Journal of Economics Bibliography
www.kspjournals.org
Volume 9 March 2022 Issue 1

Forecasting stock prices in the New York stock exchange

By Huboh S. RINGMU [a†] Saidou B. OUMAR [b]

Abstract. Stock price prediction is one of the most relevant aspects in a stock market and world economies. Price is an important variable of concern where this sector since economic and market conditions vary over time. Efficient methods are needed to describe the trends and characteristics of stock prices. The performance of different time series models for analysis of stock prices is provided to determine the feasibility of techniques for the generation of results in the wake of economic decisions. Historical time series of monthly average price of stocks for Callon, Chesapeake, General Electric and Encana in the oil and gas sector of the New York Stock Exchange were analysed for the period 2012-2019. It was discovered that the New York Stock Exchange follows a random walk. A random walk implies uncertainty. Uncertainty implies high risk. Risk is directly related to profit. The fitted autoregressive integrated moving averages model used for forecasting shows that the predicted average stock price for the period 2021-2024 for Callon, Chesapeake, General Electric and Encana may reach United States Dollars 12.14, United States Dollars 7.34, United States Dollars 21.69 and United States Dollars 23.90, respectively. Therefore, cautious trading in the New York Stock Exchange was recommended for high profits to be achieved.
Keywords. Convergence; Integration; Modelling; Stationarity; Variation.
JEL. C5, C22, C32, E27, E32.

1. Introduction

Forecasting or predicting is an important exercise to any business entity. It is like a chainsaw time series that is easy to use, hard to use well, and always dangerous (Saylors, 2017) to implement. A time series describes values of a variable that are taken over time. These values may represent stock prices that are recorded weekly, monthly, quarterly or yearly. The making of statements against future possibilities that are unknown at the time of assertion is usually referred to as forecasting. The various techniques used to model the mean daily stock price include empirical regression lines, neutral network graphs, and time series autoregressive integrated moving averages (ARIMA) also referred to as Box-Jenkins (2013) models. Time series aims at using past observations of a series of events to develop an appropriate model which describes the

[a†] Department of Banking and Finance, Faculty of Economics and Management Sciences, The University of Bamenda, Cameroon.
☎. +237674371549 ✉. hubohsamuel@gmail.com
[b] Department of Economics, Faculty of Social and Management Sciences, University of Buea, Cameroon.
☎. +237670172520 ✉. saidoubo@gmail.com

inherent structure of a series of events. This model is then used to predict future values for the series.

Many stock exchange markets exist across the globe. The 8 smallest stock exchanges in ascending order in the world include: Seychelles Securities Exchange (SSE) with a market capitalisation of United States Dollars (US$) 43 million, Cambodia Stock Exchange (CSE) worth US$176 million market capitalisation, Douala Stock Exchange (DSE) with a market capitalisation of US$317 million, Rwanda Stock Exchange (RSE) with a market capitalisation of US$480 million. Others are: Damascus Securities Exchange (DaSE) worth US$660 million market capitalisation, The Lao Securities Exchange (LaSE) with a market capitalisation of US$1.4 billion, Bolsa de Valores Stock Exchange (BSE) with a market capitalisation of US$1.5 billion, and Cyprus Stock Exchange (CySE) worth US$2.54 billion market capitalisation (Shum 2016).

The Nigerian Stock Exchange (NSE), founded in 1960 as the Lagos Stock Exchange (LaSE) provides access to capital for Nigerian companies seeking to take advantage of the capital market to fund their business expansion. Since the deregulation of the Nigerian capital market in 1993, foreign capital has freely been allowed to flow into the country, enabling investors of all nationalities to participate in the capital market. Also, to enhance liquidity, NSE has operated an automated trading system since 1999, allowing dealers to trade live through a network of computers connected to its central terminal. Some listed companies in NSE include: African Alliance Insurance, Beta Glass Company and Computer Warehouse Group. The market capitalisation of listed domestic companies in Nigeria was worth US$3.15 billion in 2018 (Shum, 2016).

Tokyo Stock Exchange (TSE) located in Japan is the third largest stock exchange in the world by aggregate market capitalisation. Its operating hours are from 8:00 to 11:30 a.m. and from 12:30 to 5:00 p.m. From April 24, 2006, the afternoon trading session started at its usual time of 12:30 p.m. Established on May 15, 1878, TSE has close to 3500 listed companies, with a combined market capitalisation in 2018 of more than US$4 trillion. It is the largest and best known Japanese giant with a global presence, including Toyota, Honda, and Mitsubishi. In March 2018, the 5 largest stocks listed on TSE included: Toyota Motor Corporation [¥222.6 million], Do Communications over the Mobile Network Company [¥105.931 million], Mitsubishi Financial Group [¥96.883 million], Softbank Group Corporation [¥87.5 million] and Keynence Corporation [¥80.3 million] (Chen, 2018).

Trading approximately US$1.46 billion shares every day, the New York Stock Exchange (NYSE) is the leading stock exchange in the world. The institution trades stocks for 2800 companies. Operating as a continuous auction floor trading stock exchange, the major players on the floor of NYSE are specialists and brokers (Chen, 2018). Brokers are employed by investment firms and trade either on behalf of their firm's clients or the firm. The broker brings buy and sell orders to the specialists. Each specialist stands in one location on the floor and deals in one or several specific stocks, depending on their trading volume. The specialists' job is to accept buy and sell orders from brokers and manage the actual auction. It is also the specialists' job to ensure that there is a market for their specified stocks at all times, meaning that they invest their own firm's capital at times to keep the market active and maintain the shares' liquidity. Specialists and brokers interact to create an effective system that provides investors with competitive prices based on supply and demand (Schmidt, 2011).

The variability of stock prices is one important challenge limiting the use of structuring equation modeling (SEM). In addition, increases in the risk and uncertainty of generating the expected stock price could function as an inhibiting factor to maintaining a certain level of profit (Sindhu *et al.*, 2017). Besides, competent forecasting and careful analysis of stock price can help reduce risk. By applying accurate stock price forecasting techniques, stock price variability patterns can be traced, thus boasting the confidence of investors to participate in the trading exercise. Accurate stock price forecasting may also exert a significant economic impact on stock market operations and substantially reduce the costs of participating in the stock exchange (Park, 2016). Drawing from these phenomena, this study mainly seeks to predict stock prices in the oil and gas sector of NYSE using ARIMA model. The specific objectives of the paper are:

- To compare the forecasting efficiency of 4 different models using model selection criteria.
- To fit NYSE stock price in an ARIMA model and forecast future stock prices.
- To establish the pattern of the stock price in NYSE.
- To offer recommendations towards improving the activities of stock market participants.

