Prince Osei Affi Journal of Engineering Research and Application *www.ijera.com*
ISSN : 2248-9622 Vol. 9,Issue 2 (Series -II) Feb 2019, pp 09-17

RESEARCH ARTICLE OPEN ACCESS

Modelling Defective Parts of Printed Circuit Boards in the Manufacturing Industries in Ghana

Affi .O. Prince[1], Mensah .A. Isaac [2], Angkyiire David[1], Lartey .N. Lloyd[1]
[1]*Department of Statistics, University of Ghana, Legon, Ghana*
[2]*Faculty Science, Jiangsu University, Zhenjiang, Jiangsu, China*
Corresponding Author: Prince Osei Affi

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 PCBs in a production process in the manufacturing industries in Ghana. The data points were assumed to exhibit Markov dependency with respective state transition probabilities matrices following the identified state space (i.e. increase, decrease or remain the same). We established a methodology for determining whether the daily number of defective parts increase, decrease or remained the same. A criterion for identifying the state(s) in which production of PCBs is cost effective based on least transition probabilities was also applied. 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 PCBs is recorded. The n-step probabilities were also determined by subjecting the transition matrix to powers. Autoregressive and moving average method was also applied. Stationarity 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
Keywords: Time series, Defective, Printed circuit boards, Transition matrix, Markov dependency

---

Date Of Submission: 24-01-2019 Date Of Acceptance:08-02-2019

I. INTRODUCTION

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 [1]. 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 [2]. 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 [2].

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 [3]. 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 aim of this study 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.

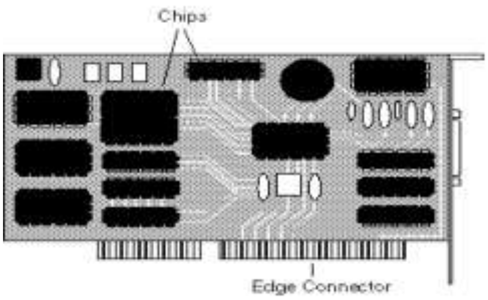

Figure 1: Sample a PCB

II. METHODOLOGY

2.1 Data Collection and Description

Secondary data is used in this study. The sample data which consist of daily productions of printed circuit boards is so large that the average production of PCBs per day will be used. Based on this, seventy-one (71) observations of the daily number of defective parts of printed circuit boards in the electronic 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.

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

$$P\left(X_{(t)} \le x \mid X_{(tn)} = x_n, X_{(tn-1)} = X_{n-1}, \ldots, X_{(t0)} = x_0\right)$$
$$= P\left[X_{(t)} \le x \mid X_{(tn)} = x_n\right] = F\left[X_n, x; t_n, t\right] \quad (1)$$

From equation (1), the relation is as follows;

$$F(X_n, x; t_n, t) = \int_{y \in S} F(y, x; \tau, t)\, \partial F(X_n, y; t_n, \tau) \quad (2)$$

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

$$P\left(X_n = j \mid X_{n_1} = i, X_{n_2} = i_2, \ldots X_{nk} = ik\right) \quad (3)$$
$$= P\left(X_n = j \mid X_{n_1} = i_1\right) = P_{ij}^{(nk,n)}$$

A stochastic process with discrete space and parameter space which displays Markov dependency as in equation (3) 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 $X_1, X_2, \ldots, X_n$ 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:

$$P\{X_{n+1} = x / X_1 = x_1, X_2 = x_2 \ldots X_n = x_n\} \quad (4)$$

$$P\{X_{n+1} = x / X_n = x_n\} \quad (5)$$

$$P\{X_1 = x_1 \ldots X_n = x_n\} > 0 \quad (6)$$

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.

2.3 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:

$$P_{ij}^{(nk,n)} = P(X_n = j \mid X_{nk} = i)$$
$$= \sum_{m \in s} P(X_n = j \mid X_r = m) P(X_r = m \mid X_{nk} = i)$$
$$= \sum_{m \in s} P_{ij}^{(nk,r)} P_{mj}^{(r,n)} \quad (7)$$

Equations (1) and (7) are termed as the Chapman-Kolmogorov equations for the process.

2.4 Assumptions of Markov Chains

We employed the following assuming using the Markov chain:

- 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,……

2.5 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

$$P_{ij} = P\{X_n = j / X_{n-1} = i\} \quad (8)$$

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.

2.6 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, the transition probabilities P_{ij} can be represented in the form of a transition matrix P such as:

$$P = \begin{pmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{pmatrix} \quad (9)$$

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

In the elements of equation (9), 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, 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.

