

**Strathmore**  
UNIVERSITY

**FORECASTING EQUITY PRICES FOR SELECTED  
COMPANIES AT THE NAIROBI SECURITIES EXCHANGE**

**ACHIENG' SANDRA OCHIENG'  
078984**

**Submitted in partial fulfillment of the requirements for the Degree of  
Bachelor of Business Science Financial Economics at Strathmore University**

**Strathmore Institute of Mathematical Sciences  
Strathmore University  
Nairobi, Kenya**

**[October, 2016]**

## **Declaration**

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the Research Proposal contains no material previously published or written by another person except where due reference is made in the Research Proposal itself.

©No part of this Research Proposal may be reproduced without the permission of the author and Strathmore University.

\_\_\_\_\_ **Name of Candidate**  
\_\_\_\_\_ **Signature**  
\_\_\_\_\_ **Date**

**This research proposal has been submitted for examination with my approval as the Supervisor.**

\_\_\_\_\_ **Name of Supervisor**  
\_\_\_\_\_ **Signature**  
\_\_\_\_\_ **Date**

## **Abstract**

Information asymmetry is the main cause of uncertainty in security exchanges all over the world. There are “informed investors” and “uninformed investors” with the latter having imperfect information. Due to this uncertainty, investors have been trying to come up with ways of predicting stock prices and to find the right stocks and perfect timing for when to buy or sell.

The primary target of this research is to construct a model that will forecast the short term stock prices for five selected companies listed in the Nairobi Securities Exchange divided into those that are highly traded, highly capitalized and highly volatile. Secondary datasets of returns on Kenyan stock market prices were retrieved from online sources such as the Nairobi Securities Exchange website and the Valuraha platform. The model employed in this paper took the form of an autoregressive integrated moving average (ARIMA). Results obtained revealed an impressive performance of the ARIMA model in stock price prediction especially when it came to the highly traded and highly capitalized stocks.

# TABLE OF CONTENTS

Declaration.....	1
Abstract.....	2
TABLE OF CONTENTS .....	3
1. INTRODUCTION.....	1
1.1. Background of the study.....	1
1.2. Motivation of the study.....	3
1.3. Problem Statement.....	4
1.4. Research Questions.....	5
1.5. Research Objectives.....	5
1.6. Value of the study.....	5
2. LITERATURE REVIEW.....	6
2.1. Introduction.....	6
2.2. Factors that affect stock prices .....	7
2.3. Creating a model for stock price prediction .....	10
3. METHODOLOGY .....	15
3.1. Empirical strategy.....	15
3.2. Data collection .....	15
3.3. Model specification .....	15
3.3.1. Autoregressive Integrated Moving Average/ Box-Jenkins technique .....	16
4. DATA ANALYSIS .....	21
A. Highly traded stocks .....	21
B. Highly capitalized stocks. ....	25
C. Highly volatile stocks .....	28
4.1. FINDINGS .....	29
5. DISCUSSION .....	34
6. CONCLUSION .....	36
7. RECOMMENDATIONS FOR FURTHER STUDIES.....	36
APPENDIX.....	37
REFERENCES.....	37

# 1. INTRODUCTION

## 1.1. Background of the study

A share is a unit of ownership that entitles the holder to a proportion of a company's capital. The shareholder is exposed to opportunities when it comes to claiming profits or loses the company has made. Shares can be categorized as either equity or preference shares. Equity shareholders get to earn voting rights in the company's decisions making in addition to sharing the profits and losses. On the other hand, preference shareholders have no voting rights and they earn fixed dividends.

When a firm goes public, it issues shares for the first time in a primary market. A primary market is where firms float new bonds and stocks to the public for the first time. At this point in time, the company issues an initial public offering (IPO). An initial public offering involves the first sale of a company's stock to outside investors, after which the stock is traded in secondary markets, as explained by Ehrhardt & Brigham (2005). An IPO consists of four groups; the issuing firm, the investment bank, the initial buyers and the larger set of investors in the secondary markets, Yalla (2014). Investment bankers, lawyers and accountants prepare a document explaining the firm's history, suggested offering, current businesses and plans for the future. This is composed of the registration statement that is presented to the stock exchange for approval. A prospectus revealing the most important statements is distributed to investors, Nderi (2009). Thereafter, investment banks assist the firm in determining the offering price range and the amount of stocks to be sold. Convincing the potential investors to buy the stock at the offering price is extremely dependent on the investment bank's reputation.

In secondary markets, firms also split their stocks to signal the management's confidence in future stock performance, obtain an optimal price range for liquidity or most favorable tick size. Chemarum & Omollo (2010). A stock split is an exchange of shares in which at least five shares are distributed for every four formerly outstanding, Eugene F. Fama (1969). Stock splits are unusually preceded by periods of unusually high rates of return on securities. However, these periods begin long before any information of a possible split hits the market. A stock split is done by firms that note that their share price levels are extremely high compared to other firms in

the same industry. Their intention is always to make their shares appear more affordable to investors even though the company value remains the same, Chemarum & Omollo (2010).

Forecasting on the other hand is the approximation of the value of a variable (or set of variables) at some point in time Choong (2012). It enables organizations to come up with future plans and is also useful in decision making. Such exercises work on the presumption that if we can predict the future we will be able to modify our behavior now to be better placed in future. It is implemented in areas such as:

- Designing economic policies – Different governments and business ventures forecast economic information such as interest rate growth, unemployment and inflation rates in future planning.
- Inventory control – Predicting future demand for a product enables manufacturers to control the stock of raw materials and finished goods and hence plan the production schedule.
- Investment policy – This involves forecasting financial information such as exchange rates and share prices, which is the main aim of this project.

### **History of equity trading**

America's first stock exchange was actually founded in Philadelphia in 1790. Soon after the New York merchant group realized that their stock exchange was in the verge of making losses following the early mayhem caused by war for bonds and stock in the Bank of the United States, they sent an informant to Philadelphia in early 1817. His return bore news of a successful Philadelphia exchange upon which there was a formal organization of the New York Stock and Exchange Board in March 8, 1817.

The birth of Wall Street was based on a wall built by Dutch settlers to protect themselves from pirates and other dangers a century earlier. Since the wall joined the banks of the East River with those of the Hudson River on the west, it had become a hot spot for commercial trade. The path had many buildings including shops and warehouses created by early merchants as well as a church and city hall. Later in March 1792, a group of twenty-four top merchants in New York City held a secret meeting at Corre's Hotel to come up with ways to organize the securities business in a more orderly framework as well as to beat their competitors and other auctioneers.

Two months later, they gathered at their regular meeting point, the Buttonwood tree and signed a document dubbed the Buttonwood Agreement. The contract required those involved to trade securities only among themselves, to set trading fees, and not to take part in other auctions of securities. The group had formed what was later to be known as the New York Stock Exchange that was later located at 11 Wall Street.

In Kenya, share trading kicked off in the 1920s when the country was still a British colony. At this point in time, a formal stock market was yet to be established, therefore no rules had been set out to govern share trading. The trade was founded on a gentleman's agreement where standard commissions were charged and clients were obliged to meet their contractual agreements of making proper deliveries and settling the required costs. As a result, there was no physical trading floor or trained stockbrokers.

Before the 1950s, shares in local companies were traded by estate agents as a supplementary part of their day to day businesses, Parkinson (1984). The exchanges would be carried out in relaxed environments where the agents and lawyers would meet over a cup of coffee. The initiation of stock brokerage firms was in the 1950s. The first firm was inaugurated by Francis Drummond in 1951. Foreign investors were the main participants in the capital markets because their high levels of income allowed them to accumulate savings as well as to take up investments in securities. In addition, they had the mastery of operating organized capital markets.

On the other hand, local citizens had very minimal participation which was attributed to their low income and statutory restrictions during the period before independence. The local stock brokerage firms desired regulations for their share dealings. They yearned for a formal organization, brokerage charges and broker/client relationships. Furthermore, the then Minister of Finance longed for a formal organization to improve the raising of government loans from local sources within Kenya. Thus, in July 1954, a new era was born and the NSE was established, Ngugi (2003).

## **1.2. Motivation of the study**

Research conducted in the past years has been based on either fundamental analysis or technical analysis. Fundamental analysis is based on the assumption that the stock price of a company is dependent on the intrinsic value of the firm and an expected return on a company's investment. It

is associated with macroeconomics in addition to the analysis of a firm's operations as well as the industry in which the company is established. An example of this is where financial news articles were used to predict stock prices, as they played a major role in influencing the movement of a stock since people react to the information.

On the other hand, technical analysis aims to forecast future stock prices using past stock prices and volume information. Here, the main assumption is that history repeats itself. Once a trend is discovered, future prices can be modelled along this path to generate profits. Examples of this include data mining and use of decision trees, Yalla (2014). However, most people believe that fundamental analysis works only on a long term basis and is unsuitable for short term predictions. Additionally, many techniques used in technical analysis have been found to be highly subjective in nature and not statistically valid, Qasem (2013)

This research project aims to strike a balance between the technical and fundamental analyses techniques, in order to create model that will predict short term stock prices for companies listed at the Nairobi Securities Exchange.

