticity are the non-linear models. Under these types of models, changes in variability are related to and/or predicted by recent past values of the observed series. Some of the non-linear models include the symmetric models such as the autoregressive conditional heteroscedastic (ARCH) model with order (p) and the Generalized ARCH (GARCH) model with order (p, q). Other asymmetric models are the Power ARCH (PARCH), Threshold GARCH (TGARCH), Exponential GARCH (EGARCH), Integrated GARCH (IGARCH) and many others. The fact remains that all these asymmetric models have order (p, q). This research would make use of the AR model, MA model and the ARMA model from the linear class of models.

3.16 Stationary Process

Actually, stationarity forms the basis of time series analysis. Normally, before a time series analysis is carried out, there is always the need to verify whether the series is stationarity or otherwise. In this research work, the Augmented Dickey- Fuller (ADF) test for stationarity was carried out to confirm the stationarity of the series. On the other hand, an assumption of stationarity is usually considered. In this section, stationarity is defined and the process described.

A series is considered to be stationarity if the mean and the auto covariance of the series do not depend on time. Stationarity exist in two way namely, strictly stationary and weakly stationary. A time series $\{X_t\}$ is said to be strictly stationary if the joint distribution of $(X_1,...,X_t)$ is the same as that of $(X_{1+t},...,X_{k+1})$ for all t, where k is an arbitrary positive integer and $(1,...,k)$ is a collection of k positive integers. The shifting of the time origin by t has no effect on the joint distribution which depends only on the intervals between the two set of points given by t which is called a lag. In this research, the weak stationarity also called stationarity in the second moment is considered.

A time series $\{X_t\}$ is weakly stationary if both the mean of X_t and the covariance between X_t and X_s are time-invariant. Thus time series $\{X_t\}$ is weakly stationary if:

$E(X_t) = \mu$, which is a constant and

$c\text{ov}(X_t, X_s) = \gamma$, is only a function of time distance between the two random variables and does not depend on the actual point in time t.

3.17 Differencing process

The fact remains that most time series data are non-stationary in nature. Stationarity therefore forms the foundation of time series analysis. There are a lot of possible causes of the non-stationarity of a time series, X_t which include the trend (long-term change in $E\left[X_t\right]$, seasonality (constant changes in $E\left[X_t\right]$ at regular periods), and non-constant variance of the random errors in X_t. One of the possible remedial measures to stationarise X_t therefore is differencing X_t to remove the trend and/or seasonality (Chatfield, 2004, p19) as cited in (Luruli, 2011). However, there is always the need to induce stationarity in the data before the analysis is conducted. Differencing of time series data can be done once or twice. Differencing is a special type of filtering, which is particularly useful for removing trend and/or seasonality in a series. For *d* a positive integer, the differencing filter $\nabla^d X_t = (1-B)^d X_t$ removes the trend of the polynomial of degree *d* in the series X_t if present, and the differencing filter $(1-B^d)X_t = X_t - X_{t-d}$ removes seasonality of period *d* in the series X_t if present.

If $X_t = \beta_\circ + \beta_t + \varepsilon_t$, then

$$(1-B)X_t = X_t - X_{t-1} = \beta_\circ + \beta_t + \varepsilon_t - \beta_\circ - (t-1) - \varepsilon_{t-1} = \beta + \varepsilon_t - \varepsilon_{t-1},$$

is free of the linear trend. The differencing was done once in this research.

3.18 Smoothing Techniques

A time series is a sequence of observations, which in description are ordered in time. Distinctive in the collection of data taken over time is some form of random variation. There exist some methods for reducing or cancelling the effect due to random variation. A widely used technique is "smoothing''. The advantage of this technique, when properly used, reveals more clearly the underlying trend, seasonal and cyclic components. Smoothing techniques are applied to in order to reduce irregularities (random fluctuations) in time series data. They provide a clearer view of the real underlying behaviour of the series. There are many different smoothing techniques which are classified into linear and non-linear time series. The moving average is a linear smoothing and that is what is applied in this study.

3.19 Moving Average Smoothing Techniques

Moving averages rank among the most popular techniques for the preprocessing of time series. They are used to filter random "white noise" from the data, to make the time series smoother or even to emphasize certain informational components contained in the time series. A moving average is obtained by calculating a series of average of different subsets of the full data set. Given a series of numbers and a fixed subset size, the first element of the moving average is obtained by taking the average of the initial fixed subset of the number series. Then the subset is modified by "shifting forward"; that is, excluding the first number of the series and including the next number following the original subset in the series. This creates a new subset of numbers, which is averaged. This process is repeated over the entire data series. The following formula is used in finding the moving average of order n, MA (n);

$$M_t = \frac{(X_t + X_{t-1} + ... + X_{t-n+1})}{n} \qquad (26)$$

3.20 Autoregressive (AR) Model

The autoregressive (AR) model uses past values of the dependent variable to explain the current value. According to Hamilton (1982) AR model is the most ordinary autoregressive models used in time series analysis. Let $\{\varepsilon_t \mid t \in T\}$ be a white noise process with mean zero and variance σ^2. A process $\{x_t \mid t \in T\}$ is said to be an autoregressive time series of order *p* (denoted AR (p)) if

$$x_t = \alpha_\circ + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + ... + \alpha_p x_{1-p} + \varepsilon_t$$

$$x_t = \alpha_\circ + \sum_{i=1}^{p} \alpha_i x_{t-i} + \varepsilon_t \qquad (27)$$

where $\alpha_\circ$ is a constant and α_i are parameters of the model.

Equation (27) can be written using the backshift operator as

$$x_i = \alpha_\circ + \sum_{i=1}^{p} \alpha_i B^i x_t + \varepsilon_t$$

Moving the summation term to the left side and using polynomial notation, AR (p) can be written as;

$$\alpha(B) x_t = \alpha_\circ + \varepsilon_t \qquad (28)$$

where $\alpha(B) = I - \alpha_1 B - \alpha_2 B^2 - \ldots - \alpha_p B^p$.

Let $\theta(B) = \alpha^{-1}(B) = I + \theta_1 B + \theta_2 B^2 + \theta_3 B^3 + \ldots$

in which there is a relationship between α_s and θ_s. Hence, equation (28) could be written as

$$\begin{aligned} x_t &= (\alpha_\circ + \varepsilon_t) / \alpha(B) \\ &= (\alpha_\circ + \varepsilon_t)\theta(B) \\ &= \mu + \varepsilon_t \theta(B) \\ &= \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \ldots \end{aligned} \qquad (29)$$

where μ is a constant and can be calculated by

$$\mu = \frac{\alpha_\circ}{(1 - \alpha_1 - \alpha_2 - \ldots \alpha_p)} \qquad (30)$$

It follows that $E(X_t) = \mu$ and the covariance function is

$$\sigma(t, t+h) = \sigma^2 \sum_{s=0}^{\infty} \theta_s \theta_{s+h} \qquad (31)$$

A model with a combination of autoregressive terms and moving average terms is termed as mixed autoregressive moving average model. The notation ARMA (p, q) is used to represent these models for our convenience, where p is the order of the autoregressive part and q is the order of the moving average part. The orders of autoregressive and moving average terms in an ARMA model are determined from the pattern of sample autocorrelation and partial autocorrelations. A model for the series X_t can be an AR (p) model or an MA (q) model or a combination of both. Hence the addition of both the AR (p) and MA (q) is termed as an autoregressive moving average of order (p, q), illustrated by ARMA (p, q), and is expressed as;

$x_t = \alpha_\circ + \alpha_1 x_{t-1} + \ldots + \alpha_p x_{t-p} + \beta_1 \varepsilon_{t-1} + \ldots + \beta_q \varepsilon_{t-q} + \varepsilon_t$,where

$\alpha_1,\ldots,\alpha_p,\beta_1,\ldots\beta_q$ are model parameters to be estimated, and ε_t is a series of random errors each with zero mean and constant variance σ^2 according to (Box, 1976).

3.21 AR (1) Model

The AR (1) denotes a special case of the general AR (p) model. The AR (1) process with $\alpha_\circ = 0$ where $p = 1$ becomes

$$x_t = \alpha x_{t-1} + \varepsilon_t \quad (32)$$

When $|\alpha| = 1, x_t$ is called a random process and then,

$$x_t = x_\circ + \sum_{i=1}^{t} \varepsilon_i$$

It follows that $E(x_t) = 0$ and $Var(x) = Var(x_\circ) + t\sigma^2$. As the variance changes with t, the process is non-stationary. When $|\alpha| > 1$ and given that $E(\varepsilon_t) = 0$, then the random term ε_t will eventually disappear and the equation becomes:

$$x_t = \alpha x_{t-1} \qquad (33)$$

The process is stationary given that $E(x_t) = 0$ and $Var(x_t) = \frac{\sigma^2}{1-\sigma^2}$.