The rest of the paper is organised as follows. Following the introduction in Section 1, Section 2 discusses the review of literature by exploring studies carried out on forecasting across the globe. Section 3 explains the methodology adopted in order to achieve the objectives of the inquiry. Then, Section 4 trails with data analysis and discussion by addressing the specific objectives of the investigation, while Section 5 offers the conclusion and recommendations of the study based on the findings of the inquiry.

2. Literature Review

Peijun (2016) analysed and predicted China's total health expenditure. His results showed that the ARIMA (5, 0, 1) model is the most appropriate to forecast total expenditure as a percentage of gross domestic product. Yue *et al.* (2015) used ARIMA models to analyse and forecast health expenditure, and in particular, hospital costs for respiratory illnesses in Shanghai, China. The monthly data from January-December 2012 used in the exercise showed that ARIMA (0, 1, 1) model proved to be the best for this projection. Adebayo (2014) used time series model to forecast stock market prices in Botswana and Nigeria. They considered the problem of selecting best ARIMA models for stock market price prediction for Botswana and Nigeria. Based on the standard model selection, 26 criteria such as AIC, BIC, RMSE and MAE were tested to evaluate the forecasting performance of different ARIMA models with a view to determine the best ARIMA model for predicting stock market prices in the 2 countries. As a result, it was found that ARIMA (3, 1, 1) and ARIMA (1, 1, 4) models were the best forecast models for Botswana and Nigeria stock markets, respectively.

Manoj & Madhu (2014) studied the forecast for sugarcane production in India. The best ARIMA model was found to be ARIMA (2, 1, 0). The forecast results showed that the annual sugarcane production was to grow in 2013, before falling in 2014, 2015 and 2017. It continuously grows with an average growth rate of approximately 3%. Okany (2014) studied the relationship between oil prices and stock prices in the Nigerian equity market using a forecasting framework. The study was driven by the need to determine whether or not the extreme volatility observed in oil prices has any significant impact on the stock price movement of a major oil exporting economy like that of Nigeria. By establishing the presence of a significant relationship between these variables, investors and policy makers alike could use oil prices as a leading indicator in producing more accurate projections of stock prices. While the results of the study recorded no cointegration between stock prices and oil prices, the use of an ARIMA and a structural ARIMA model showed that oil price is a significant exogenous variable which could improve the accuracy of stock price prediction in the Nigerian stock market to an extent.

Abdel-Aal & Al-Garni (1997) used univariate Box-Jenkins (1997) time series analysis to model and forecast monthly domestic electric energy consumption in the Eastern Province of Saudi Arabia. They found that, compared to regression and adductive network, machine learning models previously developed on the same data required less data. They had fewer coefficients, and were more accurate.

Tang *et al.* (1991) discussed the results of a comparative study of the performance of neutral networks and conventional methods of forecasting time series. Their work was initially inspired by previously published works that yielded inconsistent results about comparative performance. Experimenting with 3 time series of different complexity using different feed forward, back propagation neural network models and the standard Box-Jenkins (1976) model, they demonstrated that for time series with long memory, both methods produced comparable results. However, for series with short memory, neural networks outperformed the Box-Jenkins (1976) model. They noted that some of the comparable results arose since the neural network and time series model appeared to be functionally similar models. They found that for time series of different complexities, there were optimal neutral network topologies and parameters that enabled them to learn more efficiently. Their initial conclusions were that neutral networks were robust and provided good long term forecasting. Neutral networks represented a promising alternative for forecasting, but there were problems determining the optimal topology and parameters for efficient learning.

3. Methodology

Two approaches can be used to forecast price movement in the future. These include the intrinsic value analysis and technical analysis. Technical analysis is based on 3 assumptions, that history tends to repeat itself, the market discounts everything and the price moves in trends. Therefore, to predict stock prices is to develop acquaintance with past patterns of price behaviour in order to recognise situations of likely recurrence (Ababio, 2012).

To a fundamental investor, the market price of a stock tends to revert towards its intrinsic value. If the intrinsic value of a stock is above the current market price, the investor would purchase the stock because he believes that the stock price would rise and move towards its intrinsic value. If the intrinsic value of a stock is below the market price, the investor would sell the stock because he believes that the stock price is going to fall and come closer to its intrinsic value.

Much efforts have been made by researchers on forecasting with ARIMA models. As a result, various important time series forecasting models have evolved. To make a good forecast, data for 5 years and above are needed (Rahman, 2015). The severe limitation of ARIMA models is the presumed linear form of the associated time series which becomes inadequate in many practical situations. To overcome this drawback, various nonlinear stochastic models have been proposed. Another limitation is the fact that, not all forecasting models could fit the data under study because there are many of such models that could fit the data considered for the study.

There are many forecasting methods in projecting price movement of stocks: Box-Jenkins (1972) method, simple linear regression, and multiple linear regressions. In this study, the Box-Jenkins (1973) method is used in forecasting the price movement of certain stocks selected industries in NYSE. Before carrying out any forecasting, it is important to know the relationship between the various time series considered.

Considering 2 time series X_r and X_s, if $Cov(X_r, X_s) = E[X_r X_s] - E[X_r] E[X_s] = E[X_r] E[X_s] - E[X_r] E[X_s] = 0$, then the 2 time series have no relationship in common. These times series are said to be independent. A fundamental principle for building a proper time series model, is the principle of parsimony, which always gives priority to the model with the smallest number of parameters so as to provide an adequate representation of the underlying time series data (Adhikari, 2013). Some of the methods that are often used in the selection of models include the Akaike information criterion (AIC), Bayesian information criterion (BIC) and mean squared error (MSE). AIC helps to choose the best model that can fit NYSE stock prices. The model with the smallest AIC value is the best for it minimises the negative likelihood caused by the number of parameters

(Ababio, 2012). BIC helps to choose the model that can fit stock prices from NYSE appropriately. The model with the smallest BIC value fits the data appropriately. MSE indicates how close a set of points is to a regression line. The smaller MSE is, the closer is the fitted line to the regression line. When a time series is not stationary, the lag operator is applied to acquire a stationary series. The lag operator L is defined for a time series $\{Y_t\}$ by the following equations.

$$
\begin{aligned}
LY_t &= Y_{t-1} \\
L^2 Y_t &= LLY_t = LY_{t-1} = Y_{t-2} \\
L^k Y_t &= Y_{t-k}
\end{aligned}
$$

According to Adhikari (2013), nonlinear models are appropriate for predicting volatility (risk) changes in financial time series. Some nonlinear models are autoregressive conditional heteroskedasticity (ARCH), generalised ARCH (GARCH), threshold autoregressive (TAR) model, exponential generalised ARCH (EGARCH), nonlinear autoregressive (NAR) and the nonlinear moving average (NMA).