### **1.3. Problem Statement**

Uncertainty still remains a nightmare to most investors. This is even as technological innovation over the years as well as new free and open economic policies have led to more investors seeking to buy shares in various stock markets, as indicated by Gatua (2013). Information asymmetry, being the main cause of uncertainty, builds different levels of information. Consequently, there are "informed investors" and "uninformed investors" with the latter having imperfect information. As a result, there is a higher probability for the uninformed investors to undertake "bad" issues as the informed investors will only compete with them for the "good" ones, Qasem (2013). This competition is attributed to most people investing with the main aim of a fall back plan for later, a secure future. As a result, the number of investment banks and stock brokers in Kenya has increased relatively over the years.

Due to this uncertainty and pricing concerns caused by the growing number of investment firms, investors have been trying to come up with ways of predicting stock prices and to find the right stocks and perfect timing for when to buy or sell, Qasem (2013). Accordingly, this project aims

to identify and construct a model that can be used to forecast short term share prices for selected companies at the Nairobi Securities Exchange.

#### **1.4. Research Questions**

The following questions shall guide this study;

- 1) What factors constitute a share price?
- 2) Is it possible to create a model that could determine and predict share prices?

#### **1.5. Research Objectives**

The objectives of this thesis are;

- 1) To determine factors that affect share prices.
- 2) To identify a model that can be used to determine and predict share prices.

#### **1.6. Value of the study**

This study aims to give an understanding on pricing of stocks and investment techniques to different users but mainly to the following groups of people;

- a) Firms.
- b) Investment bankers and individual investors.
- c) Scholars.
- d) The government.

## **2. LITERATURE REVIEW**

### **2.1. Introduction**

The behavior of stock market prices is considered a controversial topic in both academic and business circles. This is because most researchers had tried to figure out to what extent the past history of a common stock's price can be used to make meaningful predictions concerning the future price of the stock. In Fama (1965) this has been answered by either chartist theories which assume that the past behavior of a security's price is rich in information concerning its future behavior or by the theory of random walk. King (1966) refers to the chartists as "trendists", who believe that an assessment of the patterns in price changes could provide future information on the direction and magnitude of price. The random walk theory on the other hand states that the series of price changes has no memory and hence the past cannot be used to predict the future. According to King, random walk theory is more associated with academicians while trendists are more of professional analysts of the stock market. In his paper, Fama looked at the theories underlying the random walk then proceeded to test the validity of the theory. The study's empirical evidence went on to present consistent and strong support for the model implying that chartist theories have no real value to an investor in the stock market.

Mandelbrot (1966) attempted to relate the behavior of stock prices to more fundamental economic "triggering" quantities. This was mainly because previous work had been focused on the statistical behavior of price series themselves. His work involved a detailed application of certain ideas current in economic theory at the time, analyzing the roles of anticipation and of expected utility. His findings depended on the performance of the underlying "triggering" variable and the relationship between that variable and price. According to him, models where the price series follows a pure random walk could be created. On the other hand, it was also possible to create models where successive price changes were dependent such that the prices were not geared towards a pure random walk, but the dependence could also not be used to increase the expected profits. In this case, the price would be considered a martingale.

The following part of this chapter looks at literature on various factors that affect stock prices as well as some of the techniques that have been used to model stock prices.

## **2.2. Factors that affect stock prices**

As security prices play an important role in directing the flow of capital in various business entities, Fisher (1961) analyzed four factors which have an impact on share prices. These were; the last declared dividends per share, the last declared undistributed profits per share, the past average annual growth in dividends per share and the size of companies to which the shares correspond. In his model, the capital market is imperfect and faced with uncertainty as the reverse of this would only mean that a security's value would be the total sum of its discounted future returns. In his analysis, he preferred the method which considered only the most active shares irrespective of the industry as compared to the one which involved choosing firms on an industry basis and thereafter rejecting those that did not involve an active market. The results showed that all the four factors had an impact on share prices though in different magnitudes. Variations in the last declared dividend per share coupled with the last declared undistributed profits per share were more significant in explaining company share prices as opposed to the independent variables alone. This was also the case when size was introduced.

Fama (1969) teamed up with Fisher, Jensen and Roll to determine the adjustment of stock prices to new information. Their study had two main objectives; to determine whether there was some "unusual" behavior in the rates of return on a split security in the months surrounding the split, and, given that a split could occur as a result of some "unusual" behavior in stock returns, to what extent could this be explained by the relationships between splits and changes in other more fundamental variables. In order to meet their first objective, they showed that prior to a stock split occurring, the rates of return incorporating dividends and capital appreciation on the securities to be split were unusually high. As a result of high periods of return long before any information hits the market, they concluded that these returns arise from the fact that during the pre-split period the companies in question experience dramatic increases in expected earnings and dividends.

Classical monetarists on the other hand believed that money supply largely contributed towards the total output in the economy. When the Neo-classists came into the picture, they analyzed the impact money supply had on individual portfolio balances. Their results proved that there was indeed a relationship between the two variables and as such, changes in the optimum portfolio would lead to changes in the market demand for securities. High demand for these securities

would therefore translate to higher stock prices. Sen (1974) also analyzed the impact of the rate of change of money supply on stock prices. His analysis also included other independent variables: the absolute level of money stock, the level of productive activity in the country and the rate of change in the productive activity. The stock of money was a summation of liquid currency and demand deposits held in banks while the level of productive activity was a function of the Gross National Product (GNP).

A linear relationship between the dependent variable (index of stock prices) and the explanatory variables (money stock, rate of change of money supply, GNP and the rate of change of GNP) was established. This allowed for a multiple regression analysis to be carried out. In his results, Sen (1974) demonstrated that there exists a relationship between the dependent and the independent variables. He found a negative relationship between stock prices and the level of money supply of the initial time period. The level of GNP however explained the most variations in stock prices. This was closely correlated with the theory given by the Keynesian economists. They stated that changes in the level of GNP have two effects on an individual's disposable income. One was the liquidity effect which influences a person's demand for assets and thus adjusts the stock market. The second involved the impact on the consumer's demand for goods and services which affects the investment climate in the economy leading to a pressurized stock market and hence more pressure on the stock prices to change.

Peavy (1992) studied the impact of both earnings and interest rates on share prices. Investors carefully investigate movements in firm earnings and interest rates so as to evaluate their next move in the stock market. This was because in the long run, share prices vary directly with company earnings and inversely with interests. Therefore, an increase in interest rates in the economy would translate to lower equity prices and vice versa. In an attempt to indicate the importance of both earnings and interest rates in evaluating stock prices, he performed a multiple regression to examine the relationship between the S&P 500 at the end of each year from 1979 to 1989 along with the corresponding year end values for the S&P 500's underlying earnings per share and interest rates of the 10 year U.S Treasury Constant Maturity Index. The results proved that earnings and interest rates do matter when explaining the variations in stock prices.

In their share price puzzle research, Edward & William (2006) agreed with Fisher's (1961) notion of an imperfect capital market. In this case, market frictions exist due to the presence of

transaction costs. In this paper they determined whether there is an optimal trading range for a firm's shares and if so why share prices differ among companies in a similar market. To move their share prices to trading ranges with higher transaction costs, firms use stock splits. This is due to incomplete information in the market which causes different investors to have differing levels of knowledge on company securities. As such, investors will only choose securities they know more about and the larger the investor base, the higher the firm's market value. Other factors considered in the paper to affect share prices include brokerage commissions, tick size and reductions in the minimum trading unit.

Edward & William (2006) stated that brokerage commissions have an inverse relationship with share prices and therefore higher brokerage commissions would translate to lower stock prices. Thus in an imperfect capital market, with imperfect information, brokers would provide more information on lower-priced firms. Additionally, reductions in the minimum trading unit (least number of a company's shares needed for a transaction in that company's stocks) would lead to an increase in the number of investors and hence a higher demand. As a result, share prices will increase.

Panda (2008) also analyzed a similar problem to Peavey (1992), the effect of interest rates on stock prices in the Indian market. Theoretically, interest rates and stock prices have a negative relationship. This is as a result of the cash flow discounting model under which the current values of stocks are computed by discounting the future cash flows at a discount rate which is a risk adjusted required rate of return and equivalent to the level of interest rates in the economy. As the discount rate rises, the current values of stocks decline. The equivalence between the discount rate and interest rates would then mean that changes in interest rates would affect stocks directly. Increased interest rates would lead to decreased present values of stocks and vice versa. The negative relationship between interest rates and stock prices in the long term was also supported by the following arguments.

