International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 4, Issue 6, November - December 2015 ISSN 2278-6856

USING TIME SERIES TO PREDICT STOCK PRICES ON THE GHANASTOCK EXCHANGE

Emmanuel Kwame Mensah[1], Joseph Kweku Arthur[2]

[1]Faculty of engineering, Ghana Technology University College
PMB 100, Tesano- Accra, Ghana

[2]Faculty of engineering, Ghana Technology University College
PMB 100, Tesano- Accra, Ghana

Abstract

Stock market prediction is the act of trying to determine the future value of a company stock or other financial instruments traded on an exchange. The successful prediction of a stock's future price could yield significant profit. The efficient-market hypothesis suggests that stock price movements are governed by the random walk hypothesis and thus are inherently unpredictable. In this paper, time series analysis was used to develop a mathematical model to predict the stock prices of stocks on the Ghana Stock Exchange. At the end of the study, the future stock prices of the listed companies were able to be determined through the developed mathematical model. Past data of UNIL, EBG and BOPP was fix into the model to predict the stock price for the next three months.

Key words: Ghana Stock Exchange, time series, ARIMA, Box-Jenkins Method.

1.INTRODUCTION

Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange. The successful prediction of a stock's future price could yield significant profit. The efficient-market hypothesis suggests that stock price movements are governed by the random walk hypothesis and thus are inherently unpredictable [1]. Others disagree and those with this viewpoint possess a myriad of methods and technologies, which purportedly allow them to gain future price information. Forecasting can estimate the statement of events according to the historical data and it is considerably important in many disciplines[2]. At present, time series models have been utilized to solve forecasting problems in various domains

Early manifestations of stock exchanges existed in the 13th and 14th centuries, first in the larger cities of northern Italy, later - after the discovery of the sea route to India - in trading towns along the coasts of Holland and Flanders. In time, the informal gatherings of merchants developed into actual stock exchanges with lively trading in goods. In Antwerp, spot and futures transactions were concluded at a very early date, i.e. in the early 16th century. These transactions were initially subject to unwritten trading customs; later on, they were governed by formal rules issued by trade association. In the early 17th century, Antwerp was replaced by Amsterdam as the home of the most important commodities exchange. When Amsterdam's exchange introduced trading in the shares of the Dutch East India Company, it became the prototype for today's securities exchanges. From the first half of the 16th century, exchanges were founded in all major trading cities, including London, Paris and the trading cities of Germany [3].Financial investors of today are facing this problem of trading, as they do not properly understand as to which stocks to buy or to sell in order to get optimum profits. Although Stock Information can be acquired by reading the business pages of the newspapers, listening or watching business news on local FM stations and on the TV,analysing all these information individually or manually is tremendously difficult.

Currency fluctuation risk; the global economic slowdown in world growth may affect Ghana exports of agricultural products, minerals and hydrocarbons. Ghana’s dependence on natural resource exports has made many countries vulnerable to commodity price shocks that are outside their control. Sudden increases in export revenues or import costs can cause currency instability and budget uncertainty[4]. Furthermore, there is strong evidence that currency depreciation has negative effect on the performance of the Ghana stock market. As such, this paper looks at using time series to predict accurate performance of stocks on the Ghana Stock Exchange (GSE). This will aid:

- Investorsto know best times to buy or sell stock.
- In analysing past behaviour of stocks
- forecasting stocks on the Ghana stock exchange

2.LITERATURE REVIEW

DETERMINANTS OF THE PRICE OF STOCK

When a company goes public through an initial public offering (IPO), an investment bank evaluates the company's current and projected performance and health to determine the value of the IPO for the business. The bank can do this by comparing the company with the IPO of another similar company, or by calculating the net present value of the firm. The company and the investment bank will meet with investors to help determine the best IPO price through a series of road shows. Finally, after the valuation and road shows, the firm must meet with the exchange, which will determine if the IPO price is fair.[5]

2.1.1DEMAND AND SUPPLY, AND THE PRICE OF SHARES

Once trading starts, share prices are largely determined by the forces of supply and demand. A company that demonstrates long-term earnings potential may attract more buyers, thereby enjoying an increase in share prices. A company with a poor outlook, on the other hand, may attract more sellers than buyers, which can result in lower prices. In general, prices rise during periods of increased demand i.e. when there are more buyers than sellers. Prices fall during periods of increased supply i.e. when there are more sellers than buyers.

2.2.DEFINITION OF TIME SERIES

Time series is a sequence of data points, typically consisting of successive measurements made over a time interval. Examples of time series are ocean tides, counts of sunspots and the daily closing value of the Dow Jones Industrial Average. Time series are very frequently plotted via line charts.

Time series are used in statistics, signal processing, pattern recognition, econometrics, mathematical, weather forecasting, earthquake prediction, electroencephalography, control engineering, astronomy, communications, and largely in any domain of applied science and engineering which involves temporal measurement.

Time series analysis comprises of methods for analysing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values. Regression analysis is often employed to test theories how current values of one or more independent time series affect the current value of another time series. This type of analysis of time series is not called "time series analysis", which focuses on comparing values of a single time series or multiple dependent time series at different points in time.

2.3REVIEW OF RELATED WORKS

2.3.1Forecasting Gold Prices using Time Series Analysis [6]

Model Used

The Authors used the original Box–Jenkins model, which uses an iterative three-stage modelling approach:

1. Model identification and model selection: making sure that the variables are stationary, identifying seasonality in the dependent series (seasonally differencing it if necessary), and using plots of the autocorrelation and partial autocorrelation functions of the dependent time series to decide which (if any) autoregressive or moving average component should be used in the model.
2. Parameter estimation using computation algorithms to arrive at coefficients, which best fit the selected ARIMA model. The most common methods use maximum likelihood estimation or non-linear least-squares estimation.
3. Model checking by testing whether the estimated model conforms to the specifications of a stationary univariate process. In particular, the residuals should be independent of each other and constant in mean and variance over time. If the estimation is inadequate, we have to return to step one and attempt to build a better model. Similar work was also done by [7]

2.3.2Time Series Analysis of Household Electric Consumption with ARIMA and ARMA Models [8]

The authors of this project found a model to forecast the electricity consumption in a household and to find the most suitable forecasting period whether it should be in daily, weekly, monthly, or quarterly.

The suitable forecasting methods were chosen for finding the method that was suitable for short-term analysis in daily, weekly, monthly, and quarterly basis.

2.4 Imports from the review

Upon reviewing the works of the authors, we discovered that the Box Jenkins ARIMA and ARMA model requires large number of observation for model Identification, hard to explain and interpret to unsophisticated users. As a result, we use non-parametricmodel, which is more a flexible approach and can be very well adapted to local features, which is very important in forecasting. The generalization to the problem of predicting Z_{t+l} (with $l \geq 1$) is straightforward, although presents an additional computational cost. In addition,non-parametric approach is less affected by noise and will help us obtain an accurate result. Non-parametric approach require no or very limited assumptions to be made about data.Non-parametric approach is effective for dealing with unexpected observations. Non-parametric approach is intuitive and is simple to use for both small and large samples. Non-parametric approach can be used for short and long term forecasting[9].