NYSE stock prices from the oil and gas sector was conveniently sampled out of 59 industries. The time horizon of the stock prices is 7 years spanning from January, 2012 to December, 2019. This paper uses the convenient sampling method. It is a nonprobability sampling method. According to John (1999), a least squares regression model can be used to value y for a fixed value of x.

Suppose $\{z_t\}$ satisfies the process: $z_t = \eta_1 z_{t-1} + \cdots + \eta_p z_{t-p} + \mu_t$, where z_t is a stationary series, η_p are constants and μ_t is the stochastic error term, then $\{z_t\}$ is called an autoregressive series of order p, denoted by AR(p). A simple case of an AR model is the random walk. Random walk is an autoregressive process of order 1 [AR (1)]. The model parameters are obtained using the least squares. A time series $\{z_t\}$, which satisfies: $z_t = \mu_t + \mu_1 v_{t-1} + \cdots + v_q \mu_{t-q}$, $v_1, \ldots, v_q$ constants, is said to be a moving average process of order q or MA(q) process (John 1999).

Suppose $\{z_t\}$ satisfies the process: $z_t = \eta_1 z_{t-1} + \cdots + \eta_p z_{t-p} + \mu_t + v_1 \mu_{t-1} + \cdots + v_q \mu_{t-q}$, $\{z_t\}$ is called an autoregressive moving average series of order (p, q), or an ARIMA (p,q) series. An ARIMA (p,q) series is stationary if the roots of the polynomial: $1 - \eta_1 Z - \cdots - \eta_p Z^p$ lie outside a unit circle. An ARIMA model is given as: $z_t = \eta_1 z_{t-1} + \eta_2 z_{t-2} + \cdots + \eta_p z_{t-p} + \mu_t - v_1 \mu_{t-1} - v_2 \mu_{t-2} - \cdots - v_q \mu_{t-q}$. Also called an ARIMA (p,d,q), meaning an ARIMA model of order p, d, q (Engle & Granger 1987).

In order to test the forecasting efficiency of certain models, it is important to study how these models are built. Box-Jenkins (2013) method is used in building an ARIMA model. The identification of the model involves determining the order of the model required in order to capture the salient dynamic features of NYSE stock prices. Before identification of model, it is important to verify if the model required for prediction is stationary. For a stationary time series, autocorrelation function (ACF) graph cuts off quickly or dies down quickly (Ababio, 2012). To make the time series stationary, finite differencing is carried out using the lag operator.

From the correllograms, ACF tails off slowly to zero, partial autocorrelation function (PACF) while the partial autocorrelation function truncates after lag 1. This gives an indication of an AR. The model must be checked for adequacy by considering the properties of the residuals. The residuals from an ARIMA model must have a normal distribution and should be random (Ababio, 2012).

For p-value associated with the Q statistic, if the p-value is less than the level of confidence (α) the model is not good. Ljung (1978) test is commonly used in ARIMA modeling applied to residuals of an already fitted ARIMA model. This paper uses AIC, BIC, MSE, RMS, R-squared and adjusted R-squared values for model selection. The analysis of data is performed with the help of STATA Version 12 software.

4. Data analysis and discussion of results

This section focuses on the analysis of NYSE stock prices and discussion of results obtained from analysing the stock prices. It examines basic correlation analysis among the variables under study. Time series graphs are used to obtain the required mathematical model describing stock prices in NYSE.

4.1. Testing for random walk

The correlation coefficients in Table 1 that are closer to 1 indicate very strong relationships between the stock prices from the selected companies in NYSE. Based on Table 1 statistics, there exists a very strong relationship between Encana (ECA) and Chesapeake (CHK), since their correlation coefficient is significantly high. This means an increase in the stock price of CPE leads to an increase in the price of CHK, GE or ECA. And a fall in the stock price of Callon (CPE) leads to a fall in the price of CHK, General Electric (GE) or ECA. Looking at the correlation coefficient between the various stock prices, on observes that it varies from 0.5 to 1. A correlation coefficient of less than 0.5 shows a very weak relationship between the variables in question. And for strong negative relationships, the variations are spread between - 0.5 and -1. A correlation coefficient of 1 represents a perfect relationship between the variables of concern. These are recapitulated in Table 1.

Table 1. *Correlation between the Selected Companies Operating in NYSE*

Company	CPE	CHK	GE	ECA
CPE	1.0000	-	-	-
CHK	0.1666	1.0000	-	-
GE	-0.2699	0.2306	1.0000	-
ECA	-0.0189	0.6758	0.2007	1.0000

Source: Authors, using STATA Version 12

Table 2 shows that there were 59 variables under observation. CPE has an average stock price code of 11.36 with a standard deviation of 0.58. This means that each stock price of CPE varies by 0.58 from the main price. The company recorded a minimum stock price of 10.23 and a maximum stock price of 12.76 from January 2019 to December 2019. CHK, GE and ECA have the same interpretation with different values.

Table 2. *Statistical Summary of Selected Companies Operating in NYSE*

Variable	Observations	Mean	Standard Deviation	Minimum	Maximum
CPE	59	11.35915	.5770923	10.23	12.76
CHK	59	4.450508	.214592	3.95	4.77
GE	59	12.20241	.5347018	10.84	13.2212
ECA	59	12.98492	.493717	11.9	14.21

Source: Authors, using STATA Version 12

Before using any stock prices for analysis, it is important that the stock prices assume stationarity. This helps to avoid having outliers, which produce nonlinear relationships. As such the Dickey-Fuller test for stationarity of Dickey & Fuller (1994) is applied to obtain the desired result. Depending on the t-statistic values and the coefficient of each variable, the stock price can be defined as stationary or not.

The coefficient of CPE after the first lag is negative 0.834 which is not significantly different from zero and the t-statistic too is not very significant. Thus, the process is not stationary. This requires second differencing so as to obtain more significant values indicating that the stock price has maintained stationarity. The coefficient -1.196 for CPE after lagging the second time with a t-statistic of -9.05 is significantly high. This gives stationarity to the lagged differenced variable, meaning that the time series is stationarised after the first difference. After differencing, the standard error drops from 0.150 to 0.145, hence the series follows a random walk. This suggests that the series was not stationary before the second difference. However, there is an exception with GE and CHK where the standard error increases after the first differencing. This implies that for the 2 stocks there is the need for differencing before achieving stationarity of the series. The differencing pattern observed with the 4 selected companies operating in NYSE reveals that NYSE stock prices follow a random walk. These are presented in Table 3.