First, as interest rates on bonds rise, bonds become more attractive to investors and their demand increases while that of stocks falls. As a result, funds are shifted from the stock market to the bond market and stock prices reduce. Firm profitability also declines as interest rates rise due to reduced cash flows which is attributed to a fall in both the firm's earnings net of interest rates and consumers' demand for the products which makes the firm to pay more in order to borrow

money. The low firm profitability then translates to reduced stock prices. Thirdly, interest rate changes also affect investor expectations on about the economy and a firm's earnings. High interest rates would lower investor expectations and thus lead to lower stock prices.

However, in the short run, interest rates affect stock prices positively. According to Peavy (1992), this was due to the following: a fast growing economy that translates to higher interest rates should also mean that a firm's earnings increases rapidly and therefore, lead to higher stock prices. In addition, such an economy would most likely be facing high inflation rates. In order to curb this, firms would then increase their price levels hence higher earnings per share. Using the stock valuation formula, a rise in the discount rate will affect the earnings per share positively and this does not necessarily guarantee low stock prices. In order to justify all these arguments, Peavy (1992) used the Johansen cointegration technique to analyze the monthly averages of the BSE SENSEX (S&P Bombay Stock Exchange Sensitive Index) and the NIFTY (an indicator of the 50 top major companies on the Indian National Stock Exchange), which were used to measure stock prices, and the month-end yields on 10-year government security and TB-15-91 (a 15-91 days treasury bill), which were proxies for the long term and short term interest rates respectively. The results showed that in the in the short run, long-term interest rates affect stock prices negatively while short term interest rates affect interest rates positively.

### **2.3. Creating a model for stock price prediction**

As the optimism for forecasting stock prices grew, new lessons for different techniques also came about. This was clearly shown in Granger's (1992) lessons for forecasters. In his paper, he states that stock price prediction is more than an analysis of simple univariate data or just inputting data in popular computer packages. He described it as more ambitious and wide ranging. There were more advantages to using longer time horizons, using disaggregated data, carefully getting rid of outliers and considering non-linear models. In addition, more sub periods were found to be more predictable than others, such as January which grew to be commonly known as the January effect, where returns were found to be higher. Empirical evidence suggested that the average of returns have regime changes with an indicator variable that records one in January and zero in the preceding months.

There was also insufficiency in producing a regression model that assessed only in sample data. This was due to small sample, in-sample biases of coefficients which could reveal extremely

encouraging outcomes. To prevent data mining as well as the above difficulties, Granger (1992) pushed for the evaluation of out-of-sample data. However, most of the forecasting techniques failed to properly input the levels of risk as well as transaction costs. As a result, such methods fail to provide any proof that the efficient market hypothesis holds. In order for Efficient Market Hypothesis (EMH) to be rejected, a model would have to exist that would continuously produce positive profits after proper adjustments of risk and transaction costs. Moreover, this model would have to be in public use for a while. Consequently, Granger believed that the justification of the efficient market hypothesis is impossible.

In his pursuit of forecasting techniques and a contrary view of the efficient market hypothesis, Granger teamed up with Timmermann (2004) to show how forecasters can implement EMH in their models or have a basis for rejecting it. According to the two, a capital market is said to be efficient if it fully and correctly reflects all relevant information in determining security prices. Thus, market returns are in line with EMH as investors would be able to make massive profits. The information set in a capital market would either portray a weak form, semi-strong form or a strong form efficient market. The three forms differ on the basis of the type of information available, with the information set containing past information, past and public information, and past, public and insider information for the weak form, semi-strong form and strong form markets respectively.

Most areas on prediction are based on the weak and semi-strong forms as accessing insider information would be a bit too costly for investors. Such transaction costs and trading restrictions also affect tests of market efficiency. In cases where transaction costs are too high, predictability would not be ruled out by arbitrage since it would be too expensive for an investor to take advantage of any opportunities in the market. Ultimately, Timmermann and Granger (2004) established that EMH does not rule out prediction of variables that may not necessarily lead to profit accumulation. However, they added that in as much as prediction models may come about, they may be short lived since the information used in these models will be incorporated into market prices after a while and cease to be successful.

The growing importance of stock price prediction among investors spurred the interest in building better models for forecasting purposes. Forecasting was done either from a statistical or artificial perspective. The artificial perspective involves the use of neural networks in areas such

as data mining especially for pattern recognition. According to Mahdi, Hamidreza & Hashemi (2010), artificial neural networks which were influenced by the activity from the human brain cells are able to memorize data patterns and derive their knowledge to recognize the future patterns. The use of neural networks as a forecasting method was attributed to the following features. The networks are self-adjusting techniques based on training data (previous data) enabling them to find solutions to the problem with minimal information about its model and without binding the prediction model by adding any additional assumptions. Subsequently, they possess the ability to recognize new patterns that had not been previously noted in the training set.

Mahdi, Hamidreza & Hashemi (2010) applied two types of neural networks; a feed forward multi-layer Perception (MLP) and an Elman recurrent network to forecast a company's stock value based on historical values. However, they were unable to determine an optimal method for prediction as both yielded different results when compared to the linear regression method. In comparison to the Elman and linear regression, the MLP recorded a lower mean squared error, mean absolute percentage error and mean absolute deviation. Therefore, the Elman and the linear regression can forecast the direction of the changes of the stock value better than the MLP but the Elman suffers from a greater error in prediction. However, the experimental results showed that the application of the MLP is more promising in predicting stock value changes rather than the Elman network and the linear regression model.

Other writers have come with other methods of predictions apart from the ones discussed above. According to Tseng, Kwon & Luna (2012), short term predictions are determined by technical analysis while longer term predictions are results of fundamental analyses. Despite that fact, stock market prices are difficult to forecast. Supporters of the efficient market hypothesis argue that the prices cannot be predicted because all the market information is reflected by the prices. The antagonists countered this argument by stating that, since the current price reflects all the market information, then both current and historical prices can be used to predict future prices.

Due to the fact that traders made investor decisions by analyzing more data patterns than fundamental factors affecting stock prices, both time series analysis of information as well as technical analysis grew in importance. Consequently, Tseng, Kwon & Luna (2012) applied the Holt/Winters models, ARIMA models, neural networks and the traditional time series

decomposition to a total of 3105 observations. The period covered both a housing and tech boom and bust, the historical 9/11 event, a recession and slow recovery for the sample prices. Due to the global financial and economic prices, there were very minimal expectations of stock price prediction. The recorded coefficient of determination was about 0.995 and hence all the three time series analyses fit the data perfectly. For the out-of-sample forecasts over 60 trading days, the mean absolute percentage errors were lower for the ARIMA, Holt/Winters model and the normalized neural network model. However the forecasting errors were large for the time series decomposition and the non-normalized neural network model.

To improve on this, Ayodele & Aderemi (2014) presented the process of building a predictive model using the Autoregressive Integrated Moving Average in order to forecast both Nokia's stock price as well as that of Zenith Bank. Before their analysis, a lot of work had been done on the artificial neural networks model whose popularity had grown mainly due to the model's competence to learn patterns from data and deduce solutions from unknown data. The inclination of the ARIMA models towards the statistical perspective promoted its extensive use in short term forecasting of financial time series data. In the model, the future value of a variable was a linear combination of past values and prior errors.

To determine the best ARIMA model among different experiments performed, Ayodele & Aderemi (2014) used Eviews software and based their results on the smallest value of the Schwarz criterion, small standard error of regression, high adjusted  $R^2$  and Q statistics and correlograms that showed no patterns in both the autocorrelation functions and partial autocorrelation functions of the residuals. Both Nokia and Zenith stock prices were found to be non-stationary and had to be converted to stationary data using differencing in order for the ARIMA model to work. The Augmented Dickey Fuller test was used to confirm stationarity in both data sets. The ARIMA (2, 1, 0) and (1, 0, 1) were selected as the best models for Nokia stock price and Zenith Bank stock price respectively. The results showed very close relations between the actual and predicted values hence showing the model's ability of short term price prediction. Additionally, they stated that the ARIMA models are superior to artificial neural networks when it came to predicting short term prices.

This section entailed the evaluation of past literature on stock prices. The beginning comprised of seminal papers on different factors that impact on stock prices including interest rates,

changes in the level of money supply, brokerage commissions and earnings. Thereafter, there was a shift towards different models that had been used by past researchers to forecast equity prices, as well as ways in which the efficient market hypothesis could be incorporated in these models, which was highlighted by mentioning Granger (1992).

This research hopes to implement the accuracy of the ARIMA model used by Ayodele & Aderemi (2014) in forecasting stock prices by using it in the Kenyan context. The purpose of this will be to create a model that can be used to predict share prices of selected companies listed in the Nairobi Securities Exchange.

## **3. METHODOLOGY**

### **3.1. Empirical strategy**

This project proposal will incorporate a quantitative research design. The research is mainly based on an analysis of historical data. It shall take the form of an autoregressive integrated moving average model of selected company stock prices in Kenya.

The modelling process involves four main steps; model identification, model estimation, diagnostic checking and forecasting which are constantly repeated in order to form a pattern that duplicates the series and attains accurate predictions.