3.22 Model identification, estimation and diagnostics

3.22.1 Model identification

A basic AR model for a given time series is identified by the sample autocorrelation function ACF and the partial autocorrelation function, PACF (Gooijer 1985). The AR (p) model for a given time series, x_t is given by

$$x_t = \alpha_\circ + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + ... + \alpha_p x_{t-p} + \varepsilon_t \ .$$

Using the backshift operator B and with $\alpha_t = 0$

$$\alpha(B) x_t = \varepsilon_t \qquad (34)$$

where $\alpha(B) = 1 - \alpha_1 B - \alpha_2 B^2 - ... - \alpha_p B^p$ is a polynomial in B of order p. An AR (p) process is said to be stationary provided that the absolute roots of the polynomial in B, $\alpha(B) = 0$ are all greater than 1 (Wei, 2006 p40) as cited in (Fululedzani, 2011). The ACF of a time series, x_t that is generated by an AR (p) process decays exponentially with lag k. Thus, if a time series x_t is generated by an AR (p) process then its sample autocorrelation function (ACF),

ρ_k = Correlation (x_t, x_{t+k}) is given as:

$$\rho_k = \frac{\gamma(t+k,t)}{\sqrt{\gamma(t+k,t+k)\gamma(t,t)}}, k = 0,1,2... \qquad (35)$$

where $\gamma(t+k,t)=$ covariance (x_t, x_{t+k}), $\gamma(t+k,t+k)=$ variance (x_{t+k}) and $\gamma(t,t)=$ variance (x_t) (Montgomery 2008). ρ_k measures the correlation between x_t and x_{t+k}. The estimate of ρ_k is given by:

$$\hat{\rho}=\frac{\hat{\gamma}(t+k,t)}{\sqrt{\hat{\gamma}(t+k,t+k)\hat{\gamma}(t,t)}}, k=0,1,2... \quad (36)$$

where $\hat{\gamma}(t+k,t)=\frac{1}{n-k}\sum_{t=1}^{n-k}\left(x_t-\bar{x}\right)\left(x_{t+k}-\bar{x}_t\right)$, $\hat{\gamma}(t+k,t+k)=\frac{1}{n-k}\sum_{t=1}^{n-k}\left(x_{t+k}-\bar{x}_t\right)^2$

and $\hat{\gamma}(t,t)=\frac{1}{n}\sum_{t-1}^{n}\left(x_t-\bar{x}\right)^2$

3.22.2 Model estimation

As a basic AR model has been identified, the next step is to estimate the parameters of the model. Generally, the basic methods for estimating model parameters are either the least squares method or the maximum likelihood methods (Box, 1976). These methods are reviewed under the assumption that the identified AR model for a given time series can be expressed as equation (34) i.e.

$$\alpha(B)x_t=\varepsilon_t$$

Let

$$SSE\left(\underset{-}{\alpha}\right)=\sum_{t=1}^{n}\varepsilon_t^2 \quad (37)$$

and

$$L\left(\underset{-}{\alpha},\sigma^2\right)=\left(2\pi\sigma^2\right)^{\frac{-n}{2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^{n}\varepsilon_t^2\right) \quad (38)$$

where $\underset{-}{\alpha} = \left(\alpha_1, \alpha_2, ..., \alpha_p\right)^T$

3.23 Least squares method

The least squares procedure for estimating the parameters of the AR model are obtained by considering values of which minimize the error sum of squares.

3.24 Maximum likelihood method

Based on the assumption $\varepsilon_t \sim N\left(0, \sigma^2\right)$, the maximum likelihood estimates of the parameters of AR model are values of σ^2 and $\underset{-}{\alpha}$ which maximize the likelihood function in equation (38). In this research, the maximum likelihood method is used to estimate the values of the parameters of the AR model.

3.25 Model diagnostics

The identified AR model may seem to be the best for the time series X_t. It may be inadequate as a result of violation of the assumption of stationarity of X_t, the presence of outliers in X_t and over or under-parameterization of the model. Checking for the adequacy of the fitted model is done to test whether or not the residuals from the fitted model are white noise and whether or not the model is the simplest best. The plot of the ACF of residuals is used for checking the model adequacy. If the AR model adequately fits the given time series, then the residuals from fitting the model should be white noise. However, a plot of the ACF of residuals should show no significant spikes at all lags *k*.

3.26 Forecasting with the AR model

The main objective of developing a time series model is for forecasting future values of the series. In this research, modelling and forecasting of the total number of defective parts of printed circuit boards in the manufacturing industry was investigated. Given the time series $X_1, X_2 ... X_t$, forecasting is done to predict the value of $x_{t+k}\left(k=1,2,3,...\right)$, the series that will be observed at time $t+k$ in the future. By the minimum mean square error criterion of forecasting, the estimated value is $\hat{x}_t(k)$, which minimizes the conditional mean square error;

$$E\left[\left(x_{t+k}-\hat{x}_t(k)\right)^2 \left|x_1, x_2, ..., x_t\right.\right] \qquad (39)$$

Therefore when we differentiate equation (39) with respect to $\hat{x}_t(k)$, then equating to zero and solving for $\hat{x}_t(k)$ gives;

$$\hat{x}_t(k) = E\left[x_{t+k}\left|x_1, x_2, ..., x_t\right.\right] \qquad (40)$$

Thus, $\hat{x}_t$ (k) is the maximum mean square error forecast of the value x_{t+k}. The focus is on the calculation of $\hat{x}_t$ (k) when the time series is generated by an AR process. We recall that the AR (p) model for a given time series x_t is given by;

$$x_t = \alpha_\circ + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + ... + \alpha_p x_{t-p} + \varepsilon_t$$

In general, a k-step ahead forecast implies that x_{t+k} is given by;

$$x_{t+k} = \alpha_\circ + \alpha_1 x_{t+k-1} + \alpha_2 x_{t+k-2} + ... + \alpha_p x_{t+k-p} + \varepsilon_{t+k} \qquad (41)$$

Using equations (40) and (41), the minimum mean square error forecast of the x_{t+k} value is given by;

$$\hat{x}(k) = E\left[x_{t+k} \mid x_1 + x_2, \ldots, x_t\right]$$

$$= \alpha_\circ + \alpha_1 E\left[x_{t+k-1} \mid x_1, \ldots, x_t\right] + \alpha_2 E\left[x_{t+k-2} \mid x_1, \ldots, x_t\right] + \ldots + \alpha_p E\left[x_{t+k-p} \mid x_1, \ldots, x_t\right]$$

$$\hat{x}_t(k) = \alpha_\circ + \alpha_1 \hat{x}_{t+(k-1)} + \alpha_2 \hat{x}_{t+(k-2)} + \ldots + \alpha_p \hat{x}_{t+(k-p)} \qquad (42)$$

For example, if $k = 2 < p$ then,

$$\hat{x}_t(2) = \alpha_\circ + \alpha_1 \hat{x}_t(1) + \alpha_2 \hat{x}_t + \ldots + \alpha_p \hat{x}_{t+2-p} \qquad (43)$$

Hence the focus error in $\hat{x}_t$ (k) from equation (43) is given as

$$e_t(k) = x_{t+k} - \hat{x}_t(k)$$

The best model for forecasting is one with the least mean square error (MSE).

The ACF and PACF are used in determining the order of the model. Several models with different orders can be considered but the ultimate model must be selected from the family of competing models that characterize the ordering data.

There are so many selection criteria that have been proposed to help in selecting the most ideal model. Among these selection criteria are the Akaike Information criterion (AIC) introduced by Akaike (1974), Bayesian Information criterion (BIC) proposed by Schwartz (1978) and many others. This research makes use of the criteria mentioned above to select the best models. The competing models are ranked according to their AIC and BIC values. However, the model with the lowest information criterion value is being considered the best. On the other hand, if it so what happens that two or more competing models have the same or similar AIC and BIC, then the principle of parsimony is applied to choose the most ideal model. Basically, the principle of parsimony states that a model with fewer parameters is normally better than one with so many parameters.

3.27 Akaike Information Criterion

The Akaike Information Criterion was the first model selection criterion to meet global standards. The Akaike Information Criterion is a measure of the relative quality of a statistical model for a given set of data. The AIC also provides a means for model selection since it is founded on information theory. Again, the AIC is an extension of the maximum principle and the maximum likelihood principle used to estimate the parameters of the model once the order(s) of the model has/have been identified. Basically, Akaike Information Criterion deals with the trade-off between the goodness of fit of the model and the complexity of the model. The AIC is defined by:

$$\begin{aligned} AIC &= -2(\log likelihood) + 2k \\ &= 2k - 2In(L) \end{aligned} \qquad (44)$$

where k is the number of estimated parameters in the model and L is the maximized value of the likelihood function for the model.

Assuming a class of competing models of various structures, the maximum likelihood estimation is used to fit the model and the AIC is computed based on each model fit. The selection of the best model is then made by considering the model with the minimum AIC value. The first term of the AIC in equation (44) measures the goodness of fit of the model whilst the second term is called the penalty function of the criterion since it penalizes a competing model by the number of parameters used.

The AIC is very vital for both in-sample and out- of-sample forecasting performance of a given model. In-sample forecasting shows how the chosen model fits the data in a given sample while out-of-sample forecasting is concerned with determining how a fitted model best forecast future values.

3.28 Bayesian Information criterion

The Bayesian Information criterion (BIC) is one other method very vital for selecting the most appropriate model out of a class of competing models. The BIC is obtained by replacing the non-negative factor 2(*k*) in the AIC formula by *k ln (n)*. Hence, the BIC is defined as;

$$BIC = -2(\log likelihood) + k.In(n)$$

$$= k.In(n) - 2In\hat{L} \qquad (45)$$

where k represents the number of parameters to be estimated in the model, *n* is the length of the time series (sample size) and $\hat{L}$ is the maximized value of the likelihood function of the model. The maximum likelihood estimation is used to fit the model. The BIC is calculated for each of the models in a class of competing models and the fitted model with the minimum BIC value is chosen as the most ideal one.