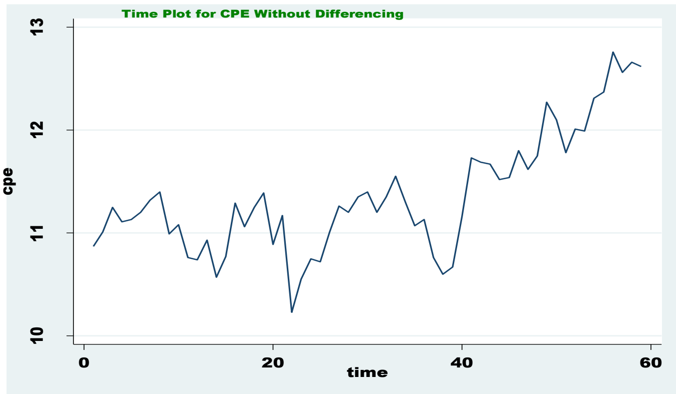
Table 3. *Dickey-Fuller Test for Stationarity*

Company	CPE	CHK	CPE	GE	ECA
Lag 1 Coefficient	-0.833	-0.001	-0.833	-0.000	-0.001
t-Statistic	-1.280	-0.220	-1.280	-0.120	-0.300
Standard Deviation	0.150	0.139	0.150	0.133	0.139
Lag 2 Coefficient	-1.196	-0.142	-1.196	0.193	0.158
t-Statistic	-9.050	-1.020	-9.050	1.340	1.150
Standard Deviation	0.145	0.141	0.145	0.139	0.138

Source: Authors, using STATA Version 12

After achieving a stationary series, it is important to provide the time series plots so as to visually appreciate the stationarity of the stock. Odd numbers panels of Figure 1 to Figure 4 are the time series plots without differencing, while the even numbers panels of Figure 1 to Figure 4 represent the time series differenced plots. Figure 1 shows the trajectory of CPE share prices from January 2012 to December 2019. From Panel 2 of Figure 1, CPE stock price was approximately constant from the first stock price to about the 40th stock price before rising. It is observed that over the years, the share prices of CPE exhibit an upwardly moving exponential trend around the mean. This implies that the stock prices are not stationary about the mean, hence the need for differencing in order to attain stationarity. Being a representation of Panel 1 of Figure 1, Panel 2 of Figure 1 is used to model the stock prices for CPE.

Panel 1



Panel 2

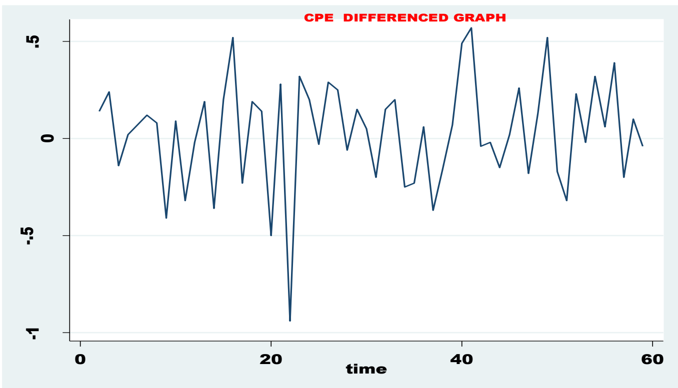
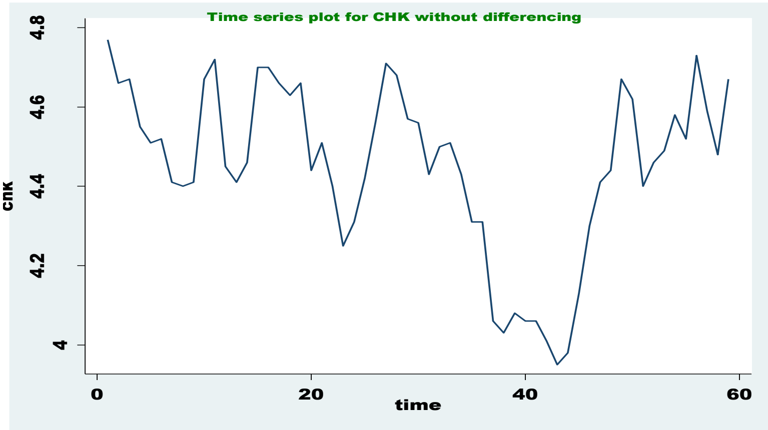


Figure 1. *Panel 1 and Panel 2: Time Series Plots for CPE*
Source: Authors, using STATA Version 12

The stock prices in Figure 2 show a seasonal pattern with a long term downward trend for CHK from January 2012 to December 2019. The seasonality of the data could be traced to the low demand for oil and gas. For example, the 2014 fall in oil prices was attributed to a lower demand for oil in Europe and China coupled with a steady supply of oil from the Organisation of the Petroleum Exporting Countries (OPEC). The excess supply of oil caused oil prices to fall sharply (Lioudis 2019). After differencing, Panel 4 of Figure 2 shows that the mean of the time series reverts about the line of origin. This means that the stock prices are now stationary.

Panel 3



Panel 4

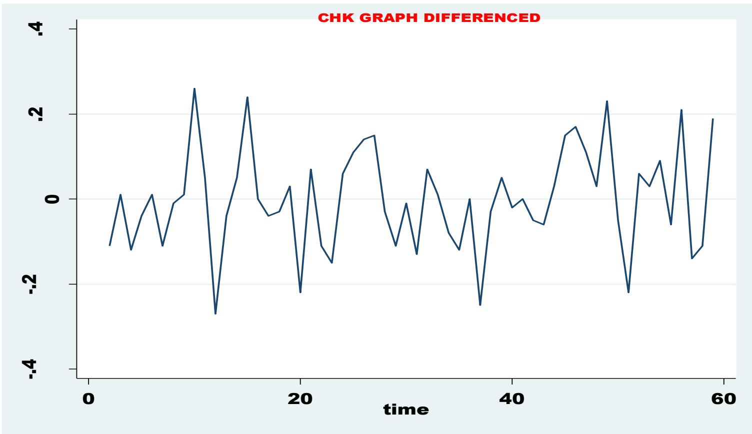
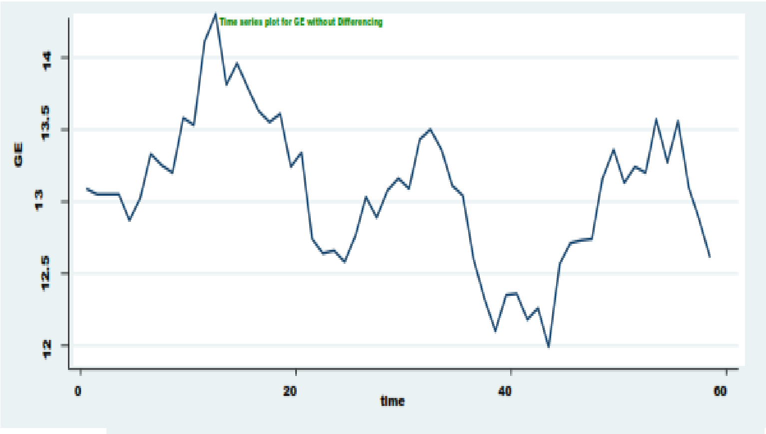


Figure 2. *Panel 3 and Panel 4: Time Series Plots for CHK*
Source: Authors, using STATA Version 12

In Figure 3, a decreasing circular trend is observed in the stock prices for GE over the period January 2012 to December 2019, a indicating a continuous fall in prices of oil and gas as time goes on due to the low demand for oil and gas. The stock prices randomly decrease from US$14.5 to US$13.5. Points that are close to each other suggest that the plotted graph meanders, meaning that it follows a random walk. Given the pattern displayed on Panel 6 of Figure 3, it can be deduced that the stock prices for GE form a random process. Therefore, they follow a random walk.