3.METHODOLOGY

This section focuses on the research methodology used for the study. The primary data was obtained from twelve (12) stocks on the Ghana Stock Exchange; TRANSOL, GOIL, GWEB, AYRTN, BOPP, CAL, HFC, CLYD, CPC, EBG, UT and UNIL from Jan 2010 to December 2014 which is approximately 5 years (60 months). These stocks were randomly selected.

3.1USE OF TIME SERIES ANALYSIS

Time series analysis is concerned with investigating the underlying characteristics that gives rise to a given series in order to:

(i) Understand past and present behaviour of that series and
(ii) Make forecast on the future behaviour of such a series.

3.1.1Definition

A time series is any sequence of measurements that take on a response variable over time.

Mathematically, given a data $Y_{i,}$ i=1, 2…….n, taken at regular intervals of time $t_{j,}$ j=1,2,……….n, then we define the time series as Y=F(t) where F(t) is a time dependent function expression from recorded past observation.

3.1.2Types of Time Series

There are two types of time series; For a given time series if forecast on future behaviour is made exactly based on knowledge of past behaviour, then it is called Deterministic series, however if knowledge of the past behaviour can only partly provide a probabilistic structure of the future then the series is a stochastic/statistical series[10].

For a Deterministic series, forecast is made using extrapolation technique without reference to the underlying characteristics of the series, which leads to less forecast accuracy than that of a stochastic series, which employs the use of appropriate forecasting model.

3.2Autocorrelation Function (ACF)

This measures the degree of correlation between neighbouring data observations of a time series. Generally, it is difficult to obtain a complete description of a series.

Example, one cannot restart the economy to see what other pattern they might have followed. Hence, the autocorrelation coefficient (ACF) assists us to obtain a partial description and evolution of the process for a forecasting model.

The autocorrelation coefficient is defined with lag k and denoted by $\rho_{k.}$

$$\rho_{k.} = \frac{E\left[\left(Y_t - \mu_t\right)\left(Y_{t+k} - \mu_y\right)\right]}{E\left[\left(Y_t - \mu_t\right)\left(Y_{t+k} - \mu_y\right)\right]^{\frac{1}{2}}} \quad (1)$$

$$\rho_{k.} = \frac{COV(Y_t, Y_{t+k})}{\sigma_{Y_t}\sigma_{Y_{t+k}}} \quad (2)$$

The ACF is estimated from sample observation as

$$r_k = \frac{\sum_{2}^{n}(Y_t - \bar{Y})\ (Y_{t-1} - \bar{Y}_{t-1})}{\left[\sum_{1}^{n}(Y_t - \bar{Y}_t)^2\right]^{\frac{1}{2}}\left[\sum_{2}^{n}(Y_{t-1} - \bar{Y}_{t-1})^2\right]^{\frac{1}{2}}} \quad (3)$$

The series Y_t is assumed to be stationary in the mean and variance thus the two means $\bar{Y}_t$ and $\bar{Y}_{t-1}$ can be assumed to be equal and the two standard deviations are estimated only once using all the known data for $Y_{t.}$ Therefore,

$$r_k = \frac{\sum_{1}^{n}(Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{\sum_{i=1}^{n}(Y_t - \bar{Y})^2} \quad (4)$$

The autocorrelation coefficients of a random data are approximately normal with $\mu_k = 0, \sigma_{pk} = \frac{1}{\sqrt{n}}$ where n is the size of the sample.

Thus for a random of 60 we expect $-2\sigma \le r_k \le +2\sigma_p$ for significance units of two standard errors which is

$$\frac{-2}{\sqrt{60}} \le r_k \le +\frac{-2}{\sqrt{60}} = -0.258 \le r_k \le +0.258$$

Hence, any value of r_k lying outside this interval is said to be significantly different from 0.

3.3Partial Autocorrelation Function (PACF)

This measures the degree of correlation between Y_t and Y_{t-k} when the effects of other time lags are held constant [11]. It is calculated specifically when the appropriate order of the autoregressive (AR) process (where the right hand side variables are merely the values of the dependent variable in previous periods) to fit the model is not known.

Considering this system of equations:

$\mathbf{Y_t = \Phi_1 Y_{t-1} + 1_t}$

$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + 1_t$

$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \ldots + \Phi_{k-1} Y_{t-k-1} + 1_t$

$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \ldots + \Phi_{k-1} Y_{t-k-1} + \Phi_k Y_{t-k} + 1_t$ (5)

Which presents an AR(k) process. We solve the system of equations to obtain the partial autocorrelation coefficients for the various time lagged (i.e. $\Phi_{1,}$ $\Phi_{2\ldots}\Phi_{k.}$).

However, this process is very complex and time consuming. An easier approach for computing the partial autocorrelation coefficient is by substituting the sample autocorrelation into the first k equations and solving for $\Phi_{1,}$ $\Phi_{2\ldots}\Phi_{k}$. Form, with lag kdefined as:

$R\Phi = r$

Where

$$R = \begin{bmatrix} 1 & r_1 & r_2 & \cdots & r_{k-1} \\ r_1 & 1 & r_1 & \cdots & r_{k-2} \\ r_2 & r_1 & 1 & \cdots & r_k \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{k-1} & \cdots & \cdots & \cdots & 1 \end{bmatrix} \quad (6)$$

is a k × k matrix

$\Phi^T = (\Phi_{1,}\ \Phi_{2,\ldots\ldots\ldots}\ \Phi_k)$ and $r^T = (r_1, r_{2,\ldots\ldots\ldots}, r_k)$

We solve for the various values of k using

$\Phi_1 = r_1$

$$\Phi_2 = \begin{bmatrix} 1 & r_1 \\ \frac{r_1}{1} & \frac{r_2}{r_1} \\ r_1 & 1 \end{bmatrix} = \frac{r_2 - r_1^{\ 2}}{1 - r_1^{\ 2}} \quad (6)$$

$$\Phi_3 = \frac{\begin{vmatrix} 1 & r_1 & r_2 \\ r_1 & 1 & r_2 \\ r_2 & r_1 & r_3 \end{vmatrix}}{\begin{vmatrix} 1 & r_1 & r_2 \\ r_1 & 1 & r_2 \\ r_2 & r_1 & r_3 \end{vmatrix}}$$

$$= \frac{(r_3 - r_1 r_2) - r_1(r_1 r_3 - r_2^2) + r_1(r_1^2 - r_2)}{(1 - r_1) - r_1(r_1 - r_1 r_2) + r_2(r_1^2 - r_2)} \qquad \textbf{(7)}$$

In general, for $\Phi_{k.}$ the determinant in the numerator has the same element as that in the denominator but with the last column replaced by column vector $r_{\mu,}$ μ=1 ….k.

Both the partial autocorrelation and the correlation play a very important role in the identification of a model.