3.29 Model diagnostic checks and adequacy

Under this, the model diagnostic checks were carried out to determine the goodness of fit of a selected model. The model diagnostic checks were carried out on the standardized residuals (Talke, 2003). The residuals are assumed to be independently and identically distributed following a normal distribution (Tsay, 2002). Hence, plots of the residuals such as the normal probability plot and the time plot of residuals could be used. If the model fits the data adequately, the normal probability plot should be a straight line while the time plot should show random variation.

3.30 Conclusion

This chapter basically has provided a summary of Markov Chains and its basic concepts as well as the transition probability matrix, steady state distribution, stochastic process, Chapman-Kolmogorov equations, n-step transition probability matrix and then the limiting distribution of a Markov chain. Additionally, a summary of time series and its basic concepts as well as a detailed description of the model order determination, estimation and forecasting of the autoregressive (AR), moving average (MA) and the autoregressive moving average (ARMA) models that are used in this research. Again, various selection methods that are of great help in the selection of the best fit model among a group of competing models are also discussed. Furthermore, model estimations are performed on the differenced series. The methods discussed under this chapter are followed in the next chapter so as to aid run the analysis for this research.

CHAPTER FOUR

DATA ANALYSIS AND DISCUSSION OF RESULTS

4.0 Introduction

This chapter presents the various analyses of the study and the discussion of the results obtained from the analyses.

4.1 Data Description

Data used for this study are daily number of defective parts of PCBs spanning from January 2009-December 2014. In effect, seventy one (71) observational data points were obtained. Table 2 illustrates the descriptive (summary) statistics of the daily number of defective parts of printed circuit boards in the manufacturing industry in Ghana for the period under discussion. Generally, the frequency of least number of defective parts of PCBs was clearly recorded over the study period. Considering the transition probabilities, the aim is to select the state(s) that recorded the least daily number of defective parts of PCBs. This in effect means that the state(s) that recorded the least daily number of defective parts respectively is cost effective on the part of producing PCBs. We expect that the choice of daily number of defective parts should not only have the highest transition probability, but should relatively possess a lower mean of defectives which signifies a decrease in defective parts and hence cost effective in terms of PCBs production. Basically, various tables were obtained using the softwares indicated in the research.

Table 2: Descriptive Statistics of daily number of defective parts of PCBs in the Manufacturing industry in Ghana (2009-2014)

Statistic	Value	Statistic	Value
Mean	27.83	Skewness	0.38
SE Mean	1.79	Kurtosis	2.19
Median	8.06	Range	62
Maximum	63	Probability	0.000
Minimum	1	Variance	46.92
Std. Deviation	6.85	Sample	71

Source: Researcher's computation based on sampled data

The results from Table 2 indicate that the mean of the daily number of defective parts of printed circuit boards is 27.83 with a standard error of 1.79 and a standard deviation of 6.85. The maximum number of defective parts of printed circuit boards was recorded as 63 with a minimum number of defective parts of PCBs as 1. Again, the range of number of defective parts over the period was 62. The number of defective parts were centred on a median of 8.06. Also, the data had a positive Skewness of 0.38 indicating that the distribution of the data based on number of defective parts of printed circuit boards is normal.

4.2 Preliminary Analysis

The reliability of a model is defined as the probability that it performs its assigned mission. By Chapman-Kolmogorov Equations, S_0 is the starting state matrix whilst P represents the transition probability matrix. In this study, we assumed the starting state matrix as

$S_{\circ} = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}$, since the last data set was zero; whilst from the data the transition probability matrix was obtained using the software as;

$$P = \begin{bmatrix} \frac{40}{71} & \frac{31}{71} & 0 \\ 0 & \frac{45}{71} & \frac{26}{71} \\ \frac{21}{71} & 0 & \frac{50}{71} \end{bmatrix}$$

To obtain the first state S_1, the starting state matrix is multiplied by the transition probability matrix. Thus, $S_1 = S_{\circ}P$ resulting in;

$$\begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \begin{bmatrix} \frac{40}{71} & \frac{31}{71} & 0 \\ 0 & \frac{45}{71} & \frac{26}{71} \\ \frac{21}{71} & 0 & \frac{50}{71} \end{bmatrix} = \begin{pmatrix} 0.563 & 0.437 & 0 \end{pmatrix}$$

From the results above, it means that the daily number of defective parts of printed circuit boards decreased from state 1 to state 2 but recorded zero number of defective parts in state 3. To obtain S_2, the equation is given as

$S_2 = S_1 P$, which also results in;

$$\begin{pmatrix} 0.563 & 0.437 & 0 \end{pmatrix} \begin{bmatrix} \frac{40}{71} & \frac{31}{71} & 0 \\ 0 & \frac{45}{71} & \frac{26}{71} \\ \frac{21}{71} & 0 & \frac{50}{71} \end{bmatrix} = \begin{pmatrix} 0.317 & 0.523 & 0.160 \end{pmatrix}$$

As illustrated above, it was observed that the daily number of defective parts of printed circuit boards increased in state two but recorded a decreased in states one and three. When the process was continued, it was realised that from the fifth state onwards, the results obtained indicated that the daily number of defective parts of printed circuit boards did not record any significant increase in the first and second states but a high number of 42% defectives in the third state. This means therefore that the daily number of defective parts of PCBs increased serially, thus from each state to the other giving a signal that production in the first and second states is cost effective since these states recorded least number of daily defective parts of PCBs. Additionally, this means that the efficiency of the machine is better during the first and second states of the production process. It also means that workers at these states are not exhausted hence concentration is high which in effect promote quality output of products. On the other hand, the daily number of defective parts of PCBs is high in the third state suggesting that the efficiency of the machine and that of workers is reduced which on the part of the workers could lead to lose of concentration. This in effect reduces quality output of products. The fifth state therefore was obtained as follows;

$S_5 = S_4 P$; resulting in

$$\begin{pmatrix} 0.217 & 0.396 & 0.390 \end{pmatrix} \begin{bmatrix} \frac{40}{71} & \frac{31}{71} & 0 \\ 0 & \frac{45}{71} & \frac{26}{71} \\ \frac{21}{71} & 0 & \frac{50}{71} \end{bmatrix} = \begin{pmatrix} 0.237 & 0.343 & 0.420 \end{pmatrix}$$

The results above suggest that prediction could best be done at the fifth state since the number of defective parts increase serially. Furthermore, the n-step probabilities were determined by subjecting the transition matrix to powers such as below:

$$P^2 = \begin{bmatrix} 0.317 & 0.191 & 0 \\ 0 & 0.402 & 0.134 \\ 0.087 & 0 & 0.496 \end{bmatrix}$$

The above results describe the probability of being the j^{th} state from state i in two steps.

$$P^3 = \begin{bmatrix} 0.179 & 0.083 & 0 \\ 0 & 0.255 & 0.049 \\ 0.026 & 0 & 0.349 \end{bmatrix}$$

Again, this means the probability of being the j^{th} state from state i in three steps. Basically, it means that on the 72nd day of production, we expect the probability of moving from state one to state two to be 31.7%. Similarly, the probability of moving from state two to state three is 49.6%. This means that on the production process, there is a higher probability that the process will remain in state three. This in effect signifies that production is cost effective in states one and two since at these states, least daily number of defective parts of PCBs are recorded as compared to state three which recorded the highest number of defective parts of PCBs. Similar situation was recorded when the transition probability matrix was subjected to the power of three. Figure 2 below illustrates the plot of the daily number of defective parts of printed circuit boards in the manufacturing industry in Ghana spanning from January 2009 to December 2014.

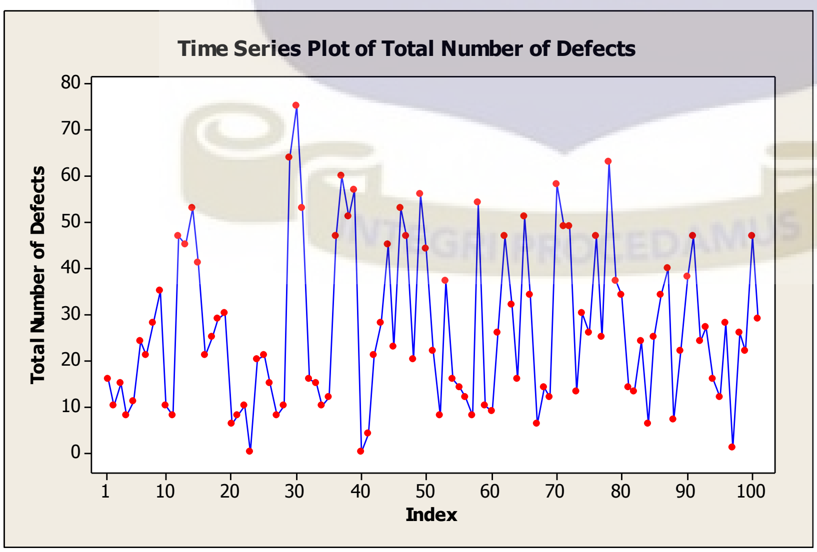

Figure 2: Time series plot of daily number of defective parts of PCBs in Ghana (2009-2014)

It is obvious from Figure 2 that the data displayed stationarity of the daily number of defective parts of printed circuit boards during the period. Furthermore, the trend analysis as illustrated in Figure 3 shows also an increasing trend.