Panel 5



Panel 6

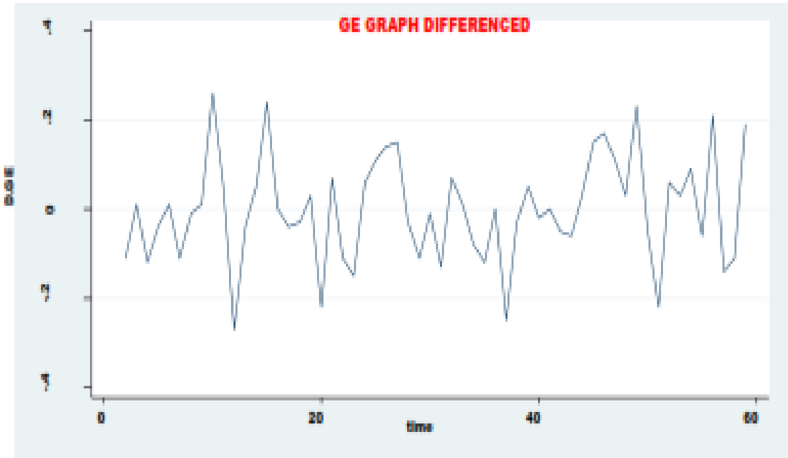


Figure 3. *Panel 5 and Panel 6: Time Series Plots for GE*
Source: Authors, using STATA Version 12

In Figure 4, a decreasing pattern is witnessed in the stock prices for ECA over the period of January 2012 to December 2019 as a result of the persistent fall in the demand for oil and gas over time. The stock prices declined from US$14.5 to US$13.2 for ECA. Points that are close to each other explain that the plotted graph trails a random walk. The trend exhibited on Panel 8 of Figure 4 confirms that hypothesis, hence the conclusion that the stock prices for ECA follow a random walk.

Panel 7

Panel 8

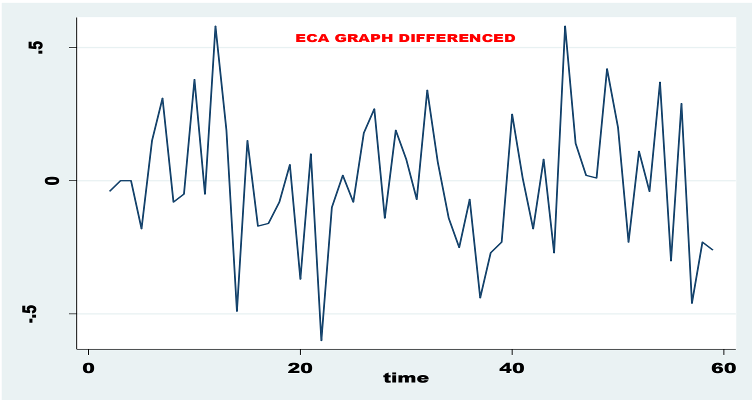


Figure 4. *Panel 7 and Panel 8: Time Series Plots for ECA*
Source: Authors, using STATA Version 12

Based on the observed pattern of the differenced plots of stock prices of CPE, CHK, GE and ECA companies operating in NYSE, one notices the formation of random processes with the time series considered for the analysis. This offers a clear evidence that NYSE stock prices follow a random walk.

4.2. Forecasting efficiency of selected models

NYSE stock prices are fitted to arbitrary models. The paper seeks to find the forecasting efficiency of 4 different models. The forecast values are fitted into ARIMA models. This is done with the help of AIC and BIC model selection criteria. The coefficients of each variable and constant terms of the 4 models are displayed in Table 4. Each coefficient is affixed L, LD2, or LD3, meaning that the variables used to generate the coefficients and constant terms are lagged variables. Based on AIC and BIC selection criteria, the model that best fits NYSE stock prices is Model (3) for it has the lowest constant term standard error (0.0194) after lagging.

Table 4. *Results for ARIMA (1, 0, 1)*

Model	(1)	(2)	(3)	(4)
Lag Model	L.CPE	L.CHK	L.GE	L.ECA
Main, _cons	11.48	-0.000372	12.36	12.97
	(0.446)	(0.00108)	(0.253)	(0.231)
ARMA L.ar	0.933	-0.383	0.895	0.864
	(0.0679)	(0.122)	(0.0680)	(0.0800)
L.ma	-0.137	-1.000	0.103	-0.0259
	(0.200)	(4910.0)	(0.195)	(0.182)
Sigma, _cons	0.269	0.146	0.227	0.242
	(0.0201)	(358.9)	(0.0194)	(0.0236)

Source: Authors, using STATA Version 12 Standard errors are in parentheses.

In Table 5, an increase is observed in the standard errors of the constant term from 0.00108, 0.446, 0.253, and 0.231 to 0.0182, 0.0191, 0.0419, and 0.0310, respectively, compared to Table 4. The increase in standard error is due to the fact that ARIMA (1, 0, 1) is closer to giving a better forecast than ARIMA (1, 1, 1). For the standard error is a measure of the efficacy of forecasting, very high standard errors indicate certain inconsistencies in the forecasting and show how the stock prices are related among themselves. Therefore, based on AIC and BIC selection criteria, the model that best fits NYSE stock prices is Model (1) because it has the lowest constant term standard error (0.0111) after lagging.

Table 5. *Results for ARIMA (1, 1, 1)*

Model	(1)	(2)	(3)	(4)
Lag Model	LD.CPE	LD.CHK	LD.GE	LD.ECA
Main, _cons	-0.00529	0.0271	-0.00517	-0.00379
	(0.0191)	(0.0182)	(0.0419)	(0.0310)
ARMA L.ar	-0.296	0.657	0.610	-0.432
	(1.390)	(0.339)	(0.809)	(1.112)
L.ma	0.406	-0.869	-0.484	0.331
	(1.271)	(0.221)	(0.896)	(1.193)
Sigma, _cons	0.116	0.265	0.231	0.250
	(0.0111)	(0.0193)	(0.0166)	(0.0275)

Source: Authors, using STATA Version 12 Standard errors are in parentheses.