2.7 The n-Step Transition Probability Matrix

Let *P* be the transition probability matrix of a Markov Chain $\{X_n, n = 0, 1, 2, \ldots\}$ 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 $P = \|P_{ij}\| = \begin{bmatrix} P_{00} & P_{01} & P_{02}\dots \\ P_{10} & P_{11} & P_{12}\dots \\ . & . & . \end{bmatrix}$ and

$$P^n = \|P_{ij}^{(n)}\| = \begin{bmatrix} P_{00}^{(n)} & P_{01}^{(n)} & P_{02}^{(n)}\dots \\ P_{10}^{(n)} & P_{11}^{n} & P_{12}^{(n)}\dots \\ . & . & . \end{bmatrix} \quad (10)$$

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

The above illustrations can obviously be indicated from the Chapman-Kolmogorov equation (7) 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 $P_{ij}^{(r)}$ (r = 3, 4, 5, ….n) is the $(i, j)^{th}$ of P^r then by the Kolmogorov equation, then $P_{ij}^{(r+1)} = \sum_{k \in s} P_{ik}^{(r)} P_{kj}$

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, ……. 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.

2.8 Limiting distribution of a Markov chain

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

$$\lim_{t \to \infty} P^t = \pi = \begin{bmatrix} \sigma \\ \sigma \\ . \\ . \\ \sigma \end{bmatrix} \quad (11)$$

Where

$\sigma = [\sigma_1, \sigma_2, \dots \sigma_m]$ 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.

2.9 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
P = The transition probability matrix, then

$S_1 = S_0P$
$S_2 = S_1P$
.
.
$S_n = S_{n-1}P$

2.10 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 [4]. Discrete time series data is considered in this study since both the parameter space and the state space of the study 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.

A lot of models are used for time series data analysis. However, these models are classified as the linear and the non-linear. Basically, in this study we will employ linear models and some of 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.

But among the linear models above we would make use of the AR model, MA model and the ARMA model from the list above. Augmented Dickey- Fuller (ADF) test for stationarity was carried out to confirm the stationarity of the series. A series is considered to be stationarity if the mean and the auto covariance of the series do not depend on time. 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. There is always the need to induce stationarity in the data before the analysis is conducted. Differencing is a special type of filtering, which is particularly useful for removing trend and/or seasonality in a series. Smoothing technique was used reveals more clearly the underlying trend, seasonal and cyclic components.

2.11 Moving Average (MA) Model

Moving averages is one of the most common 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);

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

2.12 Autoregressive (AR) Model

The autoregressive (AR) model uses past values of the dependent variable to explain the current value. According to Hamilton [5], AR model is the most ordinary autoregressive models used in time series analysis. Let $\{\mathcal{E}_t \,/\, t \epsilon\, T\}$ be a white noise process with mean zero and variance δ^2. A process $\{X_t;\, t \in T\}$is said to be an autoregressive time series of order p (denoted AR (p)) if

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

$$x_t = \alpha_\circ + \sum_{i=1}^{p} \alpha_i x_{t-i} + \varepsilon_t \qquad (12)$$

Where α_o is a constant and α_i are parameters of the model. From (12) and using polynomial notation, AR (p) can be written as;

$$\alpha(B)x_t = \alpha_\circ + \varepsilon_t \qquad (13)$$

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 αs and θs. Hence, equation (13) could be written as

$$\begin{aligned} x_t &= (\alpha_\circ + \varepsilon_t)/\alpha(B) \\ &= (\alpha_\circ + \varepsilon_t)\theta(B) \\ &= \mu + \varepsilon_t \theta(B) \\ &= \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \ldots \end{aligned} \qquad (14)$$

where μ is a constant and can be calculated by

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

It follows that E (Xt) = μ and the covariance function is

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

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 Xt 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 $\mathcal{E}_t$ is a series of random errors each with zero mean and constant variance δ^2 according to Box [6].

2.13 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 [7]. The AR (p) model for a given time series, X_t is given by

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + \varepsilon_t$$

Using the backshift operator B and with $a_t = 0$

$$\alpha(\beta)\, x_t = \varepsilon_t \qquad (15)$$

Where

$$\alpha(\beta) = 1 - \alpha_1\beta - \alpha_2\beta^2 - \; \ldots\ldots - \alpha_p\beta^p$$

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* , α(β) = 0 are all greater than 1 [8]. 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),

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

$$\rho_k = \frac{\gamma(t+k,\ t)}{\sqrt{\gamma(t+k, t+k)\,\gamma(t,t)}} \,, k = 0,1,2 \ldots (16)$$

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) [9]. ρ_k measures the correlation between x_t and x_{t+k} . The estimate of ρ_k is given by:

$$\hat{\rho} = \frac{\hat{\gamma}(t+k, t)}{\sqrt{\hat{\gamma}(t+k, t+k)\,\hat{\gamma}(t,t)}} \,, k = 0,1,2\ldots \qquad (17)$$

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

The maximum likelihood method is used in estimating AR model parameters [6].