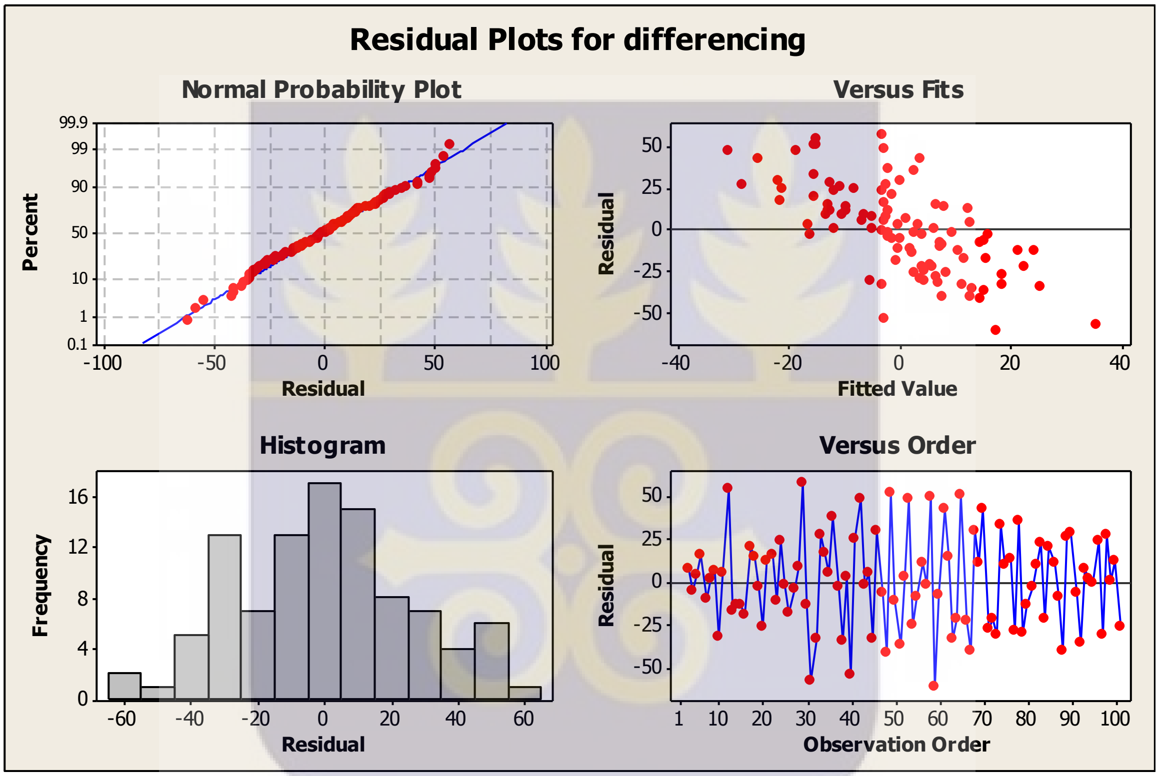

Figure 3: Trend analysis plot for daily number of defective parts of PCBs in the manufacturing industry in Ghana (2009-2014)

To obtain stationarity in time series data, there are several transformations that are normally performed to achieve a stationary data. In this piece of research, the ordinary differencing was used. A plot of the first differencing data was performed and as shown in Figure 3 above is an indication that stationarity was achieved.

To further confirm that the data really attained its stationarity, the Augmented Dickey-Fuller (ADF) test was performed to ascertain the stationarity. The plots in Figure 2 and Figure 3 appear that first ordinary differenced daily number of defective parts of printed circuit boards attained stability in both the mean and variance over time illustrating that there is stationarity. Thus, this is confirmed by performing the Augmented Dickey-Fuller (ADF) test for stationarity on the data series.

Table 3: Augmented Dickey-Fuller (ADF) Test for the daily number of defective parts of PCBs in the Manufacturing industry in Ghana (2009-2014)

	t-Statistic		Probability
Augmented Dickey- Fuller test Statistic	-7.707		0.000
Test Critical Values:	1% =	-3.552	
	5% =	-2.914	
	10% =	-2.592	

Source: Researcher's computation based on sampled data

The test for stationarity was performed on the daily number of defective parts data series and the results in Table 3 above illustrated that the series is stationary. The computed ADF test statistic (-7.707) is smaller than the critical values (-3.552, -2.914, -2.592) at 1%, 5% and 10% respectively. This implies that we can reject the null hypothesis that the first differenced daily number of defective parts of printed circuit boards' data series has a unit root supporting the idea that the series is stationary at 1%, 5% and 10% significant levels.

Additionally, a test for whether there exists serial correlation (autocorrelation) in the daily number of defective parts of printed circuit boards data series was performed to enable us identify the order(s) of the AR, MA and/or the ARMA models. This was done by obtaining the Autocorrelation function (ACF) and Partial Autocorrelation Function (PACF) plots of the daily number of defective parts of printed circuit board data series. These are shown in Figure 4 and Figure 5 respectively.

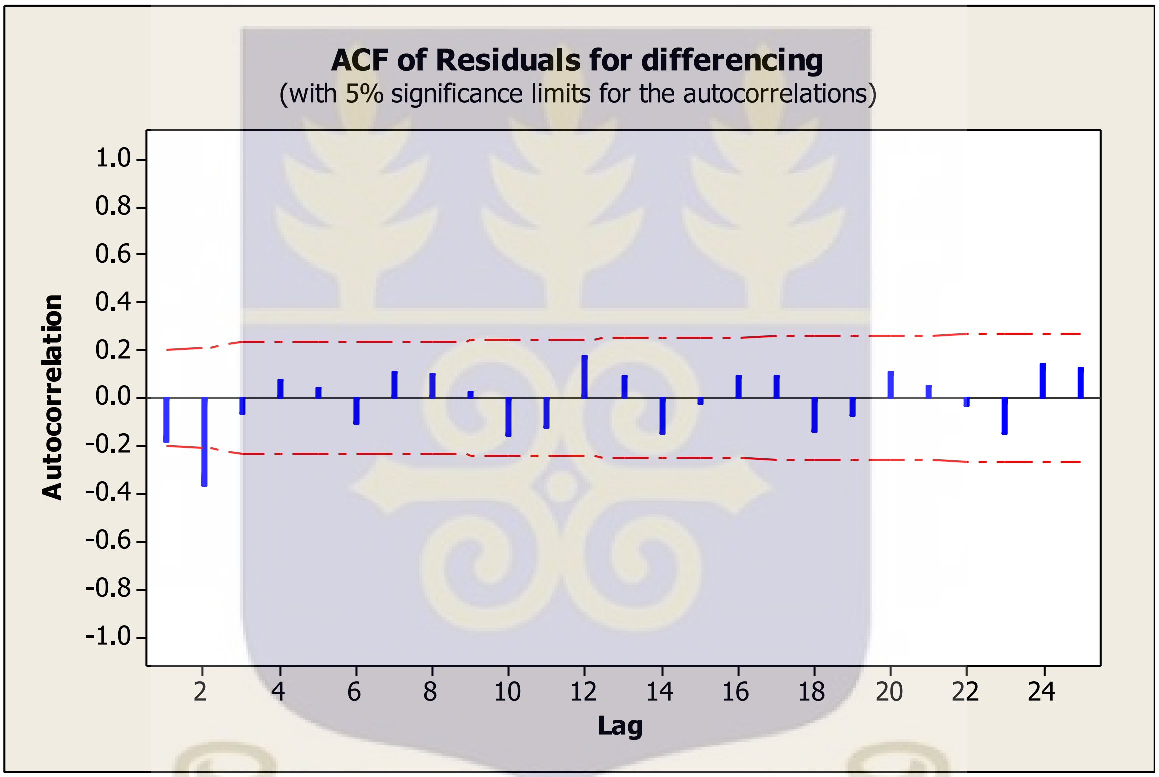

Figure 4 Autocorrelation function (ACF) plots of daily number of defective parts of PCBs in Ghana (2009-2014)