In Table 6, some of the standard errors of the constant term become very negligible as a result of second differencing of the variables of the models. This minimises the error terms, because as the number of lags increased, there is a significant drop in the error of the estimate. Hence the better the estimates of the model are when the number of lags increases. Consequently, based on AIC and BIC selection criteria, the model that best fits NYSE stock prices is Model (3) because it has the lowest constant term standard error (0.0185) after lagging.

Table 6. *Results for ARIMA (1, 2, 1)*

Model	(1)	(2)	(3)	(4)
Lag Model	LD2.CPE	LD2.CHK	LD2.GE	LD2.ECA
Main, _cons	0.000662	0.00165	0.00279	-0.000364
	(0.00114)	(0.00210)	(0.00188)	(0.00183)
ARMA L.ar	0.103	-0.190	0.0756	-0.0768
	(0.185)	(0.156)	(0.136)	(0.194)
L.ma	-1.000	-1.000	-1.000	-1.000
	(1472.5)	(0.035)	(0.030)	(211.0)
Sigma, _cons	0.117	0.270	0.231	0.253
	(86.34)	(0.0206)	(0.0185)	(26.69)

Source: Authors, using STATA Version 12 Standard errors are in parentheses.

ARIMA (1, 3, 1) in Table7 presents very negligible standard errors as the number of lags increases. The higher the number of lags, the better the model. It can equally be observed that the ratio of the standard deviation to the standard errors is higher than in any other ARIMA model used before. Accordingly, based on AIC and BIC selection criteria, the model that best fits NYSE stock prices is Model (3) because it has the lowest constant term standard error (0.0228) after the lagging.

Table 7. *Results for ARIMA (1, 3, 1)*

Model	(1)	(2)	(3)	**(4)**
Lag Model	LD3.CPE	LD3.CHK	LD3.GE	LD3.ECA
Main, _cons	-0.000372	0.000215	0.000881	-0.000675
	(0.00108)	(0.00202)	(0.00139)	(0.00159)
ARMA L.ar	-0.383	-0.564	-0.536	-0.601
	(0.122)	(0.0860)	(0.108)	(0.121)
L.ma	-1.000	-1.000	-1.000	-1.000
	(4910.0)	(0.046)	(0.034)	(0.039)
Sigma, _cons	0.146	0.353	0.263	0.298
	(358.9)	(0.0449)	(0.0228)	(0.0280)

Source: Authors, using STATA Version 12 Standard errors are in parentheses.

4.3. Fitting New York stock exchange stock price in appropriate model

With 4 different fitted models to NYSE stock prices, a question remains unanswered. Which of the models best fits the NYSE stock prices? This leads to the use of AIC and BIC values selection criterion. These values are obtained and presented in Table 8. The smallest AIC and BIC values are -42.6 and -34.6, respectively, from ARIMA (1, 3, 1). Hence, ARIMA (1, 3, 1) best fits NYSE stock prices. Based on the efficiency, ARIMA (1, 3, 1) is the most efficient for all the models tested in this paper, and it applies to NYSE stock prices. This is because as the number of lag increases the more stationary the data become and the closer to the true description of the real situation they are.

Table 8. *Values of AIC and BIC for the Selected Models*

Criterion/Company	ARIMA Model	AIC	BIC
CPE	ARIMA (1,1,1)	18.5492	26.7214
CHK	ARIMA (1,3,1)	-42.59805	-34.56872
GE	ARIMA (1,0,1)	2.569871	10.81164
ECA	ARIMA (1,2,1)	9.221534	17.46331

Source: Authors, using STATA Version 12 Standard errors are in parentheses.

Other measures that can be used in selecting the model that best fits a phenomenon are MSE and RMSE values of the stock prices. The smallest RMS indicates the best fitted model. Table 9 displays some results on these measures. In Table 9, it is observed that CHK has the smallest MSE and RMSE values. This ties with Table 8, since CHK had the lowest AIC and BIC values. The theoretical form of the model is given by Equation (1).

$$z_t = \alpha_1 z_{t-1} + \alpha_2 z_{t-2} + \alpha_3 z_{t-3} + \alpha_4 z_{t-4} - \varepsilon_t \tag{1}$$

Table 9 displays the parameter estimates of the selected model**.** After the parameters are estimated the model for the share price of CPE, CHK, GE and ECA over the period under consideration is given in Equation (2).

$$z_t = -0.564\, z_{t-1} - 0.383 z_{t-2} - 0.536\, z_{t-3} - 0.601 z_{t-4} - \varepsilon_t \tag{2}$$

Applying the Lag operator (L) gives a new equation. This is represented in Equation (3).

$$z_t = -0.564 L z_t - 0.383 L^2 z_t - 0.536 L^3 z_t - 0.601 L^4 z_t - \varepsilon_t. \tag{3}$$

The required characteristic is given in Equation (3). The values of ν can be obtained by solving the polynomial of degree 4 expressed in Equation (4). For absolute values of ν greater than 1, the model is stationary and otherwise for absolute values of ν less than 1. Adjusted R-squared is the measure of accuracy of prediction or regression. Therefore, 52%, 7%, 37% and 11% of the variation in the graphs of CPE, CHK, GE and ECA respectively, is accounted for by the regressed variables of the models.

$$0.601\nu^4 + 0.536\nu^3 + 0.383\nu^2 + 0.564\nu + 1 = 0 \tag{4}$$

Table 9. *Summarised Statistics for Selected Companies in NYSE*

Statistical Measure	MS	RMSE	R-Squared	Adj R-Squared
CPE	10.1883684	0.40017	0.5275	0.5192
CHK	0.221054817	0.20731	0.0828	0.0667
GE	6.2436263	0.42589	0.3765	0.3656
ECA	1.77050818	0.4658	0.1252	0.1099

Source: Authors, using STATA Version 12

The predicted average stock values for CPE, CHK, GE and ECA are given in Table 10. Between January and December 2020, the average stock price for

CPE was US$11.25. For CHK, it stood at US$4.66, while GE and ECA recorded each an average stock price of US$12.49. A careful observation of the predicted average stock prices for the companies reveals that between 2021 and 2024, CPE may experience an average stock price increase of 7.91%, CHK may witness an average stock price increase of 57.51%, while the average stock price increase for GE and ECA may reach 73.65% and 91.35%, respectively due to the prevailing erratic conditions surrounding the global business activities in the oil and gas sector.