The partial autocorrelation coefficient of random data are approximately normal, where n is the size of the sample. Thus for a random sample of size 60 we expect -2 $\sigma_{\Phi kk} \leq \Phi_{kk} \leq 2\, \sigma_{\Phi kk}$ for significance units of two standard errors which is

$$\frac{-2}{\sqrt{60}} \leq \Phi_{kk} \leq \frac{+2}{\sqrt{60}} = -0.258 \leq \Phi_{kk} \leq +0.258 \qquad \textbf{(8)}$$

Hence, any value of Φ_{kk}lying in this interval is said to be significantly different from zero.

3.3.1Models

The time series models to be considered are the Autoregressive AR(p), Moving Averages MA(q), Autoregressive Moving Average ARMA (p,q) and Autoregressive Integrated Moving Average ARIMA (p,d,q) based on the statistical properties of the given time series.

3.3.1.1Autoregressive Models, AR (p)

This model expresses the time series variable Y_t as a linear function of some number of actual past values of $Y_{t.}$ The general AR(p) model is given by

$$\mathbf{Y_t = \Phi_1\ Y_{t-1} + \Phi_2 Y_{t-2} + \ldots + \Phi_v Y_{t-p} + 1_t + \mu} \qquad \textbf{(9)}$$

Where p is the order of the AR model, μ is the mean of the given time series data, $1_{t,}$ the error term (which is independent to period $Y_{t-1,}$ $Y_{t-2,\ldots}Y_{t-p}$ are the values of Y_t in time lags 1,2…p respectively.

Moving Average Models MA (q)

This provides forecast of Y_t based on a linear combination of past forecast errors. The general form of an MA (q) model is given by:

$$\mathbf{Y_t = \alpha_1 1_{t-1} - \alpha_2 1_{t-2} - \ldots - \alpha_q 1_{t-q} + \mu} \qquad \textbf{(10)}$$

Where q is the order of the model, $\alpha_1,\ldots\alpha_q$ are parameters/coefficients of the model and $1_{t-1\ldots\ldots\ldots}L_{t-q}$ are residuals in the past 1…q periods respectively. We estimated the model parameters as follows:

For an MA (1) process an iterative method is used since the ordinary least square which cannot be used as the residual sum of squares is not a quadratic function. The approach suggested by Box and Jenkins is used. Given the MA (1) modelfrom (*5*

Where μ and α_1 are constants and

$$\mathbf{r_1} = \left(\frac{\alpha_1}{1+\alpha_1{}^2}\right) \qquad \textbf{(11)}$$

Select suitable values for μ and α_1 such as $\mu=\overline{Y}$ and α_1 given by the solution of r_1 = $\left(\frac{\alpha_1}{1+\alpha_1{}^2}\right)$, then the corresponding residual sum of squares is calculated using

$Y_t = \Phi_1 Y_{t-1} + \mu + l_t$

Recursively in the form

$1_t = Y_t - \mu - \alpha_1 1_{t-1}$

With 1_0=0,

We have

$1_1 = Y_{1-,}\ \mu$, $1_2 = Y_2 - \mu - \alpha_1 1_1$

$1_3 = Y_3 - \mu - \alpha_1 1_{2\ \ldots} 1_n = Y_n - \alpha_1 1_{n-1}$

Then $\sum_{t=1}^{N} l_t^2$ is calculated. This procedure is then repeated for other values of μ and α_1 and the sum of squares computed for a grid of points in the μ- α_1 plane. We then determine by inspection the least squares estimate of μ and α_1, which minimizes $\sum_{t=1}^{N} l_t^2$

Autoregressive Moving Average Models ARMA (p,q)

This is a combination of an AR(P) and MA(q) process. The general form of an AR MA (p,q) is

$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \ldots + \Phi_p Y_{t-p,} - \alpha_1 1_{t-1\ldots} - \alpha_q 1_{t-q} + \mu$ (12)

3.3.1.3The Autoregressive Integrated Moving Average ARIMA (p,d,q)

If a non-stationary time series, which has variation in the mean is differenced to remove the variation, the resulting time series is called an integrated time series. It is called an integrated mode because the stationary model, which is fitted to the differenced data, has to be summed or integrated to provide a model for the non-stationary data. Notationally, all AR(p) and MA(q) models can be represented as ARIMA models. For example, an AR(1) can be represented as ARIMA (1,0,0), that is , no differencing and no MA part. The general model is ARIMA (p,d,q) where p is the order of AR part, d the degree of differencing and q the order of the MA part i.e.

$$W_t = \vee^d\ Y_t = (1-B)^d\ Y_t \qquad (13)$$

The general ARIMA process is of the form

$$W_t = \sum_{i=1}^{p} \Phi_1 w_{t-1} + \sum_{i=1}^{q} \alpha_1 l_{t-1} + \mu + l_t \qquad (14)$$

An example of ARIMA (p,d,q) is the ARIMA (1,1,1) which has an autoregressive parameter, one level of differencing and one MA parameter is given by,

$W_t = \Phi_1 W_{t-1} + \alpha_1 1_{t-1} + 1_t$ (15)

$$(1\text{-}B)\ Y_t = \alpha_1(1\text{-}B)\ Y_t + \alpha_1 1_{t-1} + \mu + 1_t \quad (16)$$

Which can further be simplified further as

$$Y_t - Y_{t-1} = \alpha_1\ Y_{t-1} - \alpha_1\ Y_{t-2} - \alpha_1 1_{t-1} + \mu + 1_t \quad (17)$$

$$Y_t - Y_{t-1} = \alpha_1(Y_{t-1} - Y_{t-2}) + \alpha_1 1_{t-1} + \mu + 1_t \quad (18)$$

3.4 Box-Jenkins Method

This is a statistically well laid-out process of analysis in building a forecasting model which best represent a time series.To build a time series model is to fit the best linear ARIMA (p,d,q) model to a given time series and use it to predict the future movement of the series.

Box-Jenkins methodology is the best technique, which has a couple of advantages over other time series methods:

(i) It is statistically accurate and logical

(ii) It extracts enough information from the historical time series data.

(iii) It results in an increase in forecast accuracy while keeping the number of parameters to a minimum (parsimony).

Box and Jenkins in1976 put together a methodology for implementing this and the basis of this approach is described in the schematic diagram and consists of four distinct phases summarized in Figure 2.