### **3.2. Data collection**

The primary target of this research is to construct a model that will forecast the short term stock prices for selected companies listed in the Nairobi Securities Exchange. Therefore, this study is founded on secondary datasets of returns on Kenyan stock market prices. The data will be retrieved from online sources such as the NSE website and the Valuraha platform.

The stock prices are composed of four main elements; open price, low price, high price and close price. The close price is chosen to be a representative of the price index to be forecasted as it reflects all the activities of the index in a trading day. There are three categories of stock prices used in this research; those that are highly traded, those with high volatilities and those that are highly capitalized. The highly traded stocks are those that are highly demanded by investors and therefore signify a large interest of the market. Stocks with high capitalization are largely affected by market conditions and therefore should the economy change either due to political or investment factors, their level of pricing will change. Finally, stocks with high volatility signify high risk and thus, as investors aim at minimizing risk, they would highly benefit from short term price prediction. Based on this, the analysis will include five stocks; Safaricom, Equity bank, KCB Bank, East African Breweries Limited and Williamson Tea.

### **3.3. Model specification**

This section involves building the ARIMA model through step by step analysis of the autoregressive, moving average and the autoregressive moving average models.

### 3.3.1. Autoregressive Integrated Moving Average/ Box-Jenkins technique

ARIMA methods are founded on statistical concepts and propositions and are able to model a wide area of time series behavior. Choong (2012) stated that they are mainly used in forecasting information such as stock prices, housing starts and company sales. These models mainly aim at establishing the right formulas that will ensure that the errors in a data set are as small as possible and show no trends. The model-building process involves; model identification, model estimation, diagnostic checking and forecasting. ARIMA has different models to pick from and an analytical approach to identifying the right model.

#### 3.3.1.1. Basic Model

There are three types of basic models; autoregressive (AR), moving average (MA), and the autoregressive moving average (ARMA) which are identified after the series has been confirmed to be stationary. When regular differencing is applied to the ARMA it forms the ARIMA model. ARIMA models are based on a single time series variable and their univariate nature is important for several reasons. First, finding variables related to the one being forecasted may be a challenge. Also, the presence of large residuals in a univariate model may correspond to abnormal events. Furthermore, where multivariate methods are available the univariate model provides a yardstick against which the more sophisticated methods can be evaluated. However, in as much as univariate models perform well in the short term, they may be outperformed by multivariate methods at longer lead terms if variables related to the variable being forecast fluctuate in ways which are different to their past behavior.

#### 3.3.1.2. The Mathematical Model

Autoregressive moving average models are described by a set of equations which are mean-adjusted to make them simpler.

$$u(t) = U(t) - \bar{U} \quad (1)$$

Where;  $U(t)$  = Mean-adjusted series;  $u(t)$  = Original time series, and  $\bar{U}$  = the sample mean.

As a subset of the ARMA models, the AR model shows a time series as a function of its prior values, therefore, its order expresses how many lagged prior values are in the series (is a regression model in which  $u(t)$  is a function of past values). It can be linear or nonlinear. Its

simplest form is the AR (1) model, most commonly known as the first-order autoregressive given by:

$$u_{(t)} = a(1) * u_{(t-1)} + e_{(t)} \quad (2)$$

Where;  $u_{(t)}$  = mean-adjusted series in time  $t$ ;  $u_{(t-1)}$  = the series in time  $t - 1$ ;  $a(1)$  = the first lag autoregressive coefficient and;  $e_{(t)}$  = error (Have no autocorrelation and are normally distributed)

Higher-order AR models consist of more lagged values of  $u(t)$  that function as predictors. In the  $n^{th}$  order autoregressive,  $AR(n)$  lagged values exist from time  $t - 1$  to  $t - n$ .

Another subset of the ARMA model is the moving average (MA), in which the time series is an unevenly weighted function of the residuals  $e(t)$ . It includes lagged values on the errors. MA (1) is expressed as:

$$u_{(t)} = e_{(t)} + b(1) * e_{(t-1)} \quad (3)$$

Where;  $e_{(t)}$  = error at time  $t$ ;  $b(1)$  = first order MA coefficient and  $e_{(t-1)}$  = error at  $t - 1$

In an MA, the weights are unequal and do not add up to 1. This is because the number of terms in the model and the weight of each term are statistically determined by the pattern of the data. All MA models are nonlinear.

As a combination of both, the ARMA model is given by ARMA ( $p,q$ ) where  $p$  is the AR order and  $q$  is the MA order. The simplest form of the ARMA model is given as ARMA (1,1) and is expressed as:

$$u_{(t)} = d + a(1) * u_{(t-1)} + e_{(t)} + b(1) * e_{(t-1)} \quad (4)$$

Therefore ARMA ( $p, q$ ) is expressed as;

$$u_{(t)} = d + a(1) * u_{(t-1)} + a(2) * u_{(t-2)} + \dots + a(n) * u_{(t-n)} + e_{(t)} + b(1) * e_{(t-1)} + b(2) * e_{(t-2)} + \dots + b(n) * e_{(t-n)} \quad (5)$$

Where  $d$  is a constant and is calculated by:  $d = z * (1 - \sum_{i=1}^p a(p))$  where  $z$  is the mean of the autoregressive process.

### 3.3.1.3. The modelling process

#### I. Model identification

According to Choong (2012) ARIMA consists of three types of models; the AR, MA and ARMA. The ‘I’ shows that the series has been altered into a stationary time series. Given a stationary time series, we need to first determine the correct model and the number of terms required in the identified model. To achieve this, we could either compute the autocorrelation function (ACF) and the partial autocorrelation function (PACF) or use an automated ARIMA.

#### a) Method 1: Calculating Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF)

**ACF:** ACF values range between  $-1$  and  $1$  computed at different lags to measure the relevance of correlations between the current and previous observations and to assess how far back in time they are correlated. The general formula for the ACF is given as:

$$\hat{\rho}_k = \frac{\hat{y}_k}{\hat{y}_0} = \frac{COV(R_{it}, R_{i,t-k})}{var(R_{it})} \quad (6)$$

The above shows that the autocorrelation coefficient for some lag  $k$  is given by the covariance between the original series and the series without  $k$  lags divided by the variance of the initial series.

**PACF:** PACF values are given by the coefficients of a linear regression of the time series using its lagged values as the explanatory variables. The values also range between  $-1$  and  $1$  for a stationary time series. If the regression inputs only one explanatory variable of a single period lag, the coefficient of the independent variable is the first order PACF, and so on. The general formula is expressed as:

$$\pi_\tau = \frac{\rho_\tau - \sum_{j=1}^{\tau-1} \pi_{\tau-1,j} * \rho_{\tau-j}}{1 - \sum_{j=1}^{\tau-1} \pi_{\tau-1,j} * \rho_{\tau-j}} \quad (7)$$

Where  $\tau > 1$ ,  $\rho$  is the autocorrelation and  $\pi$  is the PACF.

## Using ACF and PACF functions to establish the correct model

A plot of a pair of PACF and ACF is known as a correlogram. A correlogram in which the ACF declines rapidly and the PACF has one large spike, we pick an AR(1) model for the data. If the patterns of the PACF resemble those of the ACF and those of the ACF resemble the PACF having one large spike, we pick an MA(1) model. Moreover, if the PACF in every pair is similar to the ACF, we choose the ARMA (1,1) model.

## Rules for differencing

1. To alter a non-stationary series into a stationary series by differencing, we do not require the constant
2. Time series differenced more than two times do not require a constant
3. If the initial time series is stationary with zero mean, a constant is not necessary, however, the constant is required if the mean is large
  - If the model has no AR component, the constant is equal to the mean value of the series but if the model has an AR component, the constant is given by  $d$

## Checking for zero mean

In cases where the data appears to be stationary, yet its mean is not zero, we have to ensure that the mean is at least close to zero. To do this we compute  $z_t = v_t - \bar{v}$

Where  $v_t$  is the differenced series and  $\bar{v}$  is the average. From the above, we can get the average value of the transformed series  $\bar{z}$ . We then check whether it is zero by estimating its standard error.

$$SE_{(z)} = \frac{\sigma_z}{\sqrt{n}} \quad (8)$$

Thereafter,  $\bar{z}$  is considered nonzero if  $|\bar{z}| < 1.96SE_{(z)}$

Another method of transforming a series into a stationary form is using the equation below:

$$\frac{(u_t - u_{t-1})}{u_{t-1}} \approx \log(u_t) - \log(u_{t-1}) \quad (9)$$

The log transformation could also be used to reduce heteroscedasticity.

## II. Model estimation

After determining the appropriate model, we get the conditional sum of squares of the residual which is conditional on the values of 'a' and 'b'. This is calculated as:

$$SSE(a, b) = \sum_{t=1}^n e(t) \quad (10)$$

Using solver, we determine how closely the fitted values match the original time series.