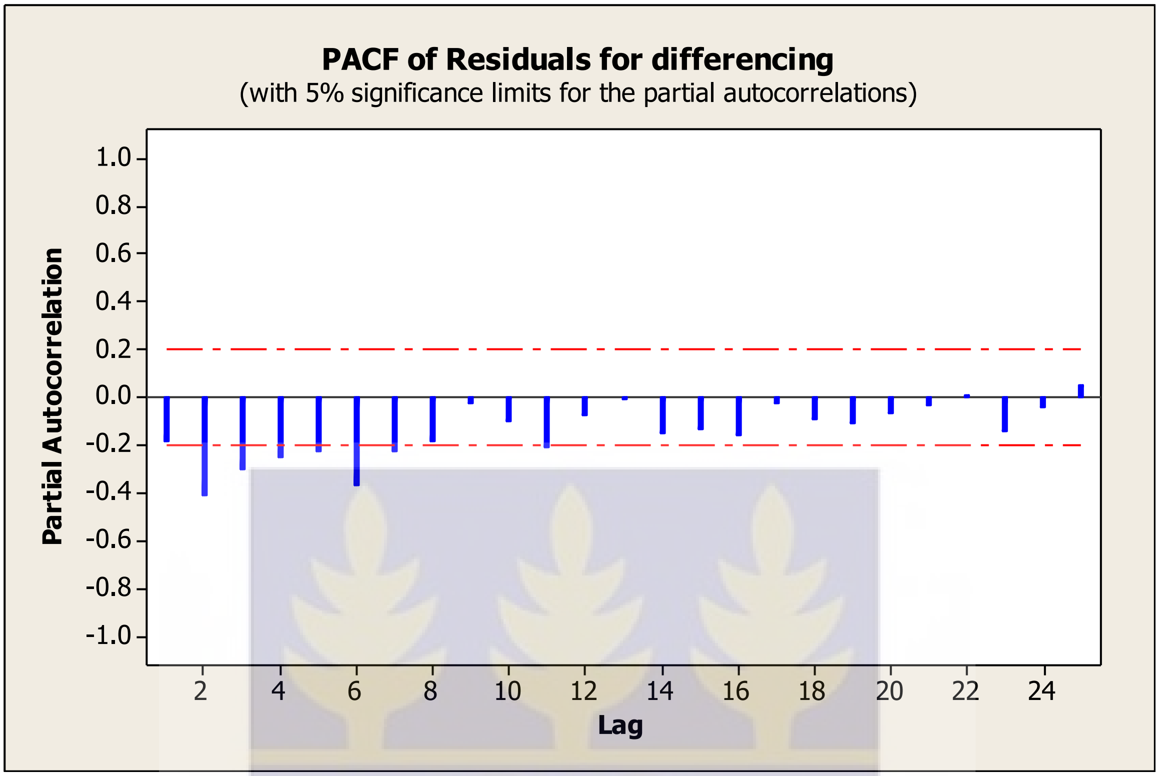

Figure 5: Partial Autocorrelation Function (PACF) plots of daily number of defective Parts of PCBs in Ghana (2009-2014)

From the plots of the ACF and PACF illustrated by Figure 4 and Figure 5, it is clear that there exists a correlation in the daily number of defective parts of PCBs. The ACF plots showed an exponential decay indicating that the daily number of defective parts of PCBs data is stationary. Due to this, there was no need to perform several differencing of the data to obtain stationarity. The ACF plots exhibits an exponential decaying and the PACF plots cut off to zero after the first lag. This means that there is no significant correlation in the daily number of defective parts of printed circuit boards. Therefore, in time series model building, the determination of the order(s) of the model is vital after the series obtained its stationarity. The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots of the series were used to determine the order. From the ACF and PACF plots of the data in

Figure 4 and Figure 5, the ACF tails of at lag 1 whilst the PACF spike at lag 1 suggesting that q=1 and p=1. Hence ARMA (1, 1) is suspected.

Again, in both the non-differencing and the differencing of the data series of the daily number of defective parts of printed circuit boards suggested the following models for the series: ARMA(1,2), ARMA(2,1), ARMA(1,5), ARMA(5,1), ARMA(2,2), ARMA(1,7), ARMA(7,1), AR(1), AR(2) and MA(1).

After the determination of the order of the model and finally the model identification has been carried out, there is the need to fit the suggested models above using the non-differencing and differencing data series of the daily number of defective parts of PCBs. To ascertain the appropriate models to be built after the series is now differenced by the moving average and the smoothing techniques, the least square method and the maximum likelihood method were used in fitting the various models. Generally, since the order(s) determined is usually a suggestion of the order(s) around which ideal model is built, several models of different orders that lie close to the suggested model of ARMA(1,1) were fitted and the most ideal was selected based on the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) significance test. By standard, the criterion is that the smaller the AIC and the BIC values the better the model. The basis of this study is to obtain a model that captures as much variation in the data series as possible. In this work, STATA 12.1 software was used to carry out the modelling process.

4.3 Model fitting and estimation using moving average on the daily number of defective Parts of printed circuit boards (PCBs)

The ARMA (p, q) was fitted using the daily number of defective parts of printed circuit boards to ascertain the true order(s) of p and q respectively. Table 4 illustrates the various suggested models for the ARMA model with their respective fit statistics.

Table 4: Comparison of suggested ARMA (p, q) model with fit statistics

Model	AIC	BIC
ARMA (1, 2)	552.00	563.97
ARMA (2, 1)	550.74	569.06
ARMA (2, 2)	550.93	553.46
ARMA (1, 1)	450.09	400.11
ARMA (1, 5)	556.62	581.21
ARMA (5, 1)	556.12	553.96
ARMA (1, 7)	554.84	574.70
ARMA (7, 1)	553.11	566.49
AR (1)	548.62	547.56
AR (2)	550.11	589.02
MA (1)	548.45	546.21

Source: Researcher's computation based on sampled data

From Table 4, AR (1), MA (1) and ARMA (1, 1) had the least AIC and BIC values. Hence, any of the models could be appropriate.

The model outputs for AR (1), MA (1) and ARMA (1, 1) were used to estimate the most ideal model.

Table 5: Model Output for AR (1)

Variable	Coefficient	Std-Error	t-Stat	Prob
α_1	0.7116	0.1196	0.8425	0.0647

Source: Researcher's computation based on sampled data

From Table 5 above, it showed that the estimate of the P- value of AR (1) model is not statistically significant; hence we failed to reject the null hypothesis.

Table 6: Model Output for MA (1)

Variable	Coefficient	Std-Error	t-Stat	Prob
α_1	0.4193	0.1127	0.5086	0.0170

Source: Researcher's computation based on sampled data

Again, Table 6 above also clearly indicated that the estimate of the P- value of MA (1) model is statistically significant; hence we reject the null hypothesis.

Table 7: Model Output for ARMA (1, 1)

Variable	Coefficient	Std-Error	t-Stat	Prob
$\alpha_\circ$	-0.6431	0.3106	-0.0567	0.0000
α_1	0.4408	0.2017	0.5342	0.0261
β_1	-0.5662	0.7714	-2.0321	0.0421

Source: Researcher's computation based on sampled data

From Table 7 however, the estimate of the P-values of ARMA (1, 1) model is statistically significant. Hence ARMA (1, 1) model has met the general requirement of an ideal model. Clearly, this showed that an AR (1) model is the best model under the moving average smoothing of order 1.Conclusively therefore, from Table 7, the model equation for an AR (1) under the moving average smoothing of order 1 given as ARMA (1, 1) could be expressed as; $X_t = -0.6431 + 0.4408x_{t-1} - 0.5662x_{t-1} + \varepsilon_t$.

where the X_t is the process (total number of defective parts of printed circuit boards in the manufacturing industry in Ghana), x_{t-1} is the lag (interval between two set of points) and ε_t as the error term.

4.4 Diagnostic and Adequacy Checks

Model diagnostic checks were carried out to determine the model adequacy among the most ideal models identified. These checks were done through the analysis of the residuals from the fitted model. Suppose the model fits the data series well, the residuals are expected to be random, independent and identically distributed following the normal distribution. Plots of the residuals such as the normal probability plot and the time plot of residuals were used. The time plot of residuals was used to check for residuals randomness whilst the normal probability plot was used to check for normality.

4.4.1 Model Diagnostic Checks for ARMA (1, 1) Model

Figure 3 illustrates the time plot of the standardized residuals of ARMA (1, 1). This was used to check for residual randomness. From the Figure, it was observed that the results indicated random variation about the mean. The Figure also gives the normal probability plot of the standardized residuals for ARMA (1, 1). The normal probability plot of the standardized residuals was linear. Hence, the linearity of the plot therefore signified that the distribution of the standardized residuals is normal.

4.4.2 Model Diagnostic Checks for AR (1) Model

Again, Figure 3 illustrates the time plot of the standardized residuals for AR (1). The time plot of standardized residuals was done to check whether standardized residuals were random. It is clear that the residuals display random variation about their mean indicating that the residuals are random. The normal probability plot of the standardized residuals for AR (1) was linear. This therefore is an indication that the distribution of the residuals is normal.

4.4.3 Model Diagnostic Checks for MA (1) Model

Figure 3 illustrates the time plot of standardized residuals for the MA (1) model. From the plot, it is obvious that the residuals display random variation about their mean indicating that the residuals are random. Additionally, the normal probability plot of the standardized residuals for MA (1) also was linear, indicating that the distribution of the residuals is normal.

4.5 Selection of Most Ideal Model

From the model diagnostic and adequacy checks for the ARMA (1, 1), AR (1) and MA (1) models aforementioned above, it was observed that all the three models represent the data adequately. Therefore, to choose the most ideal model among these models we considered the AIC and the BIC values of these models. The Table below displays the AIC and BIC values of the models.

Table 8: Selection of Most Appropriate Model from ARMA (1, 1), AR (1) and MA (1)

Model	AIC	BIC
ARMA (1, 1)	450.09	400.11
AR (1)	548.62	547.56
MA (1)	548.45	546.21

Source: Researcher's computation based on sampled data

From Table 8, it is evident that the ARMA (1, 1) had the least AIC and BIC values among the selected models.