2.14 Forecasting with the AR model

The main objective of developing a time series model is for forecasting future values of the series. In this paper, modelling and forecasting of the total number of defective parts of printed circuit boards in the manufacturing industry was investigated. Given the time series $X_1, X_2, \ldots, X_t$, forecasting is done to predict the value of $x_{t+k}\ (k=1,2,3,\ldots)$, the series that will be observed at time x_{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;

$$E\left[\left(x_{t+k} - \hat{x}_t(k)\right)^2 \middle| x_1, x_2, \ldots, x_t\right] \quad (18)$$

Therefore when we differentiate equation (18) with respect to $\hat{x}_t(k)$,then equating to zero and solving for $\hat{x}_t(k)$ gives;

$$\hat{x}_t(k) = E\left[x_{t+k} \middle| x_1, x_2, \ldots, x_t\right] \quad (19)$$

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

$$X_{t+k} = \alpha_0 + \alpha_1 X_{t+k-1} + \alpha_2 X_{t+k-2} + \ldots + \alpha_p X_{t+k-p} + \varepsilon_{t+k} \quad (20)$$

We employed Akaike Information criterion (AIC) and Bayesian Information criterion (BIC) in the selection of the best models [10].

III. RESULTS

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 $S_0 = (1,0,0)$, since the last data set was zero; whilst from the data the transition probability matrix is obtained 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 S1, the starting state matrix is multiplied by the transition probability matrix. Thus $S_1 = S_0 P$ resulting in;

$$(1 \quad 0 \quad 0)\begin{vmatrix} \frac{40}{71} & \frac{31}{71} & 0 \\ 0 & \frac{45}{71} & \frac{26}{71} \\ \frac{21}{71} & 0 & \frac{50}{71} \end{vmatrix} = (0.563 \quad 0.437 \quad 0)$$

It was observed 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. It was realized that from the fifth state, $S_5 = (0.237 \quad 0.343 \quad 0.420)$

Indicating that the daily number of defective parts of PCBs 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 that production in the first and second states is cost effective.

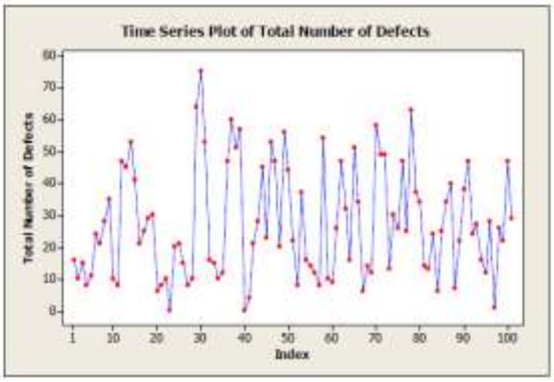

Figure 2: Time series plot of daily number of defective parts of PCBs in Ghana (2009- 2014).

Fig 2 above 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. It is obvious from Fig 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 Fig 3 shows also an increasing trend.

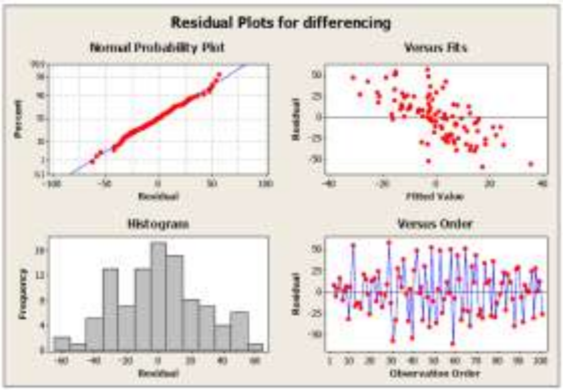

Figure 3: Trend analysis plot for daily number of defective parts of PCBs.

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 Fig 3 above is an indication that stationarity was achieved.