## III. Diagnostic testing

Here, we determine whether the model reflects the actual time series. One of the conditions is that the error term mean should be zero or at least close to zero. To ensure this, we compute the standard error of the mean error.

$$\sigma_e = \sqrt{\frac{\sum_{t=1}^n (e_t - \bar{e})^2}{n}} \quad (11)$$

$$SE_{\bar{e}} = \frac{\sigma_e}{\sqrt{n}} \quad (12)$$

Where;  $\sigma_e$  is the residual standard error,  $\bar{e}$  is the mean error,  $n$  is the number of errors and  $SE_{\bar{e}}$  is the standard error of the mean error. If  $\bar{e} > 1.96SE_{\bar{e}}$  then it is significantly non-zero.

Another test is using the Durbin-Watson (DW) statistic whose values lie between 0 and 4. It is used to detect the presence of autocorrelation in the data. If the test statistic is 2, then there is no autocorrelation. A value of DW less than 2 indicates positive serial correlation. However, if the values are greater than 2, this shows negative autocorrelation hence an underestimation of the level of statistical significance.

## IV. Forecast.

We are now able to make predictions. The equation can be applied to derive  $u(t)$  from observed  $u(t - 1)$ .

## 4. DATA ANALYSIS

This section entails a detailed explanation of how ARIMA works depending on the stock type (either highly traded, highly capitalized or highly volatile). The sample period varied with the highly traded and capitalized stocks covering the period from 5<sup>th</sup> October, 2010 to 14<sup>th</sup> July, 2016, while that of Williamson Tea covered 24<sup>th</sup> September, 2012 to 24<sup>th</sup> August, 2015. In trying to attain short term prediction using the model, we first carried out diagnostic tests on the data using the Augmented Dickey-Fuller unit root test. Thereafter, we conducted automatic ARIMA forecasting using eviews 9.5 on each data set, to determine the autoregressive and moving average orders (p, q). From our analysis, we obtained the following findings;

### A. Highly traded stocks

#### Safaricom

##### i. Augmented Dickey-Fuller unit root test (ADF test)

The ADF test evaluates the null hypothesis of a unit root in a univariate time series and involves carrying out the following regression;

$$y_t = \beta + \rho y_{t-1} + \delta_t + u_t, \quad (13)$$

Where,  $\beta$  is the constant term and  $\rho$  is the trend or unit root.

In order to account for serial correlation, the ADF test incorporates lags of the first differences of  $y_t$ . We fail to reject the null when the absolute value of the ADF test statistic is lower than any test critical value and vice versa.

*Table 1: Safaricom ADF Test*

		<b>t-statistic</b>	<b>P-value</b>
<b>Augmented Dickey-Fuller test statistic</b>		-16.79232	0.0000
<b>Test critical values</b>	1% level	-3.964873	
	5% level	-3.413150	
	10% level	-3.128588	

*Ho: D(RETURNS) has a unit root*

From our analysis, the ADF test statistic, in absolute terms, is 16.79232 and is greater than each test critical value at all the levels. Therefore, we reject the null hypothesis of the existence of a unit root. In addition the zero probability value shows that there is a 0% chance that we could make an error in estimating the statistic. The result thus establishes that the returns are stationary with the first order of differencing

**ii. Automatic ARIMA forecasting**

This entails selecting the best ARIMA model based on the lowest Akaike Information Criterion (AIC). The AIC estimates the quality of a model relative to others by choosing a model that is well fitted. This ensures that the model does not lose generality or fail to capture the exact nature of the variability in the outcome variable. The best ARIMA model thus chosen for Safaricom was given by ARIMA(2,0,2). The numbers represent the orders of the autoregressive model, differencing and the moving average model respectively from left to right. Formally, we express this as;

$$y_t = \beta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \quad (14)$$

**Table 2: Safaricom Automatic ARIMA forecast**

Dependent Variable: DRETURNS  
Method: Least Squares  
Date: 10/25/16 Time: 22:39  
Sample (adjusted): 10/11/2010 12/31/2015  
Included observations: 1364 after adjustments  
Estimation settings: tol= 0.00010, derivs=analytic (linear)  
MA derivatives use accurate numeric methods  
Initial Values: C(1)=2.3e-06, C(2)=0.00250, C(3)=0.00250, C(4)=0.00250,  
C(5)=0.00250  
Convergence achieved after 19 iterations  
MA Backcast: 10/07/2010 10/08/2010

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.78E-06	3.32E-06	0.536624	0.5916
AR(1)	0.998442	0.050985	19.58307	0.0000
AR(2)	-0.179157	0.027314	-6.559130	0.0000
MA(1)	-1.874767	0.045704	-41.02014	0.0000
MA(2)	0.876062	0.045454	19.27373	0.0000
R-squared	0.444742	Mean dependent var		2.25E-06
Adjusted R-squared	0.443108	S.D. dependent var		0.021277
S.E. of regression	0.015878	Akaike info criterion		-5.444079
Sum squared resid	0.342628	Schwarz criterion		-5.424951
Log likelihood	3717.862	Hannan-Quinn criter.		-5.436919
F-statistic	272.1278	Durbin-Watson stat		2.001170
Prob(F-statistic)	0.000000			
Inverted AR Roots	.76	.23		
Inverted MA Roots	.99	.89		

Table 2 **Error! Reference source not found.** shows the results of estimating the above equation in e-views. The coefficient results are briefly described as follows according to Eviews (2016):

R squared measures the strength of predictability of the values of the dependent variables within the model and ranges between 0 and 1, with values closer to 1 representing a higher explanatory power of the coefficients within the model and therefore high predictability. It is computed as;

$$R^2 = 1 - \frac{SSE}{TSS} \quad (15)$$

Where,  $SSE$  = sum of squared errors and  $TSS$  = total sum of squares.

The adjusted R-squared also works like the  $R^2$  but then adjusts for a number of terms in the model by penalizing additional regressors which do not add the explanatory power of the model. The standard errors measure the statistical consistency of the coefficient estimates. The lower the values, the less the noise in the estimates. A Durbin-Watson statistic is used to measure the level of autocorrelation within the residuals. A DW statistic ranges between 0 and 4 with a value of 2 signifying no autocorrelation in the data while a figure less than 2 signifies positive serial correlation. The F-statistic tests the null hypothesis that all regression coefficients, except the constant, are equal to zero. It is computed as;

$$F = \frac{R^2/(k-1)}{(1-R^2)/(T-k)} \quad (16)$$

Where,  $T$  is the number of observations and  $k$  is the number of regressors.

To test the significance of the f test, we examine the probability value below the statistic. The null is rejected if the p-value is less than the specified level of significance such as 0.1 or 0.05. The log likelihood function is estimated at the value of the coefficients and is used to calculate both the Schwarz Criterion (SC) and the Hannan-Quinn (HQ) Criterion. These are used to impose penalties for additional coefficients into the model, with the SC imposing the larger penalty. They therefore serve as criteria for model selection among a finite set of models and are given by:

$$SC = -2 \frac{l}{T} + \frac{(k \log T)}{T}, \text{ and } HQ = -2 \frac{l}{T} + 2k \log \frac{(\log(T))}{T} \quad (17)$$

Where,  $l$  is the log likelihood function.

From the above, our analytical study on the models is based on the R-squared, adjusted R-squared, F- statistic and the Durbin Watson statistic.

From Table 2, all the autoregressive and moving average coefficients were significant as they all recorded zero probability values. There was also no autocorrelation in the data as the DW statistic was 2.001170. However, the low r-squared value of 44.47% indicates a low explanatory power of the model and a probability of not attaining precise predictions. Nonetheless, given this factor, the sampling variation when estimating the data will only be 1.59% as seen from the standard error of regression.

## Equity Bank

### i. Unit root test

*Table 3: Equity returns ADF test*

		<b>t-statistic</b>	<b>P-value</b>
<b>Augmented Dickey-Fuller test statistic</b>		-16.76918	0.0000
<b>Test critical values</b>	1% level	-3.964857	
	5% level	-3.413143	
	10% level	-3.128584	

*Ho: D(RETURNS) has a unit root*

As seen above, returns from equity's stock prices become stationary after 1<sup>st</sup> level differencing. This is because the ADF statistic in absolute terms is 16.76918 and is greater than all the test critical values. In addition, it is also statistically significant given the zero probability value.

### ii. Automatic ARIMA forecast

Dependent Variable: DRETURNS  
Method: Least Squares  
Date: 11/03/16 Time: 09:05  
Sample (adjusted): 10/08/2010 12/31/2015  
Included observations: 1365 after adjustments  
Estimation settings: tol= 0.00010, derivs=analytic (linear)  
MA derivatives use accurate numeric methods  
Initial Values: C(1)=-4.4e-06, C(2)=0.00250, C(3)=0.00250  
Convergence achieved after 40 iterations  
MA Backcast: OFF (Roots of MA process too large)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.99E-06	3.22E-06	-1.239148	0.2155
AR(1)	0.241983	0.027037	8.949967	0.0000
MA(1)	-1.007213	0.005163	-195.0671	0.0000
R-squared	0.426945	Mean dependent var		-4.37E-06
Adjusted R-squared	0.426103	S.D. dependent var		0.021765
S.E. of regression	0.016489	Akaike info criterion		-5.370111
Sum squared resid	0.370288	Schwarz criterion		-5.358641
Log likelihood	3668.101	Hannan-Quinn criter.		-5.365818
F-statistic	507.3669	Durbin-Watson stat		2.170364
Prob(F-statistic)	0.000000			

**Table 4: Equity automatic ARIMA forecast**

An automatic ARIMA forecast of Equity returns in e-views allowed us to adopt an ARIMA(1,0,1). This shows that the autoregressive, differencing and moving average orders were 1, 0 and 1 respectively.