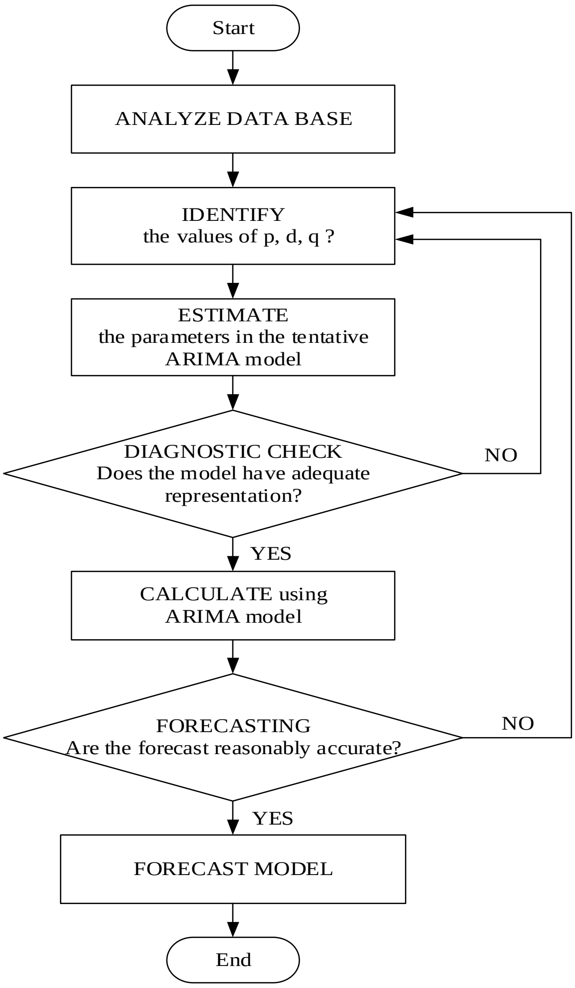


Figure 2 Schematic Diagram of BOX-JENKINS Methodology

This methodology assumes no particular pattern in the historical data of the time series to be forecasted. With an iterative approach, the procedure identifies a possible useful model from a general class of ARIMA (p,d,q) models. The chosen model is adequately checked against the historical time series to see if it accurately describes the time series, after the estimation. If the chosen model is not satisfactory, the Box-Jenkins process is repeated until a satisfactory model is found. However, if the residuals between the forecast and actual series are small, randomly distributed and independent, the chosen ARIMA (p,d,q) model is said to be a good fit.

3.4.1 Identification

The purpose of identification is to select the most appropriate orders of (p,d,q) to enable us choose a specific model from the general class of ARIMA model as well as an initial estimate of parameters. The first step in identification process is to determine whether the series is stationary. If the series is non-stationary, then it can be made stationary by the method of differencing.

Upon obtaining a stationary and/or invertible series, we go on to identify the tentative model to fit the time series. To identify the values of (p,q) we carefully examine the behavior of both sample ACF's and PACF's. Identify whether the function decays or cuts off. A survey of behaviour of ACF's and PACF's is shown in Table 2

Table 2 Behaviour of ACF's and PACF's

Model	Stationary Conditions	Invertible Conditions	Theoretical Function ACF Coefficient	Theoretical Function PACF Coefficient
1.AR (p)	Yes	No	Dies down	Cut off after lag p
2.MA (q)	No	Yes	Cut down after lag q	Dies down
3.AR MA	Yes	Yes	Dies down	Dies down

Table 3 below gives the identification for specific Time Series Model.

Table 3 Identification Table

Model	Stationary Conditions	Invertible Coefficients	Theoretical Function ACF	Theoretical Function PACF
AR(1) OR ARIMA (1,0,0)	$-1< \Phi <1$	None	Dies down	Cut off after lag 1
AR(2) OR ARIMA (2,0,0)	$\Phi_1+\Phi_2<1$ $\Phi_1-\Phi_2<1$ $-1<\Phi_2<1$	Yes	Dies down	Cut off after lag 2

MA(1) or ARIMA (0,0,1)	None	$\alpha_1<1$	Cut off after lag q	Dies down
MA(2) or ARIMA (0,0,2)	None	$\alpha_1+\alpha_2<1$ $\alpha_2-\alpha_1<1$	Cut off after two lags	Dies down
ARMA (1,1)	$-1<\Phi_1<1$	$-1<\alpha_1<1$	Dies down	Dies down

Thus the form of model can only be identifies by a careful observation of ACF and PACF by the use of SPSS. If ACF trails off impartially to zero, an AR model is identified. Similarly, if the PACF trail off to zero, a mixed ARIMA model is identified. The order of an AR is indicated by the number of the PACF and the order of the MA, by the number of ACF, that are statistically different from zero, as shown figure 2

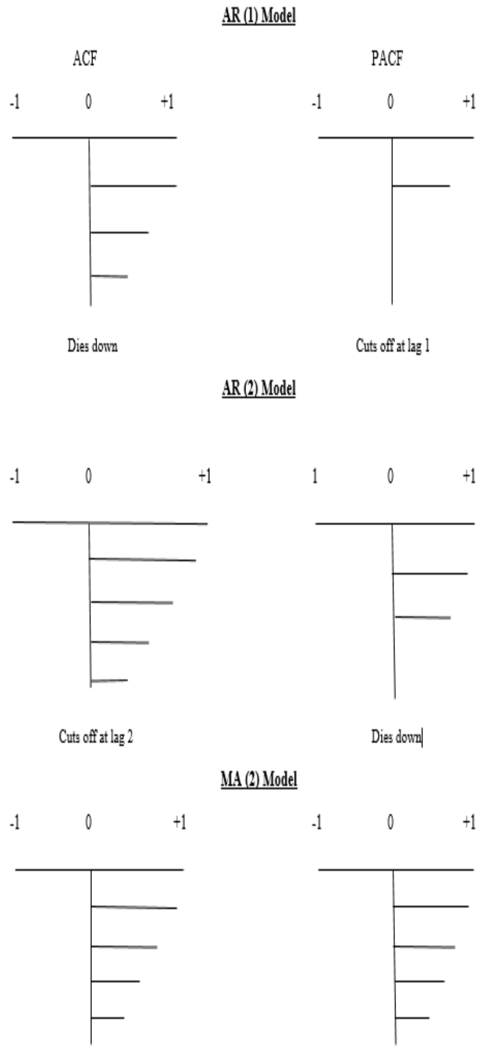


Figure 3. The order of AR Model

If the ACF or PACF at a lag comes out of the bartlet line then it is statistically different from zero but if it lies between the bartlet line then it is said to be statistically zero.

3.4.2 Estimation Process

The purpose of estimation is to calculate the parameters for the tentative model using various statistical inferences. This procedure is iterative and aims at minimizing the error term.

3.4.3Test for Adequacy

This is done before the model is used for forecasting. It is achieved by examining the error terms 1_t to be sure that they are random. If the error terms are statistically different from zero, the model is considered inadequate and an alternative model must be selected. To check for adequacy, the autocorrelations of the residual are diagnostically examined by Q-statistic given byequation 20

$$Q= n\,(n+2) \sum \left(\frac{r^2_i}{n-k} \right) \qquad (19)$$

This is approximately distributed as a Chi-Square distribution with k-p-q degrees of freedom. n is the length of the series , k is the 1st k autocorrelations being checked, r_i is the estimated autocorrelation coefficient of the ith residual term. If the calculated value of Q is greater than the corresponding Chi-Square value, the model used for forecasting it is achieved by examining the error term.

Assume a ready to forecast for a chosen model:

$$Y_t = \Phi_1 Y_{t-1}+\ldots+\Phi_p Y_{t-p}-\alpha_1 1_{t-1}-\ldots-\alpha_q 1_{t-q}+\mu \quad (20)$$

Y_t=value of observed variable at time t.

$\Phi_1 \ldots \Phi_q$ parameters of the MA model$1_{t-1\ldots}1_{t-q}$

error at the lag q.