Table 2: Augmented Dickey-Fuller (ADF) Test for the daily number of Defective parts of PCBs in the Manufacturing Industry in Ghana

	t-Statistic	Probability
Augmented Dickey-Fuller	-7.707	0.000

The test for stationarity was performed on the daily number of defective parts data series and the results in Table 2 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 Fig 4 and Fig 5 respectively below

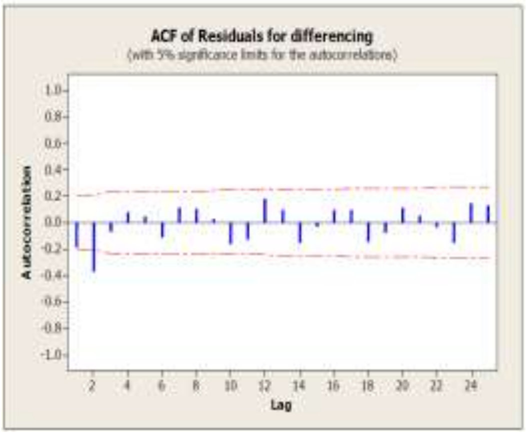

Figure 4: Autocorrelation function (ACF) plots

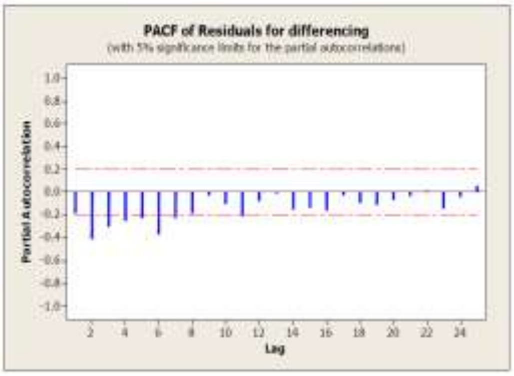

Figure 5: Partial Autocorrelation Function (PACF) plots

From the plots of the ACF and PACF illustrated by Fig 4 and Fig 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 thatcaptures as much variation in the data series as possible. In this work, STATA 12.1 software was used to carry out the modelling process.

3.1 Model fitting

The three models identified are:

i. AR(1):
ii. MA(1): $X_t = \alpha_1 x_{t-1} + \varepsilon_t$
iii. ARMA(1,1): $X_t = \alpha_1 x_{t-1} + \varepsilon_t$

$$X_t = \alpha_\circ + \alpha_1 x_{t-1} + \beta_1 x_{t-1} + \varepsilon_t$$

Table 3: Selection of Most Appropriate Model from ARMA (1, 1), AR (1) and MA (1)

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

From the Table 3 above it was observed that the ARMA(1,1) model was the model with the smallest value of both the AIC and BIC values hence was the most appropriate model identified.

Table 4: 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	126.11 %	0.65
Rank	1	3	2

From the Table 4 above we ranked the three models we considered suitable for the forecasting for the defective PCBs and from the ranking in the Table 3 the ARMA(1,1) model was ranked the best in terms of forecasting performance.

2.2 Model output for ARMA (1,1)

The model output for ARMA(1,1) was obtained as:

$$X_t = -0.6431 + 0.4408 x_{t-1} - 0.5662 x_{t-1} + \varepsilon_t$$

where 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 $\mathcal{E}_t$ as the error term.

IV. 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 observed that there exists limiting distribution. Markov chain as a tool aids in improving manufacturers' ideas and chances of reducing the daily number of defective parts of PCBs through best choice decisions. The ARMA (1, 1) model was considered as the best fit model among the ARMA (p, q) models.

The goodness of fit statistics 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. The AR(1), MA(1) and the Chapman-Kolmogorov Equations are equally good for modelling the defective parts of PCBs.

ACKNOWLEDGEMENTS

We say a very big thank you to Maha Yen Samuel for his support.

REFERENCES

[1]. Ghana, B. O. (2009). Bank of Ghana Monetary Policy Report. Business News, vol.3 No.2.
[2]. Nexus. (2014). Find Industry and Manufacturing expertise in Ghana. Nexus Strategic Partnership Limited.
[3]. Gebus, S. (2000). Process Control Tool for a Production Line at Nokia. Control Engineering Laboratory, 27, 42-43.

[4]. Akgun, B. (2003). Identification of periodic autoregressive moving average models. Middle East: Akgun.
[5]. Hamilton, J. D. (1982). Time Series Analysis. Priceton University Press.
[6]. Box, G. E. (1976). Time Series analysis: Forecasting and control. Holden: San. Franscisco California.
[7]. Gooijier, J. A. (1985). Methods for determining the order of an autoregressive-moving average process. A survey Int.Statits, Rev.53, pages 301-329.
[8]. Fululedzani, L. L. (2011). ARIMA forecast of the number of Beneficiaries of social security grants in South Africa. University of South Africa.
[9]. Montgomery, D. J. (2008.). Introduction to time series forecast. Solar Energy, Vol.83, pages 342-349.
[10]. Schwartz, G. E. (1978). Estimating the Dimension of a Model. Annals of Statistics, Vol. 6(2), pages 461-464.

Prince Osei Affi" Modelling Defective Parts of Printed Circuit Boards in the Manufacturing Industries in Ghana" International Journal of Engineering Research and Applications (IJERA), vol. 9, no.2, 2019, pp 09-17