## B. Highly capitalized stocks.

### KCB Bank

#### i. Unit root test

**Table 5: KCB Bank ADF test**

	t-statistic	P-value
<b>Augmented Dickey-Fuller test statistic</b>	-17.31657	0.0000
<b>Test critical values</b>		
	1% level	-3.964847
	5% level	-3.413138
	10% level	-3.128581

***Ho: D(RETURNS) has a unit root***

We reject the null hypothesis of the presence of a unit root in the series as the ADF test statistic in absolute terms, 17.31657, is greater than all the test critical values.

#### ii. Automatic ARIMA forecast

**Table 6: KCB automatic ARIMA forecast.**

Dependent Variable: DRETURNS  
Method: Least Squares  
Date: 10/25/16 Time: 22:56  
Sample (adjusted): 10/14/2010 12/31/2015  
Included observations: 1361 after adjustments  
Estimation settings: tol= 0.00010, derivs=analytic (linear)  
MA derivatives use accurate numeric methods  
Initial Values: C(1)=4.0e-20, C(2)=0.00250, C(3)=0.00250, C(4)=0.00250,  
C(5)=0.00250, C(6)=0.00250, C(7)=0.00250, C(8)=0.00250,  
C(9)=0.00250  
Convergence achieved after 27 iterations  
MA Backcast: 10/11/2010 10/13/2010

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-8.95E-07	1.54E-06	-0.582577	0.5603
AR(1)	-0.840188	0.037943	-22.14369	0.0000
AR(2)	-0.664182	0.041595	-15.96786	0.0000
AR(3)	0.170991	0.039904	4.285010	0.0000
AR(4)	-0.098481	0.035932	-2.740729	0.0062
AR(5)	-0.084809	0.028120	-3.016020	0.0026
MA(1)	0.085032	0.026389	3.222311	0.0013
MA(2)	-0.130716	0.024684	-5.295506	0.0000
MA(3)	-0.949359	0.025083	-37.84898	0.0000
R-squared	0.389134	Mean dependent var	5.61E-20	
Adjusted R-squared	0.385519	S.D. dependent var	0.017608	
S.E. of regression	0.013803	Akaike info criterion	-5.721319	
Sum squared resid	0.257574	Schwarz criterion	-5.686827	
Log likelihood	3902.358	Hannan-Quinn criter.	-5.708407	
F-statistic	107.6564	Durbin-Watson stat	2.005304	
Prob(F-statistic)	0.000000			
Inverted AR Roots	.32-.37i	.32+.37i	-.37	-.56+.81i
Inverted MA Roots	-.56-.81i	1.00	-.54+.81i	-.54-.81i

An automatic ARIMA forecast on the company's stock price returns portrayed ARIMA(5,0,3) as the best model for forecasting. This exhibits that the autoregressive, differencing and moving average orders were 5, 0 and 3 respectively. Thus in a forecasting form, the model would be expressed as;

$$y_t = \beta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \phi_4 y_{t-4} + \phi_5 y_{t-5} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} \quad (18)$$

All the variables recorded excluding the constant were significant at 5% as is portrayed by the minimal probability values of the coefficients. The r-squared recorded was lower than that of Safaricom. The value achieved at 0.389134 implies that 38.9% of the sampling variation will be explained by the model while the other percentage remains unexplained. This explanatory power is supported by the significant probability value of the F-statistic. There was no serial correlation within the model, which was depicted by the DW statistic of 2.005304.

## East African Breweries Limited

### i. Unit root test

*Table 7: EABL unit root test*

	t-statistic	P-value
<b>Augmented Dickey-Fuller test statistic</b>	-17.78896	0.0000

<b>Test critical values</b>	1% level	-3.964842	
	5% level	-3.413135	
	10% level	-3.128580	

***Ho: D(RETURNS) has a unit root***

The series became stationary after having been differenced once. As seen above, the ADF test statistic is statistically significant at both 5% and 1%, revealing that no error can be made in estimating it. Also, the test statistic in absolute terms is greater than all the test critical values and thus we reject the null.

**ii. Automatic ARIMA forecast**

An automatic ARIMA forecast revealed the best model as ARIMA(1,0,1) based on the lowest Akaike Information Criterion value which was -5.672158. The model if formally expressed as;

$$y_t = \beta + \phi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1} \tag{19}$$

We note that all the orders of the AR and MA processes were significant with extremely minimal probability values and standard errors. However, the probability of the constant was greater than 0.05. The standard error of regression shows that there is a 1.42% chance of sampling variation when estimating the data. From the r-squared, only 38.45% of the variations in the data will be explained by the model and as such the predictions will not be exact. Nonetheless, the model does not suffer from autocorrelation as the Durbin-Watson statistic is stated at 2.022129.

***Table 8: EABL automatic ARIMA forecast***

Dependent Variable: DRETURNS  
Method: Least Squares  
Date: 10/25/16 Time: 23:20  
Sample (adjusted): 10/08/2010 12/31/2015  
Included observations: 1365 after adjustments  
Estimation settings: tol= 0.00010, derivs=analytic (linear)  
MA derivatives use accurate numeric methods  
Initial Values: C(1)=1.4e-19, C(2)=0.00250, C(3)=0.00250  
Convergence achieved after 15 iterations  
MA Backcast: 10/07/2010

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.02E-06	1.07E-06	-0.948952	0.3428
AR(1)	0.228454	0.026320	8.679739	0.0000
MA(1)	-0.999594	0.001108	-902.5526	0.0000
R-squared	0.384523	Mean dependent var	-6.10E-20	
Adjusted R-squared	0.383620	S.D. dependent var	0.018058	
S.E. of regression	0.014177	Akaike info criterion	-5.672158	
Sum squared resid	0.273755	Schwarz criterion	-5.660688	
Log likelihood	3874.248	Hannan-Quinn criter.	-5.667865	
F-statistic	425.4596	Durbin-Watson stat	2.022129	
Prob(F-statistic)	0.000000			
Inverted AR Roots	.23			
Inverted MA Roots	1.00			

### C. Highly volatile stocks

#### Williamson Tea

##### i. Unit root test.

After 1<sup>st</sup> level differencing of Williamson Tea's stock price returns, we reject the null of no stationarity, i.e. presence of a unit root. This is because the ADF test statistic at 14.59279 is greater than all the test critical values at all the levels.

**Table 9: Williamson Tea unit root test**

	<b>t-statistic</b>	<b>P-value</b>
<b>Augmented Dickey-Fuller test statistic</b>	-14.59279	0.0000
<b>Test critical values</b>		
	1% level	-3.973874
	5% level	-3.417546
	10% level	-3.131192

*Ho: D(RETURNS) has a unit root*

##### ii. Automatic ARIMA forecast.

**Table 10: Williamson Tea automatic ARIMA forecast**

Dependent Variable: DRETURNS  
Method: Least Squares  
Date: 10/25/16 Time: 23:46  
Sample (adjusted): 9/28/2012 12/31/2014  
Included observations: 589 after adjustments  
Estimation settings: tol= 0.00010, derivs=analytic (linear)  
MA derivatives use accurate numeric methods  
Initial Values: C(1)=5.7e-05, C(2)=0.00250, C(3)=0.00250, C(4)=0.00250,  
C(5)=0.00250, C(6)=0.00250, C(7)=0.00250  
Convergence achieved after 33 iterations  
MA Backcast: 9/24/2012 9/27/2012

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.30E-05	1.23E-05	1.059094	0.2900
AR(1)	0.735581	0.301196	2.442198	0.0149
AR(2)	-0.149386	0.127499	-1.171663	0.2418
MA(1)	-2.267699	0.300076	-7.557093	0.0000
MA(2)	2.111415	0.466823	4.522942	0.0000
MA(3)	-1.126312	0.281295	-4.004022	0.0001
MA(4)	0.284143	0.097938	2.901262	0.0039
R-squared	0.759070	Mean dependent var		5.67E-05
Adjusted R-squared	0.756587	S.D. dependent var		0.104314
S.E. of regression	0.051465	Akaike info criterion		-3.084007
Sum squared resid	1.541526	Schwarz criterion		-3.031972
Log likelihood	915.2402	Hannan-Quinn criter.		-3.063734
F-statistic	305.6073	Durbin-Watson stat		1.989736
Prob(F-statistic)	0.000000			
Inverted AR Roots	.37-.12i	.37+.12i		
Inverted MA Roots	.99	.65	.31-.59i	.31+.59i