Then forecasting for t = 20…25

$$Y_{20}= \Phi_1 Y_{19}+\ldots+ \Phi_p Y_{20-p}-\alpha_1 1_{19}-\ldots-\alpha\, 1_{20-q} \quad (21)$$

$$Y_{25} = \Phi_1 Y_{24} +\ldots+ \Phi_p Y_{21-p} - \alpha_1 1_{20-p}-\ldots-\alpha_q 1_{20-q} \quad (22)$$

3.4.4 Forecasting

After identifying and validating a model, forecast for one or several periods ahead can be made. As the forecast period become further apart, the chances of forecast error become large.

As new observations for a time series are obtained,, the model should be re-examined and checked for accuracy. If the series seems to be changing over time, the parameters of the model should be recalculated or an entirely new model may have to be developed. When small differences in forecast error are observed, then only the parameters in the model should be recalculated. On the other hand, when large differences are found in the forecast error, this gives an indication that new forecasting model must be constructed.

3.4.4.1 Selection Criteria For Computing Models

a) Bayesian Information Criterion(BIC)

The BIC is an asymptotic result derived under the assumptions that the data distribution is in the exponential family. Let: n= the number of observations, or equivalently, the sample size; k= the number of free parameters to be estimated. If the estimated model is a linear regression, k is the number of regressor, including the constant, L= the maximized value of the likelihood function for the estimated model.

The formula for the

$$BIC=-2*\ln(L) + k \ln(n) \quad (23)$$

n = sample size
k = the number of free parameters to be estimated
L = the maximized value of the likelihood function for the estimated model

Under the assumption that the model errors or disturbances are normally distributed, this becomes

$$BIC = n \ln(RSS/n) +k \ln(n) \quad (24)$$

RSS = residual sum of squares from the estimated model.
Given any two estimated models, the model with the lower value of BIC is the one to be preferred.

The BIC is an increasing function of RSS and an increasing function of k. That is, unexplained variation in the dependent variable and the number of explanatory variables increase the value of BIC. The BIC penalizes free parameters more strongly than does the Akaike information criterion. It is important to keep in mind that the BIC can be used to compare estimated models only when the numerical values of the dependent variable are identical for all estimates being compared. The models being compared need not be nested, unlike the case when models are being compared using an F or likelihood ratio test.

b) Parsimony

A model is said to be parsimonious when it contains a minimum member of parameters. Given any two competing models, the principle of parsimony requires that the model with fewer parameters is preferred to the one with more parameters.

c) Q-Statistic

After obtaining the Q-statistic associated with a model, we test it against corresponding Chi-Square statistic at its degree of freedom given by k-p-q. Any model with its Q-value greater than the corresponding Chi-Square value at the specified degrees of freedom is considered inadequate and hence discarded. For example, if ARIMA (1,0,0) is distributed at X^2_{23} and its critical value at 5% level is 35.172, then the Q-value greater than 35.172 will lead to the rejection of the ARIMA (1,0,0) model as fit for the data.

4.RESULTS, FINDINGS AND ANALYSIS

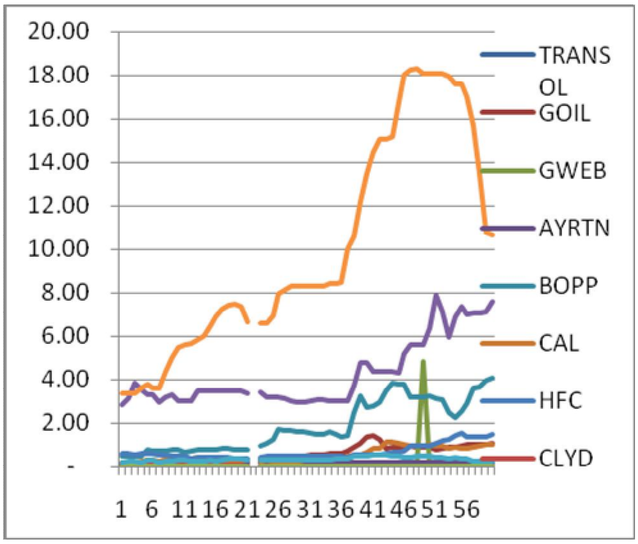
This chapter seeks to analyze the data collected and to also examine the various stock performances. SPSS was used to analyze the data.

4.1GRAPHICAL ANALYSIS OF STOCK PRICES

Table 4 consists of change in stock prices of 12 different stocks from different companies within a period of five years starting from January 2010 to December 2014.Raw data was collected from the Ghana Stock Exchange, and analysed to get the change stock prices for 12 companies chosen at random. The data contains the prices of listed companies on the stock market.

Table 4 Changes in the prices of stocks within the period of 5 years

STOCKS	Change in Price within five (5) years
TRANSOL	-0.06
GOIL	0.84
GWEB	-0.02
ARYTN	0.05
BOPP	3.59
CAL	0.81
HFC	0.88
CLYD	-0.05
CPC	-0.01
EBG	4.75
UT	0.02
UNIL	7.31



Month

Graph 1 Performances of Stock Price

TRANSOL, CLYD and CPC had a poor performance with a negative change in stock prices which indicates that their end stock prices fell below their starting stock prices.UNIL, EBG and BOPP had a good performance with positive change in stock prices within a period of five years indicating that their end stock prices exceeded the starting stock prices.

ARIMA MODEL

Table 5 UNIL ARIMA MODEL

Model	BIC	Q-STATS	Parameter	Constant
ARIMA (1,0,0)	-0.414	71.352	Lag1=0.969	4.120
ARIMA (1,1,0)	-0.993	14.148	Lag1=0.627	0.418
ARIMA (1,1,1)	-0.859	11.463	Lag1=-0.501 Lag1= -0.216	0.410

S T O C K P R I C E

ARIMA(2,0,0)	-0.977	12.860	Lag1=1.627 Lag2=0.702	3.293
ARIMA(0,0,2)	-0.042	104.572	Lag1=1.600/ Lag2=0.910	2.447

The best fitting models according to normalized BIC (smaller values indicate better fit) from the table ARIMA (1,1,0) is the best model.

Table 6 EBG ARIMA MODEL

Model	BIC	Q-STATS	Parameter	Constant
ARIMA (1,0,0)	-1.537	23.395	Lag1=0.897	2.124
ARIMA (1,1,0)	-1.527	31.747	Lag1=0.066	-0.024
ARIMA (1,1,1)	-1.657	18.163	Lag1=-0.412/lag1=-0.879	-0.036
ARIMA (2,0,0)	-1.463	22.753	Lag1=1.015/lag2=0.131	2.090
ARIMA (0,0,2)	-1.308	63.780	Lag1=-1.393/lag2=0.518	2.118

The best fitting models according to normalized BIC (smaller values indicate better fit) from the table ARIMA (1,1,1) is the best model.