4.6 Forecasting Evaluation and Accuracy Selection Method

The forecasting evaluation and accuracy method were applied for the selection of the most ideal model. The models were evaluated in terms of their forecasting capability of future daily number of defective parts of printed circuit boards. The common measures of forecast evaluation which include the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE), the Mean Absolute Percent Error (MAPE) and the Theil's Inequality Coefficient (TIC) were applied. With this, the model with the least values of the error measurements is considered as the most ideal. The Table below therefore displays the error measurement for the selected models.

Table 9: Forecast Performance of Selected Models

Measure	ARMA(1,1)	AR(1)	MA(1)
Root Mean Square Error (RMSE)	0.42	0.63	0.61
Mean Absolute Error (MAE)	0.46	0.51	0.54
Mean Abs. Percent Error (MAPE)	101.74%	126.11%	116.21%
Theil's Inequality Coefficient (TIC)	0.68	0.81	0.65
Rank	**1**	**3**	**2**

Source: Researcher's computation based on sampled data

From Table 9, the results of the forecast performance indicated that the ARMA (1, 1) model performed better than the other models with the least values of error measurements. The ARMA (1, 1) model had the least RMSE value of 0.42 and MAPE value of 101.74% whilst the MA (1) model had only the least TIC value of 0.65.

4.7 Discussion of Results

Only results obtained from the analyses performed in various stages in this research work are discussed. The study was suggested on three states. Thus, the daily number of defective parts of printed circuit boards could increase, decrease or remain the same from one state to the other. The initial state was represented as $S_{\circ}$ whilst the transition probability matrix was represented as P.

Again, the initial state was assumed as $S_{\circ} = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}$, since the last data set was zero and the transition probability matrix calculated from the daily number of defective parts of printed circuit boards' data series as;

$$P = \begin{bmatrix} \frac{40}{71} & \frac{31}{71} & 0 \\ 0 & \frac{45}{71} & \frac{26}{71} \\ \frac{21}{71} & 0 & \frac{50}{71} \end{bmatrix}$$

This was achieved because of the fact that the parameter space which is the daily number of defective parts of printed circuit boards is discrete and that of the state space which is of three states is also discrete. By the application of Chapman-Kolmogorov Equations, the various states of whether the daily number of defective parts of printed circuit boards increase. decrease or remain the same were obtained as;

$$S_1 = S_\circ P$$

Where S_1 is the first state, S_0 is the initial state and P the transition probability matrix. By computation from the daily number of defective parts of printed circuit boards, the first state (S_1) was obtained as

$$S_1 = \begin{pmatrix} 0.563 & 0.437 & 0 \end{pmatrix}$$

From the above results, it is obvious that the daily number of defective parts of printed circuit boards decreased from state one to state two but recorded zero number of defective parts in state three. Also, on the part of the second state (S_2), we applied the expression $S_2 = S_1 P$ which yielded the second state (S_2) as

$$S_2 = \begin{pmatrix} 0.317 & 0.523 & 0.160 \end{pmatrix}$$

The results showed again that the daily number of defective parts of printed circuit boards increased from state one to state two but recorded a decreased in state three. As this process continued, it was realised that at the fifth state (S_5), the daily number of defective parts of printed circuit boards increased serially from each state to the other, thus

$S_5 = \begin{pmatrix} 0.237 & 0.343 & 0.420 \end{pmatrix}$

From the analysis, it is evident that the daily numbers of defective parts of PCBs increase serially from one state to the other giving a signal that production in the first and second states is cost effective since least number of defective parts are recorded in these states. This again means that the performances of the machines are optimal at the first and second states coupled with the fact that operators of the machines (workers) seem not to be exhausted hence concentration is high which in effect leads to quality output. It is therefore prudent for the said industries to produce more at the first and second states since products at these states are quality, thus contain least number of defective parts of PCBs.

The n-step probabilities were determined by subjecting the transition matrix to powers. The results indicated that the probabilities of being the j^{th} state from state i in two steps and three steps respectively were obtained as follows:

$$P^2 = \begin{bmatrix} 0.317 & 0.191 & 0 \\ 0 & 0.402 & 0.134 \\ 0.087 & 0 & 0.496 \end{bmatrix} \text{and } P^3 = \begin{bmatrix} 0.179 & 0.083 & 0 \\ 0 & 0.255 & 0.049 \\ 0.026 & 0 & 0.349 \end{bmatrix}$$

These therefore are the probabilities of daily number of defective parts at each state during the second and third steps respectively.

The daily number of defective parts of printed circuit boards over the period was generally high and centred around 8% with a maximum of about 63% and a minimum of about 1%. From the results, stationarity was attained before and after the data was differenced once. This was further confirmed by the Augmented Dickey-Fuller test for stationarity. The data series actually assumed normality before and after the first difference. This is confirmed also by the histogram plot (with normal curve) of the first difference of the daily number of defective parts of printed circuit boards' data series (Appendix 1A). The ACF and PACF plots of the data were employed to estimate the AR order, AR (P) and MA order, MA (q) around which all the necessary models were built. The plots indicated that AR (1) and MA (1) seemed rather suitable.

The smoothing techniques studied were the linear smoothing method such as the moving average. After the data had been proved to attain stationarity, the ARMA was fitted purposely to determine the order of the moving average and the order of the AR model. From the results obtained, the most ideal ARMA model which adequately represents the data was identified as ARMA (1, 1). This indicates that the moving average smoothing is of order 1 with AR (1) as the best model.

For a fair comparison of the models AR (1), MA (1) and ARMA (1, 1), the results showed that the ARMA (1, 1) had the least AIC and BIC values an indication that it is the most ideal model than AR (1) and MA (1) models. The results from the analyses also revealed that the ARMA (1,1) model was the optimal among the most appropriate models. From the most ideal model selection criterion carried out on the bases of the AIC and BIC values, it was observed that the ARMA (1, 1) model had the least AIC and BIC values of 450.09 and 400.11 respectively making it the best of the bests among the suggested models. Additionally, the forecast experiments conducted on the selected models based on the measurements error to find the optimal model among the three best models AR (1), MA (1) and ARMA (1,1), indicated that ARMA (1,1) had the least RMSE, MAE and MAPE values of 0.42, 0.46 and 101.74% respectively.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.0 Introduction

This chapter gives a summary of the results obtained from the study as well as the conclusions and the recommendations of the areas for future study. Thus, the main results and findings on the performance of the Chapman-Kolmogorov Equations, the two selected data smoothing procedures as well as the four selected models for predictions are presented.

5.1 Summary

Most empirical researches conducted to model stochastic time series data have had to assume the initial state matrix and that of calculating the transition probability matrix from the given data series. Again, empirical studies carried out to model stochastic time series data have had to smoothen the data by applying the traditional procedures such as moving averages smoothing, moving averages and many others. Basically, the modelling of time series was done using the Box-Jenkins methodology until the introduction of the Heteroscedastic models by Bollerslev., (1986). In Ghana and other parts of the World, little researches have been conducted in the area of stochastic time series modelling on the part of the daily number of defective parts of printed circuit boards in the manufacturing industries. However, detecting the daily number of defective parts of printed circuit boards in the manufacturing industries by the use of stochastic time series modelling so far is uncommon.

From the study, it was observed that at the fifth state, the daily number of defective parts of printed circuit boards increased serially from one state to the other, thus $S_5 = (0.237\ 0.343\ \ 0.420)$.

This research was conducted to predict the daily number of defective parts of printed circuit boards in the manufacturing industry using stochastic time series modelling. The research also sought to find the optimal model from the four selected models in the prediction of the daily number of defective parts of printed circuit boards in the manufacturing industry. The modelling was carried out on the non-differenced and the first differenced on the daily number of defective parts of printed circuit boards since both produced the same results from the data spanning from January 2009 to December 2014. The average daily number of defective parts of printed circuit boards for the period under study was 27.83 with a standard error of 1.79 and a standard deviation of 6.85.

The value of the standard deviation signified that there was some amount of variability among the daily number of defective parts of printed circuit boards in the manufacturing industry in Ghana over the period under study. The distribution of the daily number of defective parts of printed circuit boards was positively skewed. The distribution again assumes stationarity and normality after the first difference. A lot of tentative ARMA models were developed based on the different values of the order (p, q) to determine the ideal order for which the moving average smoothing must be carried out. The orders (1), (1) and (1, 1) were developed based on the suggested AR (p) , MA (q) and ARMA (p,q) values from the ACF and PACF plots. The model with the least AIC and BIC values is adjudged the best fit model. AR (1), MA (1) and ARMA (1, 1) had AIC and BIC values of 548.62 and 547.56, 548.45 and 546.21 and 450.09 and 400.11 respectively.

Hence, the ARMA (1, 1) model was considered the best fit since it had the least AIC and BIC values. The forecast experiments conducted on the selected models based on the measurements errors among the three best models AR (1), MA (1) and ARMA (1, 1) evidently showed that ARMA (1,1) had the least RMSE, MAE and MAPE values of 0.42, 0.46 and 101.74% respectively.

5.2 Conclusion

The Markov process provides a reliable approach for successfully analyzing and predicting the daily number of defective parts of PCBs which reflects Markov dependency. The study observes that all states obtained communicate and are aperiodic and ergodic hence possessing limiting distributiuons. Basically, our choice of Markov chain as a tool in effect aids in improving manufacturers' ideas and chances of reducing the daily number of defective parts of PCBs given cost effective in terms of producing PCBs through best choice decisions.