The equation estimated above is expressed as;

$$y_t = \beta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \theta_4 \varepsilon_{t-4} \quad (20)$$

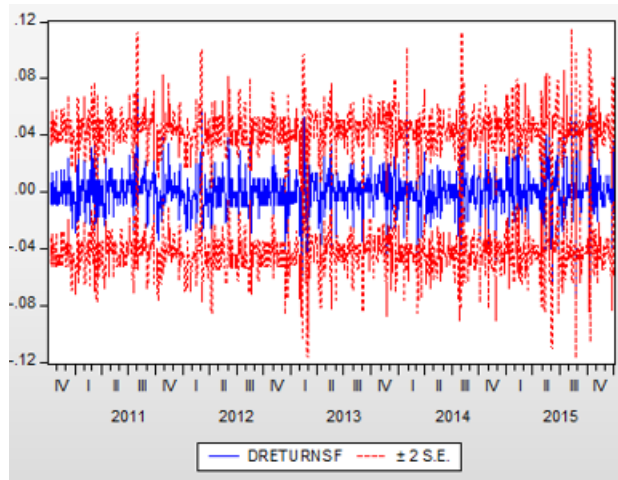
This particular stock recorded the highest r-squared value at 0.75907 value and the highest standard error of regression at 0.051465 once we estimated its ARIMA(2,0,4) equation. This means that out of the three groups of stock, this would record the highest sampling variation at 5.15% and of these variations, 75.9% would be explained by the model. The low p-values show that the t-statistics of the coefficients were significant. However, that of the constant and the AR(2) were insignificant at a 5% level of significance. The Durbin-Watson statistic indicated that the errors in the data were not related across different periods since the DW statistic of 1.9897 is extremely close to 2. The F-statistic probability is statistically significant as it is less than 0.05 hence we reject the null hypothesis that all regression coefficients are zero.

## 4.1. FINDINGS

In this section, we examine the Goodness of fit of the models basing on their values of bias proportions, variance proportions and covariance proportions. This is to show how well our in sample observations can be used as prediction proxies for the out of sample periods of each of the stocks. The bias proportion tells us the difference of the mean between the actual and the

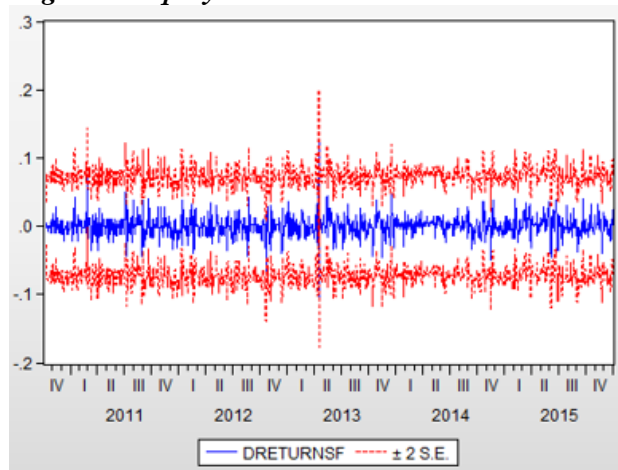
forecasted series. The variance proportion shows the extent of variations between the predicted and actual series while the covariance proportion measures the remaining unsystematic forecasting errors. All the proportions add up to one and a model is considered to have good forecasting abilities if both the bias and variance proportions are as small as possible and the noise is concentrated around the covariance proportion, (Eviews, 2016).

**a. Highly traded stocks.**



Bias proportion = 0.000111  
 Variance proportion = 0.211102  
 Covariance proportion = 0.7888786

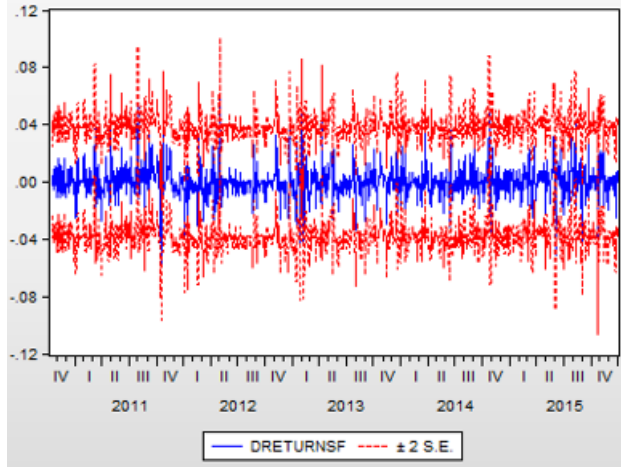
*Figure 1: Equity Bank*



Bias proportion = 0.000000  
 Variance proportion = 0.199922  
 Covariance proportion = 0.800078

*Figure 2: Safaricom*

**b. Highly capitalized stocks**

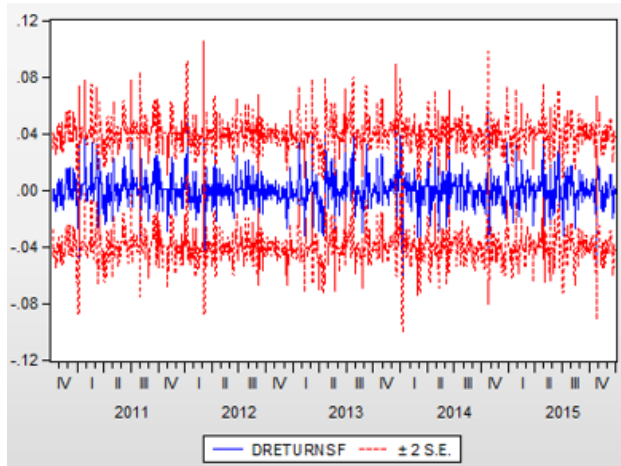


Bias proportion = 0.000058

Variance proportion = 0.232169

Covariance proportion = 0.76773

**Figure 3: KCB Bank**



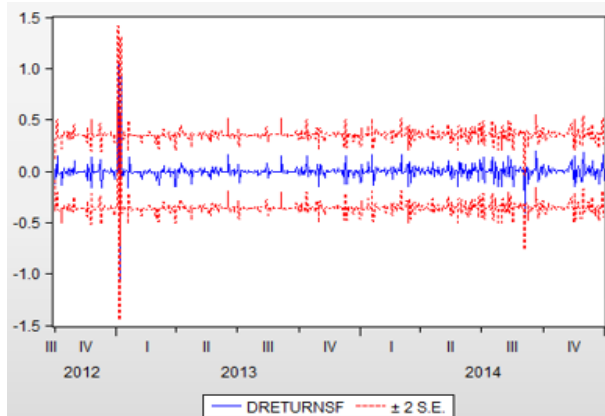
Bias proportion = 0.000299

Variance proportion = 0.232689

Covariance proportion = 0.767012

**Figure 4: EABL**

**c. Highly volatile stocks**



**Figure 5: Williamson Tea**

Bias proportion = 0.000357

Variance proportion = 0.072431

Covariance proportion = 0.927212

The results above thus show that the models generated from the in-sample periods could be used for forecasting purposes owing to their low bias and variance proportions and high covariance proportions. The model with the highest covariance proportion was that of Williamson Tea which recorded a value of 0.927212, and this goes to prove its robustness in forecasting, which was also shown by its high r squared value. From the figures we can also tell when the values dipped most to when the stock returns recorded their highest values. The strength of the models is also shown by the minimal variations computed by getting the difference between the actual and predicted values shown appendix a.

## **5. DISCUSSION**

Stock markets can either be order driven, quote driven or hybrid systems (a mixture between the two). In an order driven market, interested buyers and sellers submit bid and ask prices for a given stock to a central location where the orders are matched by a broker who does not own the stock but act as a facilitating agent. Prices are sold to the investor with the highest bid price and bought from the seller with the lowest offering price. In case the securities are neither bought nor sold, the broker does not bear the risk as long as there exists no other party with a similar deal, Robert et al (2014). In this market, prices are determined by the forces of demand and supply.

On the other hand, in a quote driven market, individual dealers provide liquidity for investors by buying and selling the shares of stock themselves. In this system there will be numerous dealers who will compete against each other to provide the highest bid prices when you are selling and the lowest asking price when you are buying stock, where the difference is the profit margin, Robert et al (2014) .It is a very decentralized system that derives its benefit from competition among the dealers to provide the best price for the buyer or seller.

The Nairobi Securities Exchange consists of buyers, sellers and brokers who act as intermediaries between the two parties. The exchange is order-driven, Ngugi (2003), as the various market participants enter their orders during the given trading sessions i.e. pre-open and open session. The buy and sell orders determine the opening prices, after which they are matched during trading hours following fixed rules and the closing prices are set. The demand and supply forces that drive prices in the NSE are as a result of the economic performance in the country owing to variables such as interest rates, company profitabilities and tax conditions.