Table 7 BOPP ARIMA MODEL

Model	BIC	Q-STATS	Parameter	Constant
ARIMA (1,0,0)	-2.465	32.822	Lag1=0.840	0.072
ARIMA (1,1,0)	-2.438	25.949	Lag1=0.239	0.015
ARIMA (1,1,1)	-2.369	27.336	Lag1=-0.635/lag1=-0.860	0.016
ARIMA (2,0,0)	-2.501	17.272	Lag1=1.128/lag2=-0.342	0.059
ARIMA (0,0,2)	-2.339	33.624	Lag1=-0.962/lag2=-0.662	0.039

The best fitting models according to normalized BIC (smaller values indicate better fit) from the table ARIMA (2,0,0) is the best model.

Table 8 ACF and PACF for UNIL ARIMA(1,1,0)

Residual ACF

Model		1	2	3	4	5	6	7
UNIL-Model_1	ACF	.070	-.048	-.062	-.112	.123	.024	-.020
	SE	.131	.132	.132	.133	.134	.136	.136

Residual ACF

Model		8	9	10	11	12	13	14
UNIL-Model_1	ACF	.095	.030	-.005	.047	-.157	-.177	-.029
	SE	.136	.138	.138	.138	.138	.141	.145

Residual ACF

Model		15	16	17	18	19	20	21
UNIL-Model_1	ACF	.131	.068	-.174	-.153	-.100	-.067	.045
	SE	.145	.147	.147	.151	.154	.155	.155

Residual ACF

Model		22	23	24
UNIL-Model_1	ACF	-.133	.075	-.026
	SE	.155	.157	.158

(a)

Residual PACF

Model		1	2	3	4	5	6	7
UNIL-Model_1	PACF	.070	-.053	-.056	-.107	.135	-.009	-.022
	SE	.131	.131	.131	.131	.131	.131	.131

Residual PACF

Model		8	9	10	11	12	13	14
UNIL-Model_1	PACF	.104	.045	-.020	.058	-.141	-.180	-.027
	SE	.131	.131	.131	.131	.131	.131	.131

Residual PACF

Model		15	16	17	18	19	20	21
UNIL-Model_1	PACF	.135	-.027	-.202	-.075	-.074	-.111	.042
	SE	.131	.131	.131	.131	.131	.131	.131

Residual PACF

Model		22	23	24
UNIL-Model_1	PACF	-.135	.083	-.064
	SE	.131	.131	.131

(b)

Table 9 ACF and PACF of EBG ARIMA(1,1,1)

Residual ACF

Model		1	2	3	4	5	6	7
EBG-Model_1	ACF	-.123	-.350	-.015	.175	-.021	-.106	-.056
	SE	.131	.133	.148	.148	.152	.152	.153

Residual ACF

Model		8	9	10	11	12	13	14
EBG-Model_1	ACF	.068	-.025	-.043	.029	.204	-.051	-.134
	SE	.154	.154	.154	.154	.154	.159	.159

Residual ACF

Model		15	16	17	18	19	20	21
EBG-Model_1	ACF	-.040	.102	-.059	-.012	-.021	-.051	-.005
	SE	.161	.161	.162	.163	.163	.163	.163

Residual ACF

Model		22	23	24
EBG-Model_1	ACF	.016	-.026	.035
	SE	.163	.163	.163

(a)

Residual PACF

Model		1	2	3	4	5	6	7
EBG-Model_1	PACF	-.123	-.371	-.141	.019	-.041	-.055	-.103
	SE	.131	.131	.131	.131	.131	.131	.131

Residual PACF

Model		8	9	10	11	12	13	14
EBG-Model_1	PACF	-.032	-.094	-.063	-.018	.195	.042	.001
	SE	.131	.131	.131	.131	.131	.131	.131

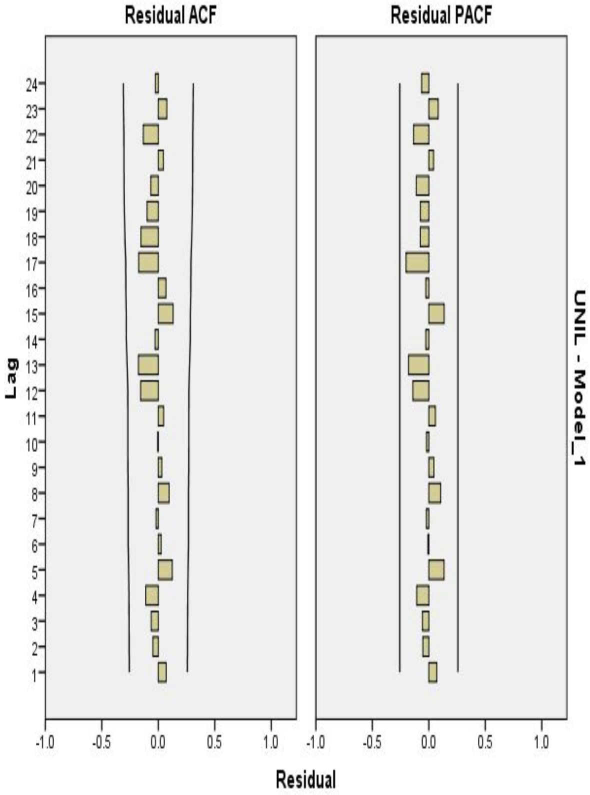
Residual PACF

Model		15	16	17	18	19	20	21
EBG-Model_1	PACF	-.073	-.019	-.101	.023	-.027	-.102	-.066
	SE	.131	.131	.131	.131	.131	.131	.131