The ARMA (1, 1) model was considered as the best fit model among the ARMA (p, q) models when the data was smoothed by the moving average smoothing procedure. Based on the AIC and BIC values as well as the estimates of the coefficients obtained from the models output, the best models were chosen.

The three best fit models selected were AR (1), MA (1) and ARMA (1, 1). These models were then compared based on their goodness of fit statistics. The goodness of fit statistics that were applied included the AIC and BIC values, the root mean squared error, mean absolute error, mean absolute percent error and the Theil's Inequality Coefficient. Basically, the ARMA (1, 1) model was considered the optimal among the three best fit models since it had the least values for all the goodness of fit statistics. The ARMA (1,1) model under scrutiny is therefore considered superior in predicting the daily number of defective parts of printed circuit boards in the manufacturing industry in Ghana under this research. This is becauuse it has the higher chance of reducing the number of defective parts of PCBs given cost effective in terms of production. It is therefore adjudged the best model and recommended for manufacturers instead of the Markov Chain model.

5.3 Recommendations

On the grounds of the summary and the conclusions drawn from the study, the following recommendations are considered for future research works.

Researchers and all those with special interest in stochastic time series modelling in general and daily number of defective parts of printed circuit boards (PCBs) in the manufacturing industry in particular should consider using the moving average smoothing techniques and differencing to smoothen the data to obtain stationarity and normality before performing the modelling exercise. This is because it produces better models. Model adequacy and checks and forecasting evaluation and accuracy selection criteria should be examined critically before the best model is selected among the numerous models obtained.

Careful study should be done on the Chapman-Kolmogorov equations to be able to precisely predict the state(s) at which production would be cost effective. To determine the n-step probabilities, one should subject the transition matrix to powers. Manufacturers of PCBs should use the ARMA (1,1) model since it is cost effective and also gives least number of defective parts of PCBs.

Again, the Box-Jenkins methodology of modelling time series and that of Heteroscedastic models should both be used in the modelling process when the data is smoothed using the moving average smoothing techniques.

REFERENCES

Kumar, A. (2008). "Computer-vision-based fabric defect detection: A Survey". *IEEE Trans. Industrial Electronics.*, Vol. 55, No. 1, P. 348-363.

Aas, E. A. (1988). 'A closer look at multiple faults and defect levels in Digital Circuits. *International Society for Computers and Their Applications, ISCAS'*, 441-444.

Aas, E. J. (1989). Defect level calculation. *The importance of accurate models for defect distribution and multiple faultvcoverage in low yield situations.IEEE OF ISCA.*

Agrawal, V. D. (1982). Fault Coverage requirements in production testing of LSI Circuits. *IEEE Journal of Solid - state Circuits.*, Vol. SC-17, No. 1, P. 57-61.

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE. Transactions on Automatic Control.*, Vol. 19(6). pages 719-723.

Akgun, B. (2003). Identification of periodic autoregressive moving average models. *Middle East: Akgun.*

Altinkemer, K. K. (2000). Optimization of PCB manufacturing: Integrated modelling and algorithms. *European Journal of operation Research.*, Vol.124(2), P.409-421.

Atuah, S. B. (2014). Statistical alysis of L inear and Non-Linear Smoothing Techniques under the AR Model and the GARCH Model.

Bhat, U. N. (Second Edition). *Elements of Applied Stochastic Process.*

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics.*, Vol, 31., Pages 307-327.

Box, G. E. (1976). *Time Series analysis: Forecasting and control.* Holden: San. Franscisco California.

Bulgak, A. A. (1992). Impact of quality improvement on optimal buffer design and productivty in automatic assembly systems. *Journal of Manufacturing systems.*, Vol. 11, P. 124-136.

Park,C.K. and Harashima, F. (2007). "Development of an automatic stencil inspection system using modified hough transform and fuzzy logic". *IEEE Trans. Industrial Electronics.*, Vol.55, No. 1, P. 604-611.

Stapper, C.H. (1983). Integrated Circuit Yield Statistics. *Proceedings of the IEEE.*

Chan, F. T. (2005). Design of a PCB plant with expert system and simulation approach. *Expert systems with Applications.*, Vol. 28(3), P. 409-423.

Cheng, C. H. (2000). Effect of straight- and U-shaped lines quality:. *IEEE Transactions on Engineering Management.*, Vol. 47, P. 321-334.

Chern, C. A. (1999). Determining a threshold control policy for an imperfect production system with reworkable jobs:. *Naval Research Logistics.*, Vol.46,P. 273-301.

Chevalier, P. B. (1997). Inspection for Circuit Board Assembly. *Management Science*, Vol. 43, No.9, P. 1198-1213.

Chung, Y. F. (1997). Productivity and undesirable outputs: A directional distance function approach. *Journal of Environmental Management.*, Vol. 51(3), P. 229-240.

Clark, A. J. (1995). A Quality Modelling System for Predicting the yield of Assembly and Test Processes in the Printed Circuit Assembly Industry. *Journal of Electronics Manufacturing.*, Vol.5, No.2 P.75-87.

Corsi, F. S. (1993). Defect level as a function of fault coverage and yield. *IEEE- Proceedings of the Third European Test Conference*, 507-508.

de Wet, T. A. (2006). Smoothing Sequence of data by extreme selectors. *ICOTS*, 7.

Deliman, N. C. and Feldman. (1996). Optimazation of process improvement and inspection location for serial manufacturing. *International Journal of production Research.*, Vol. 34, P. 395-405.

Dengiz, B. A. (2000). Computer simulation of a PCB production line:. *Metamodelling approach, International Journal of production Economics.*, Vol. 63(2), P. 195-205.

Doniavi, A. M. (2000). A systems approach to photolithography process optimization in an electronics manufacturing environment. *International Journal of Production Research.*, Vol.38(11), P.2515-2528.

Fare, R. A. (2004). Modelling undesirable factors in efficiency evaluation. *European Journal of Operational Research,* Vol. 157(1), P. 242-245.

Fare, R. G. (1989). Multilateral Productivity comparisons when some outputs are undesirable: A non parameteric approach. *The review of Econonmics and statistics.*, Vol.71(1), P.90-98.

Fare, R. G. (1996). An activity analysis model of the environmental performance of firms-application to fossil-fuel-fired electric utilities. *Ecological Economics.*, Vol. 18(2), P. 161-175.

Finkelshtein, A. H.-G. (2005). Economic optimatization of off-line inspection in a process subject to failure and recovery. *IIE Transactions.*, Vol. 37, P. 995-1009.

Fululedzani, L. L. (2011). *ARIMA forecast of the number of Beneficiaries of social security grants in South Africa.* University of South Africa.

Gardiner, J. (1987). *"Detecting Small Shifts in Quality Levels in a Near-Zero Defect Environment for Integrating Circuits.".* Seatle, Washington: University of Washington.

Gebus, S. (2000). Process Control Tool for a Production Line at Nokia. *Control Engineering Laboratory*, 27, 42-43.

Ghana, B. O. (2009). Bank of Ghana Monetary Policy Report. *Business News*, vol.3 No.2.

Giachetti, R. E. (2001 and 2007). An object-oriented information model for manufacturability analysis of printed circuit board fabrication:. *Computers in industry.*, Vol. 45(2), P. 177-196.

Gooijier, J. A. (1985). Methods for determining the order of an autoregressive-moving average process. *A survey Int.Statits*, Rev.53, pages 301-329.

Grunow, M. G. (2000). Simulation-based Peformance analysis and optimization of electronics assembly equipment:. *International Journal of Production Research.*, Vol. 38(17), P. 4247-4259.

Stapper, C. H. (1989). Fault Simulation programs for integrated circuit yield estimates. *IBM Journal of Research and Development.*, Vol.33, No.6, P.647-652.

Hamilton, J. D. (1982). *Time Series Analysis.* Priceton University Press.

Han, M. S. (1998). Perfomance analysis of serial production lines with quality inspection machines. *International Journal of system science.*, Vol. 29, P. 939-951.

Hawkins, D. M. (1998). *Cummulative Sum Charts and Charting for Quality Improvement.* New York, NY.: Springer-Verlag.

Haynes, K. E.-S. (1997). Toward a pollution abatement monitoring policy: Measurements, model mechanics and data requirements. *The Environmental Professional.*, Vol.16, P. 292-303.

Henshaw, B. (1989). How much fault coverage is enough? *1989 International Test Conference*, 939-940.

Hiller, F. S. (1990). *Introduction to Operations Research. 5th edition.* New York, NY.: McGraw-Hill.

Holzmann, M. T. (2004). "Automatic Optical inspection of circuit assemblies in a high mix/low volume environment". *Procs of the Surface Mount Technology Association.*

Hroundas, G. (1990). New directions in electrical testing of PCBs for the 1990's. *Circuit World 1990.*, Vol.16, No.4,P.29-31.

Inman, R. B. (2003). Designing production system for quality: Research opportunities from automotive industry perspective. *International Journal of production Research*, Vol. 41, P. 1953-1971.