Panda (2008) showed that high interest rates lower investor expectations on a company's earnings and lower demand for stocks relative to that of bonds. Due to the high bargaining power of buyers, funds are shifted from the stock market to the bond market and the stock market indices for companies listed in the NSE lose points, Kitati et al (2015). In this case the bid prices stated by buyers affect the final execution prices of the securities that will be given by the brokers. Conversely, high firm profitability translates to increased company stock prices, Panda (2008). This is because of the high bargaining power of the sellers who then determine the prices to be stated by the brokers and thus the final market value.

This research uses the set closing prices in its analysis of the five companies and the results show an element of mispricing in each of the stocks, as the predicted values differ from those revealed by the market and hence the variations which portray existence of arbitrage opportunities in the given companies' stock price returns.

From the analysis, the highly traded and highly capitalized companies experienced higher levels of predictability from past prices as compared to the volatile stock. This implies that these stocks do not follow either the efficient market hypothesis or the random walk theory. The minimal variations within the model could be attributed to the informationally inefficient stock prices. The inefficiency is as a result of the low investor confidence and awareness, low level of capital market liquidity, a minimal competitive pressure in the local market due to few products and unavailability of products such as derivatives and options and market vulnerability to shocks, Sheila & Odhiambo (2014), which inhibit the speed at which new information is reflected in prices. Moreover, lack of awareness by investors promotes information asymmetry between the buyers, sellers and brokers prior to a transaction taking place that could provide for exploitation of arbitrage opportunities.

Both KCB and EABL are highly capitalized firms, however, EABL experiences higher sampling variations between its actual and predicted values. Therefore it has both windows of predictability and non-predictability. This is as a result of macroeconomic variables such as interest rates having different impacts on the companies' performance levels due to their operations in different industries. KCB is in the Kenyan banking industry while EABL is one of the main players in the beverage industry. Williamson Tea recorded the highest variation values hence depicting that high volatility stocks portray an element of randomness.

## **6. CONCLUSION**

Stock price prediction has always been a major topic in economics. In the past, a lot of models were used to forecast prices and they included the Monte-Carlo simulation and neural networks. In trying to come up with future prices our main objective was set on creating a model that can be used for short term prediction of selected company stock prices within the Nairobi Securities Exchange, in this case the ARIMA model.

Our findings were in line with past literature such as (Dyakove, 2014) which pointed out that Kenya is one of the most informationally inefficient markets and therefore, stock price predictability within the economy is high. However, this argument does not necessarily apply to Williamson Tea which had the highest sampling variations owing to its volatile nature and is therefore efficient in the weak form. The analysis thus signified the strength of the ARIMA model in short term prediction basing on stocks that are highly volatile and also those that are highly capitalized.

## **7. RECOMMENDATIONS FOR FURTHER STUDIES**

The above model could be strengthened by the inclusion of other variables that affect stock prices, however, frequencies of factors such as dividend yields and macroeconomic variables such as Growth Domestic Product and exchange rates are either monthly, quarterly or annual which would not match daily stock price data. Moreover, conversion of daily to monthly stock price data will only smooth out the data which would equally affect the analysis. Therefore, more analytical work could be done to create a model that merges the above specifications.

## APPENDIX

*Appendix a: Variations between the actual and forecasted stock price values.*

PRICE VARIATIONS BETWEEN ACTUAL AND PREDICTED STOCK PRICE VALUES				
HIGHLY TRADED		HIGHLY CAPITALIZED		HIGHLY VOLATILE
SAFARICOM	EQUITY	KCB	EABL	WILLIAMSON TEA
0.08	0.00	0.34	-0.13	-4.86
-0.16	0.00	0.11	0.87	23.15
0.00	-0.55	0.28	-0.87	-2.45
0.06	0.30	-0.15	-0.13	-30.32
0.12	0.25	0.08	1.87	122.49
-0.06	-2.61	-0.35	0.87	-22.11
0.30	0.75	0.22	-0.89	-22.16
0.11	0.79	0.04	4.97	-23.56
0.06	0.04	0.75	8.25	12.46
-0.09	-1.04	-0.44	-2.85	-15.47
-0.08	-0.54	-0.41	2.08	-20.23
0.37	0.78	-0.10	1.95	-22.76
-0.16	-0.25	0.17	0.89	9.32
-0.03	0.25	0.43	3.84	17.30
-0.12	0.25	-0.43	-5.14	5.66
0.12	-0.05	0.49	-0.11	1.09

The variations above were computed by getting the difference between the actual and predicted values.

## REFERENCES

1. Allan Timmermann\*, C. W. (2004). Efficient market hypothesis and forecasting. *International Journal of Forecasting* 20, 15-27.
2. Ayodele A. Adebisi, A. O. (2014). Stock Price Prediction Using the ARIMA Model. *16th International Conference on Computer Modelling and Simulation*, 105-110.
3. Brooks, C. (2014). *Introductory Econometrics for Finance*. Cambridge: MPG Printgroup Ltd.
4. Choong, J. (2012). *Powerful forecasting with MS Excel*.

5. Dyakove, G. S. (2014). African stock markets; Efficiency and relative predictability. *South African Journal of Economics*.
6. Edward A. Dyl, W. B. (2006). The share price puzzle. *The Journal of Business*, 2045-2066.
7. Edward Kitati, E. Z. (2015). Effect of Macro-Economic Variables on Stock Market Prices for the Companies Quoted on the Nairobi Securities Exchange in Kenya. *International Journal of Science; Basic and Applied research*, 235-263.
8. Eugene F. Brigham, M. C. (2005). *Financial Management: Theory and Practice*. South Western College.
9. Eugene F. Fama, L. F. (1969). The adjustment of stock prices to new information. *International Economic Review*, 1-21.
10. Eugene F. Fama, L. F. (1969). The Adjustment of stock prices to new information. *International Economic Review*, 1-21.
11. *Eviews*. (2016, November 3). Retrieved from Eviews website: [http://www.eviews.com/help/helpintro.html#page/content/Forecast-Forecast\\_Basics.html](http://www.eviews.com/help/helpintro.html#page/content/Forecast-Forecast_Basics.html)
12. *Eviews website*. (2016, November 3). Retrieved from User's Guide : Basic Single Equation Analysis : Basic Regression Analysis : Equation Output: [http://www.eviews.com/help/helpintro.html#page/content/Regress1-Equation\\_Output.html](http://www.eviews.com/help/helpintro.html#page/content/Regress1-Equation_Output.html)
13. F. Fama, E. (1965). The Behavior of Stock-Market Prices. *The Journal of Business*, 34-105.
14. F. King, B. (1966). Market and industry factors in stock price behavior. *The Journal of Business*, 139-190.
15. Fisher, G. (1961). Factors influencing share prices. *The Economic Journal*, 121-141.
16. Gatua, F. K. (2013). Analysis of share price determinants at Nairobi Securities Exchange. *University of Nairobi repository*.

17. Granger, C. (1992). Forecasting stock market prices: Lessons for forecasters. *International Journal of Forecasting* 8, 3-13.
18. Kuo-Cheng Tseng, O. K. (2012). Time series and neural network forecast of daily stock prices. *Investment Management and Financial Innovations*, 32-54.
19. Mahdi Naeini, H. T. (2010). Stock market value prediction using neural networks. *International Conference on Computer Information Systems and Industrial Management Applications*, 132-136.
20. Mandelbrot, B. (1966). Forecasts of Future Prices, Unbiased Markets and Martingale Models. *Journal of Business*, 242-255.
21. N.M.Odhiambo, S. N. (2014). The Dynamics of stock market development in Kenya. *The Journal of Applied Business Research*, 73-82.
22. Nderi, T. M. (2009). Firm Specific determinants of Underpricing IPOs in the Nairobi Securities Exchange. *University of Nairobi Repository*.
23. Ngugi, R. W. (2003). Development of the Nairobi Stock Exchange: A historical perspective. *KIPPRA Discussion paper No.27*, 7-11.
24. Omollo, C. &. (2010). Market reaction to stock splits. *African Journal of Business & Management*.
25. Panda, C. (2008). Do interest rates matter for stocks? *Economic and political weekly*, 107-115.
26. Parkinson, J. (1984). The Nairobi Stock Exchange in the context of development of Kenya. *Savings and Development*, 363-372.
27. Peavy, J. W. (1992). Do interest rates and earnings really matter? *Financial Analysts Journal*, 10-12.
28. Qasem A. Al-Radaideh, A. A. (2013). Predicting stock prices using data mining techniques. *The international Arab conference on Information Technology*, 1-8.

29. Robert Silikhe Simiyu, S. M. (2014). The effect of automation on trading volumes on listed companies at the Nairobi Securities Exchange, Kenya. *Research Journal of Finance and Accounting*, 21-33.
30. Sen, P. (1974). Stock Price Level: A predictive analysis. *Economic and political weekly*, 81-86.
31. Yalla, D. (2014). Determinants of stock price behaviour after initial public offer for firms quoted at the NSE. *University of Nairobi repository*.