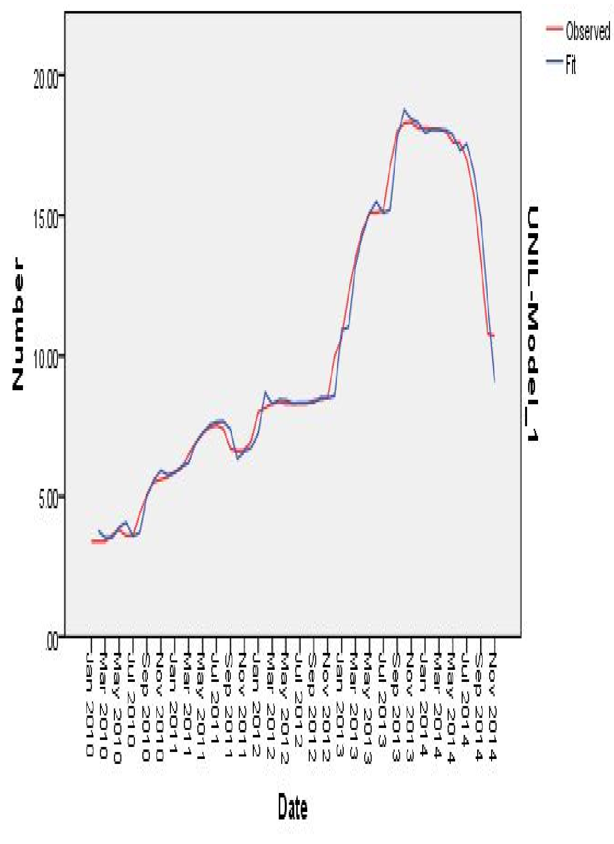
Residual PACF

Model		22	23	24
EBG-Model_1	PACF	-.096	-.119	-.072
	SE	.131	.131	.131

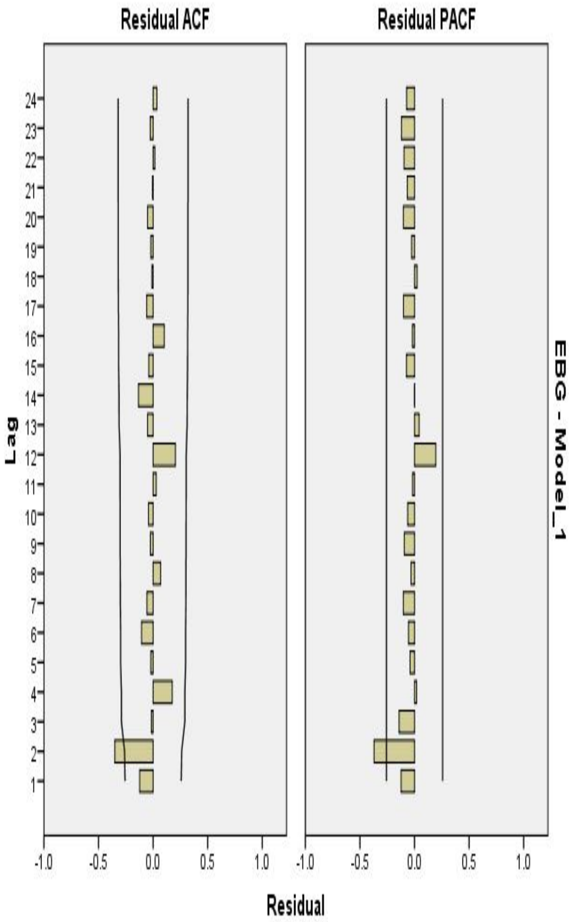
(b)



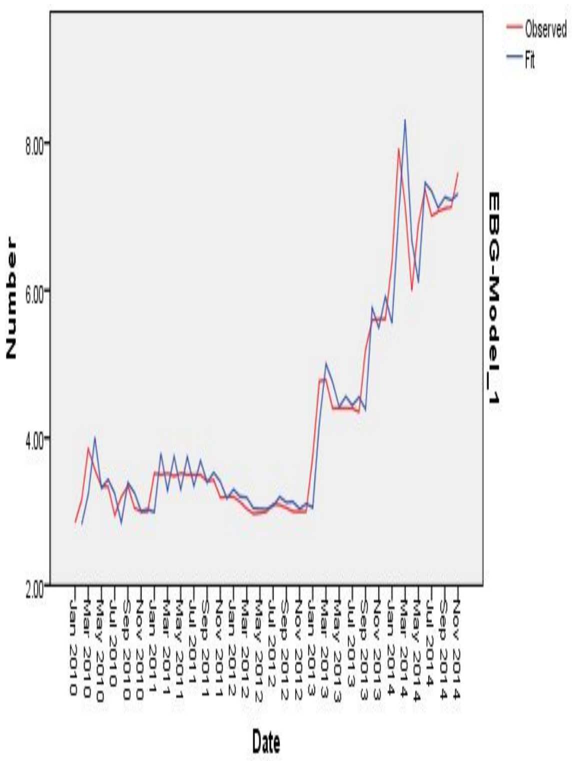
(c)



(d)



(c)



(d)

Table 10 ACF and PACF for BOPP ARIMA(2,0,0)

Residual ACF

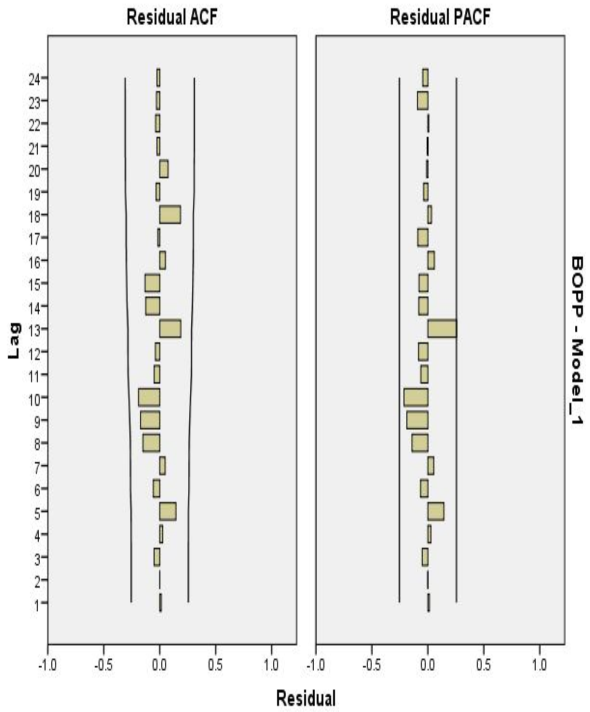
Model		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
BOPP-Model_1	ACF	.014	.001	-.051	.025	.144	-.057	.048	-.150	-.172	-.190	-.050	-.039	.187	-.125	-.132	.050	-.018	.185	-.032	.073	-.025	-.037	-.030	-.025
	SE	.130	.130	.130	.131	.131	.133	.134	.134	.137	.140	.145	.145	.145	.149	.151	.153	.153	.153	.157	.157	.158	.158	.158	.158

(a)

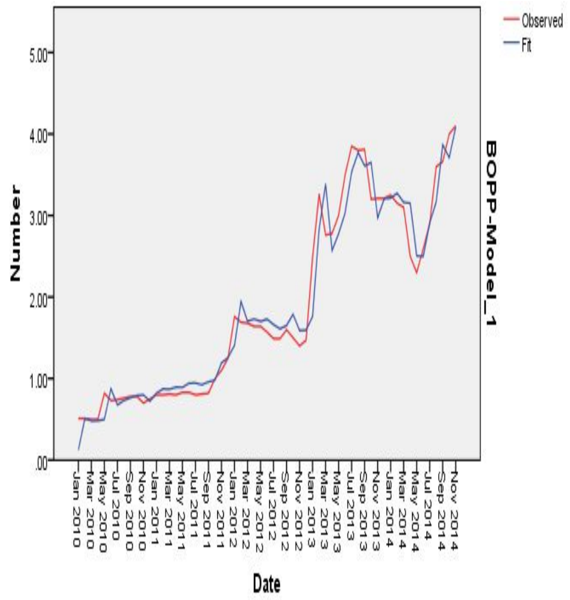
Residual PACF

Model		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
BOPP-Model_1	PACF	.014	.000	-.051	.027	.144	-.065	.053	-.141	-.188	-.214	-.063	-.083	.257	-.081	-.079	.058	-.090	.033	-.038	-.011	-.005	.008	-.093	-.046
	SE	.130	.130	.130	.130	.130	.130	.130	.130	.130	.130	.130	.130	.130	.130	.130	.130	.130	.130	.130	.130	.130	.130	.130	.130

(b)



(c)



(d)

4.3FORMULATION OF THE PREDICTION EQUATION

Table 11 Best fit models for UNIL, EBG and BOPP

Model ARIMA	BIC	Parameters	Constant
UNIL(1,1,0)	-0.993	Lag1=0.627	0.418
EBG(1,0,0)	-1.537	Lag1=-0.412 lag1=-0.879	-1.537
BOPP(2,0,0)	-2.501	Lag1= 1.128 Lag2=-0.342	0.059

4.3.1PREDICTION FORMULA

$$\mathbf{Y_t = \alpha\, Y_{t-1} + \beta\, Y_{t-2} + \mu;\ t\text{-}1 < t > t\text{-}2} \quad (25)$$

Where Y_t is the predicted stock price for a particular month.