Jackson, E. J. (1972). "All Court Distribution Are Not Alike". *Journal of Quality Technology*, Vol.4, No.2. 86-92.

Jacobs, D. A. and Meerkov. (1991). Assympototically reliable serial production lines with quality control systems, Computers Mathematics and Applications . 85-90.

Jacobs, D. A. (1998). Asympototically reliable serial production lines with quality control systems. *Computers, Mathematics and Applications.*, Vol.21, P. 85-90.

Johnson, N. L. (1969). *Discrete Distribution.* Boston, MA.: Houghton Mifflin Company.

Joseph, L. A. (1990). Modelling and Analysis of a New Product Development in Electronics Sub-Assembly Manufacturing. *Proceedings of the 1990 Winter Simulation Conference.*, 899-903.

Khouja, M. R. (1995). Optimal robot operation and selection using quality and output trade-off. *International Journal of Advanced Manufacturing Technology.*, Vol. 10, P.342-355.

Kim, J. A. (2005). Integrated quality and quantity model of a production line. *Spectrum*, Vol.27, P. 287-314.

Wadsack, R. L.S (1981). VLSI: How much fault coverage is enough? *Proceedings of the International Test Conference.*, 547-554.

Li, J. A. (2007). Quality evaluation in flexible manufacturing systems:. *A Markovian approach. Mathematical Problems in Engineering*, Article 57128.

Li, J. A. (2007). Manufacturing system design to improve quality buy rate. *IEEE Transaction on Automation Science and Engineering.*, Vol.4, P. 75-79.

Li, Y. R. (1993). Design Factors and their effect on PCB assembly yield- statistical and neural network predictive models. *IEEE-CHMT. International Electronics Manufacturing Technology Symposium.*, 353-361.

Limited, N. S. (2014). Find Industry and Manufacturing expertise in Ghana. *Nexus Strategic Partnership Limited.*

Linn, R. (1996). Component placement error model for surface mounted technology. *5th Industrial Engineering Research Conference Proceedings.*, 91-96.

Liu, J. J. (1996). Optimal lot-sizing with in an imperfect production system with homogeneous reworkable jobs: . *European Journal of operational Research.*, Vol. 91, P.517-527.

Liu, W. B. (1999). DEA models via goal programming. In: Westerman, G. (ED.), Data Envelopment Analysis in the public and private sector.,. *Deutscher Universtats-verlag.*

Lovell, C. A. (1995). Units invariant and translation invariant DEA models. *Operations Research Letters.*, Vol. 18(3), P. 147-151.

Luruli, F. L. (2011). *ARIMA Forecasts of the Numbe rof Beneficiaries of Social Security Grants in South Africa.* Pretoria: (Unpublished thesis) University of South Africa.

Martiz, J. (1970). *Empirical Bayes Methods.* London, UK.: Methuen.

Matanachai, S. A. (2001). Balancing mixed model assembly lines to reduce work overload. *IIE Transactions.*, Vol. 33, P. 29-42.

Millman, S. D. (1994). Improving Quality:Yield versus Test Coverage. *Journal of Electronics testing 1994.*, Vol.5, P. 253-261.

Montgomery, D. C. (1996). *Introduction to Statistical Quality Control. 3rd edition.* New York, NY.: John Wiley and Sons.

Montgomery, D. J. (2008.). Introduction to time series forecast. *Solar Energy*, Vol.83, pages 342-349.

Moyer, L. K. (1997). An efficient assembly sequencing heuristic for printed circuit boards configurations:. *Journal of Electronics Manufacturing.*, Vol. 7, No.2, P.143-160.

Nakahara, H. (1999). PCB Output 1998. *Printed Circuit Fabrication*.

Narahari, Y. A. (1994). Transient analysis of manufacturing systems performance. *IEEE Transactions on Robotics and Automation.*, Vol.10, P.230-244.

Narahari, Y. A. (1996). Modelling reentrant manufacturing system with inspection stations. *Journal of manufacturing systems.*, Vol. 15, P. 367-378.

Nexus Strategic Partnership Limited, (2014). Find Industry and Manufacturing expertise in Ghana. *Nexus Strategic Partnership Limited*.

Neyman, J. (1939). "On a new Class of "Contagious" Distribution.Applicable in Entomology and Bacteriology". *Annals of Mathematical Statistics.*, Vol.10.35-57.

Orsejo, S. (1998). Test Strategies for PCB Manufacturing:. *Future Circuits International.*, No.3, P. 235-242.

Kripper,P.and Beer,D. (2004).'AOI testing position in comparison'. *Circuits Assembly.*,26-32.

Parker, E. S. (1989). A Statistical Model for delay-fault testing. *IEEE Design and Test of Computers.*, Feb., P. 45-55.

Parker, E. S. (1992). The total delay fault model and statistical delay fault coverage. *IEEE Transaction on Computers*, Vol.41, No.6, P. 688-698.

Patil, G. C. (1968). *A Dictionary and Bibliography of Discrete Distribution.* New York, NY.: Hafner publishing Company.

Pendse, R. D. (1994). Demountable tab-a new path for tab technology. *IEEE Transactions on components packaging and Manufacturing Technology Part A-Advanced Packaging.*, Vol. 17, P. 527-536.

Pitman, R. W. (1983). Multilateral Productivity comparisons with undesirable outputs. *The Economic Journal.*, Vol. 93(372), P. 883-891.

Richard, S. C. (1999). 'The complete solder paste printing processes'. *Surface Mount Technology*, Vol.13, P. 6-9.

Ryan, A. A. (2007). Manufacturing an environmentally friendly PCB using existing industrial processes and equipment:. *Robotics and computer-integrated manufacturing.*, Vol.23(6), P. 720-726.

Scheel, H. (2001). Undesirable Outputs in efficiency evaluation. *European Journal of Operational Research.*, Vol. 132(2), P. 400-410.

Schwartz, G. E. (1978). Estimating the Dimension of a Model. *Annals of Statistics*, Vol. 6(2), pages 461-464.

Seiford, L. M. (2002). Modelling undesirable factors in efficiency evaluation. *European Journal of operational Research.*, Vol. 142(1), P. 16-20.

Sheaffer, R. A. (1976). "The Negative Binomial Model for Counts In Units of Varying Size.". *Journal of Quality Technology.*, Vol.8, No.3 pp158-163.

Sherbrooke, C. C. (1966). *"Discrete Compound Poisson Process and Tables of The Geometric Poisson Distribution."*. Santa Monica, CA.: Memorandum RM-4831-PR,The Rand Corporation.

Son, Y. K. and Park (1987). Economic Measure of Productivity, quality and flexible in advanced Manufacturing Systems. *Journal of Manufacturing systems.*,Vol.6, P.193-207.

Sousa, J. T. (1996). Realistic Fault Extraction for Boards. *IEEE Computer Society Press*, 612.

Stapper, C. H. (1985). "The effects of Wafer to Wafer Defect Density Variations on Integrated Circuit Defect and Fault Distribution". *IBM Journal of Research Development*, Vol.29. No.1.87-97.

Talke, I. S. (2003). *Modelling Volatility in Time Series Data.* Kwa- Zulu Natal: Unpublished Thesis.

Tsai, C. Y. (2005). A case-based reasoning system for PCB defect prediction. *Expert Systems with Applications.*, Vol. 28(4), P. 813-822.

Tsay, R. S. (2002). *Analysis of Financial Time Series.* New York: John Wiley and Sons Inc.

Tyteca, D. (1997). Linear Programming models for the measurement of environment performance of firms-concepts and empirical results. *Journal of Productivity Analysis.*, Vol. 8(2), P. 183-197.

Urban, T. L. (1998). Analysis of production systems when run length influences product quality. *International Journal of production Research.*, Vol. 36, P. 3085-3094.

Velleman, P. F. (1981). *Applications, basics, and computing of exploratory data analysis.* Duxbury Press.

Viswanadham, N. S. (1996). Inspection allocation in manufacturing systems using stochastic techniques. *IEEE Transactions on systems, man and cybernetics- Part A: Systems and Humans.*, Vol.36, P.222-230. .

Wadsack, R.(1978). Fault Coverage in digital integrated circuits. *Bell System Technological Journal.*, Vol. 57, No. May-June, 1475-1488.

Wadsack, R. L. (1981). How much fault coverage is enough? *Proceedings of the International Test Conference.*, 453-470.

Wang, R. H. (2000). Measuring production and marketing efficiency using grey relation analysis and data envelopment analysis. *International Journal of production Research.*, Vol.48(1), P. 183-199.

Williams, T. W. (1985). Test length in a Self-testing environment. *IEEE Design and Test.*, April, P. 59-63.

Williams, T. W. and Brown, N. C. (1981). Defect Level as a Function of Fault Coverage. *IEEE: Transactions on Computers*, Vol.C-30, No.12, P.987-988.

Wu, C. (2007). Multifactor performance measure model with an application to printed circuit board industry. *International Journal of Business Performance Management.*, Vol.9(1), 43-57.

Takagi, Y. (1990). "Visual inspection machine for solder joints using tiered illumination". *In SPIE Machine Vision Systems Integration in Industry.*, Vol. 1386, P. 21-29.

APPENDIX

Figure 1A: Histogram (with Normal curve) of the first difference of daily Number of defective parts of PCBs in the manufacturing indutry in Ghana (2009-2014)