Table 10. *Predicted Average Stock Values for CPE, CHK, GE and ECA for the period 2021-2024*

Company -	Average Stock Price (US$) in 2020	Predicted Average Stock Price (US$) for 2021-2024	Percentage (%) Increase (+) or Decrease (-)
CPE	11.25	12.14	+ 7.91
CHK	4.66	7.34	+ 57.51
GE	12.49	21.69	+ 73.65
ECA	12.49	23.90	+ 91.35

Source: Authors, using STATA Version 12

5. Conclusion and recommendations

The paper has established how to select a model that best fits NYSE stock prices. With reference to the first objective of this paper, it is empirically evident that the stock prices in NYSE follow a random walk based on the comparison of the standard errors before and after differencing. In order to forecast the stock prices in any stock market, it important to test the forecasting efficiency of different models to ensure their reliability in predicting future stock prices of any securities. Once the reliability yardstick is achieved, it becomes necessary to understand whether or not the pattern of the stock prices in the stock market follow a random walk. When the stock prices in the stock market trail a random walk, the implication is that it is impossible for traders to outperform the overall stock market average except by pure luck. Based on the findings of this study, investors are advised to apply the principles of the random walk theory. They should buy securities now and hold them to sell in the future in a selection of securities that represent the overall stock market since the stock prices in NYSE assume an upward moving trend.

Appendices

The properties of autocorrelation function (ACF) and partial autocorrelation function (PACF) for NYSE stock prices are presented in Appendix 1 to guide the choice of ARIMA model fitting the stock price of NYSE. It shows that ACF dies down slowly and cuts of at the 9th lag. The stock price is slightly stationary. The Bartlett's formula is used to obtain the moving average component of the data. This is done at a 95% confidence interval. This shows ACF from 1st January 2012 to 30th December 2019. ACF dies down at a very slow rate confirming the existence of a trend in the stock price. But it cuts off at the initial line. This means that, initially the time series was not stationary.

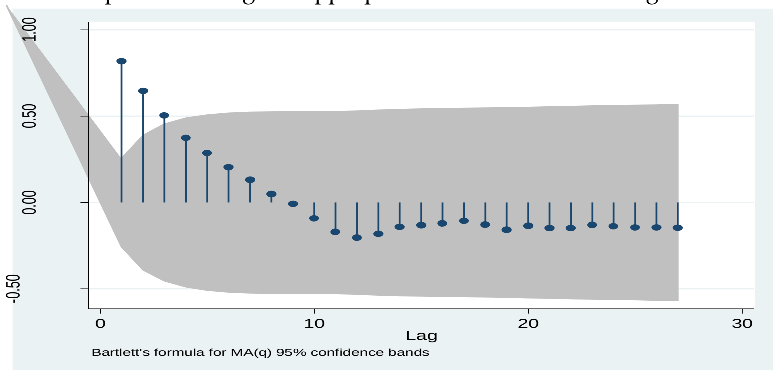
Appendix 1. *Autocorrelation Function and Partial Autocorrelation Function Properties*

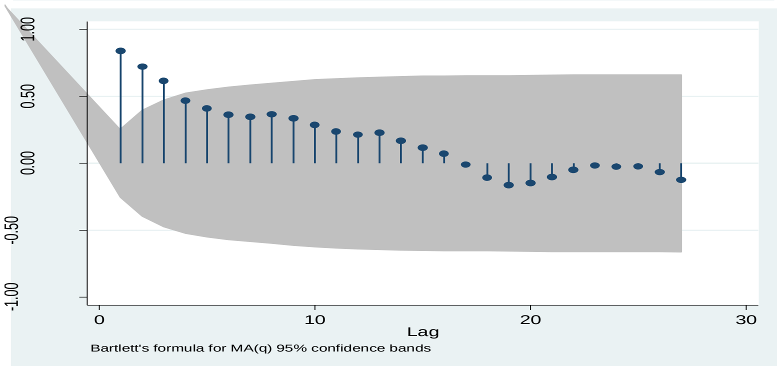
Graph Type	AR(p)	MA(q)	ARMA(p,q)
ACF	Tails off	Cuts of at lag q	Tails off
PACF	Cuts off at lag p	Tails off	Tails off

Source: Authors, using STATA Version 2012

Appendix 2. *ACF for CHK and CPE*

Appendix 2 shows ACF of ECA and GE share prices for the period under consideration. ACF dies down or tails off after the first lag at a very slow rate confirming the existence of a trend in the stock price. This means the stock price for ECA and GE are AR processes. This helps in selecting the appropriate model for forecasting.

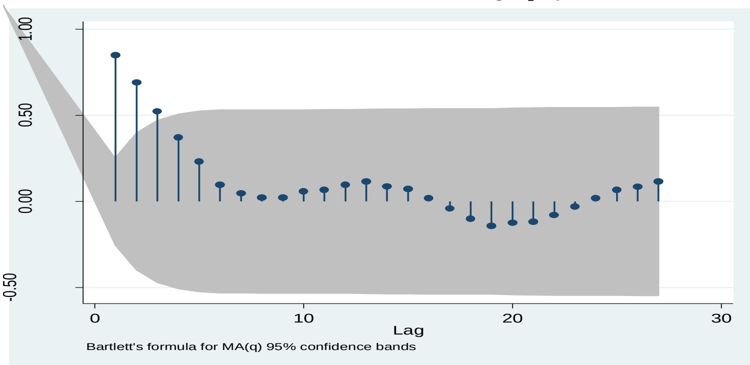


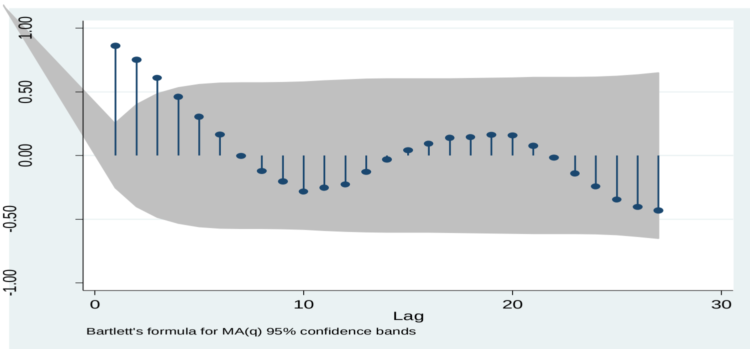


Source: Authors, using STATA Version 12

Appendix 3. *ACF Plots for ECA and GE*

ACF for ECA truncates after lag 1 indicating an AR (1) model. ACF is used to obtain trend in the stock prices. The graph cuts off after the first lag, and then begins to die off gradually. This indicates that at particular time frame, the stock price falls. This is due to political instability and too much supply of the stock in the market.