Y_{t-1} /Y_{t-2} is the stock price for a particular month.
α is the Lag 1
β is the Lag 2
μ is the constant
t is a particular month

4.3.1.1Calculating stock prices UNIL, EBG and BOPP for the next three month.

UNIL ARIMA (1,1,0)

61st month	62nd month	63rd month
t = 61, t-1 = 60	t = 62, t-1 = 61	t = 63, t-1 = 62
α = 0.627, and μ=0.418	α = 0.627, and μ=0.418	α = 0.627, and μ=0.418
Y_{60} = 10.7	Y_{61} = 7.1269	Y_{62} = 4.8866
Y_{61} = 0.627 (10.7) + 0.418	Y_{62} = 0.627 (7.1269) + 0.418	Y_{63} = 0.627 (4.8866) + 0.418
Y_{61} = 7.1269	Y_{62} =4.8866	Y_{63}=3.4819

Therefore, the stock prices of UNIL for the next three will be 7.1269, 4.8866 and 3.4819 respectively indicating fall in stock prices.

EBG ARIMA (1,0,0)

61st month	62nd month	63rd month
t = 61, t-1 = 60, t-2 =59	t = 61, t-1 = 60, t-2 =59	t = 61, t-1 = 60, t-2 =59
α = 0.897 and μ=2.124	α = 0.897 and μ= 2.124	α = 0.897 and μ= 2.124
Y_{60} = 7.6	Y_{61} = 8.9412	Y_{62} = 10.1443
Y = 0.897(7.6) + 2.12	Y_{62} = 0.897(8.9412) + 2.124	Y_{63} = 0.897(10.1443) + 2.124
Y_{61} =8.9412	Y_{62} = 10.1443	Y_{63} = 11.2243

Therefore, the stock prices of EBG for the next three will be 8.9412, 10.1443 and 11.2243 respectively indicating rise in stock prices

BOPP ARIMA (2,0,0)

61st month	62nd month	63rd month
t = 61, t-1 = 60, t-2 =59	t = 61, t-1 = 60, t-2 =59	t = 63, t-1 = 62, t-2 =61
α = 1.128, β = -0.342 and μ= 0.059	α = 1.128, β = -0.342 and μ= 0.059	α = 1.128, β = -0.342 and μ= 0.059
Y_{60} = 4.1 and Y_{59}=4	Y_{60} = 4.1 and Y_{59}=4	Y_{60} = 4.1 and Y_{59}=4

Y_{61} = 1.128 (4.1) - 0.342 (4) + 0.059	Y_{62} = 1.128 (3.3158) - 0.342 (4.1) + 0.059	Y_{63} = 1.128 (2.397) - 0.342 (3.3158) + 0.059
Y_{61} = 3.3158	Y_{62} = 2.397	Y_{63} = 1.6288

Therefore the stock prices of BOPP for the next three will be 3158, 2.397 and 1.6288 respectively indicating rise in stock prices.

5.CONCLUSION

At the end of the study, the future stock prices of the listed companies were able to be determined through the developed mathematical model. Past data of UNIL,EBG and BOPP was fix into the model to predict the stock price for the next three months. The output result of UNIL for the 61st month, 62nd month and 63rd month were 7.1269, 4.8866 and 3.4819 respectively. The predicted stock price for EBG were 8.9412, 10.1443 and 11.2243 respectively, that of BOPP were 3.3158, 2.397 and 1.6288 respectively. According to the predicted results of the following companies, the stock prices of UNIL and BOPP will decrease as time goes by. EBG showed steadily increase in stock prices as time goes by. This shows that EBG is the best company to invest in order to obtain optimum profit at the time of conducting this research.

References

[1]. A. Addo and F. Sunzuoye, “Short term interest rates impact in Ghana Stock Exchange,” European Journal of Business and Econimics, 2012.

[2]. S. Chen, X. Lan, Y. Hu, Q. Liu and Y. Deng, “The time series forecasting: from the aspect of network,” Cornell University, New York, 2014.

[3]. B. M. Smith, A history of the Global stock market, Chicago: University of Chicago Press, 2003.

[4]. S. Antwi, E. F. E. A. Mills and X. Zhao, “An Empirical Analysis of the Performance of the Ghana Stock Exchange and Treasury Bills,” International Journal of Business and Social Science, vol. 3, no. 23, pp. 97-108, 2012.

[5]. R. C. Enyaah, “An analysis of the effect of interest rate and exchange rate changes on stock market returns:Empirical evidence of Ghana Stock Exchange,” Kwame Nkrumah University of Science and Technology, Kumasi, 2011.

[6]. A. Anand, “Forecasting Gold Prices using TimeSeries Analysis,” Piyush Dharnidharka, India, 2012.

[7]. M. M. A. Khan, “Forecasting of Gold Prices (Box Jenkins Approach),” International Journal of Emerging Technology and Advanced Engineering, vol. 3, no. 3, 2013.

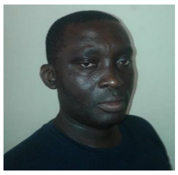
[8]. P. Chujai, N. Kerdprasop and K. Kerdprasop, “Time Series Analysis of Household Electric Consumption with ARIMA and ARMA Models,” International Association of Engineers, 2013.

[9]. J. M. Vilar-Fernandez and R. Cao, “Nonparametric forecasting in time series. A comparative study,” Facultad de Inform´atica, University of. A Coruna, Coruna,Spain, 2005.

[10]. T. Awokuse and T. Ilvento, “Using Statistical Data to Make Decisions,” University of Delaware, Delaware, 2010.

[11]. S. K. Amponsah, D. Otoo and C. A. K. Todoko, “Time series analysis of water consumption in the Hohoe municipality of the Volta region, Ghana,” International Journal of Applied Mathematical Research, vol. 4, no. 2, pp. 393-404, 2015.

Authors



Emmanuel Kwame Mensah is a lecturer at Ghana Technology University College. Heobtained both BSc and MSc from Kwame Nkrumah University and Science Technology Ghana in the field of Mathematics. He specialized in Mathematical Control Theory. Moreover, he has another degree in Postgraduate Diploma in Education from University of Education, Winneba - Ghana. His research area is stochastic optimal control theory.

Kweku Arthur is a final year PhD student at the ST. Petersburg State University of Telecommunications. He obtained his BSc and MSc degree in the same university. He lectures at the Ghana Technology university college. His research areas are wireless technology and emerging trends in mobile networking