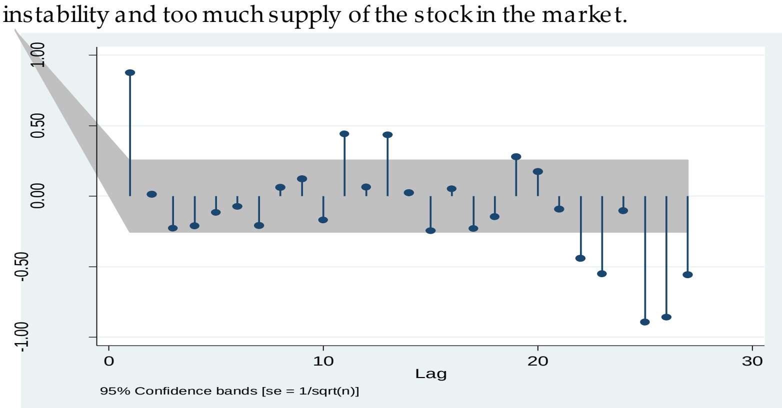


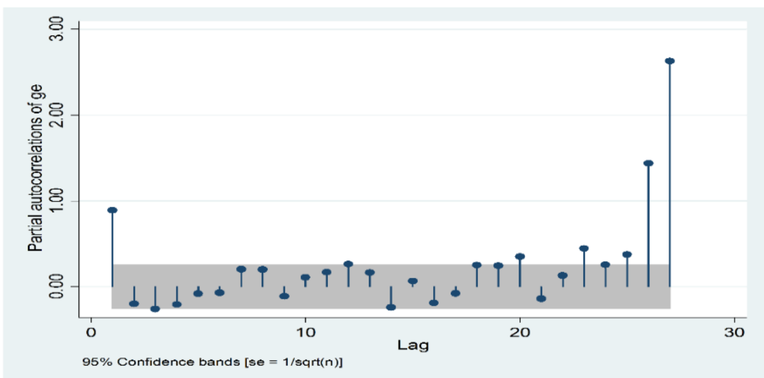


Source: Authors, using STATA Version 12

Appendix 4. *PACF Plots for ECA and GE*

PACF for ECA truncates after lag 1 indicating an AR (1) model. PACF is used to obtain the order of AR model. The graph cuts off after the first lag, and then begins to die off gradually. This indicates that at particular time frame, the stock price falls. This is due to political instability and too much supply of the stock in the market.





Source: Authors, using STATA Version 12

References

Ababio, K.A. (2012). *Comperative Study of Stock Prices Forecasting using ARIMA Models*. Kumasi, Ghana: Kwame Nkrumah University of Science and Technology.

Adebayo, A.F (2014). Forecasting stock market series with ARIMA model, *Journal of Statistical and Econometric Methods*, 3(3), 65-77.

Adhikari, R. (2013). *An Introductory Study on Time series Modeling and Forecasting*, Germany: LAP Lambert Academic Publishing.

Abdel-Aal, R.E. & Al-Garni, A.Z. (1997). Forecasting monthly electric energy consumption in Eastern Saudi Arabia: Univariate time-series analysis, *Energy*, 22(11), 1059-1069. doi. 10.1016/S0360-5442(97)00032-7

Arsham, H. (2015). *Time Critical Decision Making for Business Administration.* 22 August 2020. [Retrieved from].

Box G.E.P. & Jenkins G.M. (1972). *Time Series Analysis: Forecasting and Control*. Revised Edition, San Francisco, USA: Holden-Day.

Box G.E.P. & Jenkins G.M. (1973). *Time Series Analysis: Forecasting and Control*. Revised Edition, San Francisco, USA: Holden-Day.

Box G.E.P. & Jenkins G.M. (1976). *Time Series Analysis: Forecasting and Control*. Revised Edition, San Francisco, USA: Holden-Day.

Box G.E.P. & Jenkins G.M. (1997). *Time Series Analysis*: *Forecasting and Control*. Revised Edition, San Francisco, USA: Holden-Day.

Box G.E.P. & Jenkins G.M. (2013). *Time Series Analysis*: *Forecasting and Control*. Revised Edition, San Francisco, USA: Holden-Day.

Chen, J. (2018). *Tokyo Stock Exchange*: *What is the Tokyo Stock Exchange*? 15 September 2021. [Retrieved from].

Dickey, D.A. & Fuller, W. A. (1994). Testing for a Unit Root in Time Series with Pretest Data-Based Model Selection, *Journal of Business & Economic Statistics*, 12(4), 461-470. 10.1080/07350015.1994.10524568

Engle, R.F. & Granger, C.W.J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55, 251-176. doi. 10.2307/1913236

John, C. (1971). *How to Choose the Right Forecasting Technique*, 4e, New York, USA: Wiley.

John, M. (1999). Dow theory basic tenets*:* Technical analysis of the financial market. 18 August 2020 [Retrieved from].

Lioudis, N. (2019). What causes oil prices to fluctuate? Global trade guide. 20 September 2021. [Retrieved from].

Ljung, M.G. (1978). On a measure of lack of fit in time series models, *Biometrica*, 65(2), 297-303. doi. 10.1093/biomet/65.2.297

Manoj, K. & Madhu, A. (2014). An application of time series ARIMA forecasting model for predicting sugarcane production in India, *Studies in Business and Economics*, 9(1), 81-94.

Okany, T.C. (2014). Effect of oil price movement on stock prices in the Nigerian equity Market, *Research Journal of Finance and Accounting*, 5(15), 61-78.

Park, E. (2016). Sustain. Energy: Case study, *Renewable Electricity Generation Systems*, 8, 393-401.

Peijun, C. (2016). Predictive analysis of Chinese total health expenditure based on ARIMA model, *Medicine and Society*, 3(3), 334-344.

Rahman, A. (2015). A single-group pre-test post-test design that adjusts for the passage of time: Using forecast modeling to evaluate treatment effects in single-group interrupted time serie analysis. 21 August 2020. [Retrieved from].

Shum, M. (2016). The 8 smallest stock markets in the World: A list of frontier markets in the World. 21 August 2020. [Retrieved from].

Sindhu, S. (2017). Solar energy deployment for sustainable future of India: Hybrid SWOC-AHP analysis, *Renewable and Sustainable Energy Reviews*, 72, 1138-1151. doi. 10.1016/j.rser.2016.10.033

Saylors, E. (2017). *The Fallacy of Forecasting in a Complex World*. Towards Data Science. 20 November 2021. [Retrieved from].

Schmidt, B.A. (2011). *Financial Markets and Trading: An Introduction to Market Microstructure and Trading Strategies* USA: John Wiley and Sons Inc.

Tang, Z., De Almeida, C. & Fishwick, P. A. (1991). Time series forecasting using neutral networks vs. Box-Jenkins methodology, *Simulation*, 57(5), 303-310. doi. 10.1177%2F003754979105700508

Yue, Z., Shengnan, W. & Yuan, L. (2015). Application of ARIMA model on predicting monthly hospital admissions and hospitalization expenses for respiratory diseases, *Chinese Journal of Health Statistics*, 2(2), 197-200.