

# Analysis of Stock Return in Nairobi Securities Exchange Using Autoregressive Integrated Moving Average Model

Simon Oluoch Ondiwa\* Dr. David Oima Dr. Fredrick Aila  
School of Business and Economics, Maseno University Private Bag Maseno, Kenya

## Abstract

Nairobi Securities Exchange exhibited erratic performance ten years after the 2007 post-election violence. Investors, listed firms and the government had hard time in making investment decisions and policies given the volatility of Nairobi Securities Exchange during this period. This trend worsen in 2011 and 2015 when the burse experienced bear run leading to a loss of investors wealth in terms of market capitalization. The Nairobi Securities Exchange 20 Share Index recorded high of 5,346 points during the first quarter of the year but dipped to 4,040 points in December 2015, representing 2.0 percent compared to December 2014. With observed volatility witnessed during the study period, analysis of stock return in Nairobi Securities Exchange using autoregressive integrated moving Average model is essential. The model output using ARIMA (p, d, q) will be used useful to managers, investors, investment analysts and policy makers. Using Autocorrelation functions and Partial Autocorrelation functions, the appropriate ARIMA (p, d q) model identified to be suitable for forecasting stock return is ARIMA (1, 1, 1) this is because the autocorrelation function drops sharply after p lags meaning the AR term exists and was be included.

**Keywords:**ARIMA, Investors, Stock Return, Partial Autocorrelation, Bear run, NSE-20 Share Index, Stock Return, Volatility

**DOI:** 10.7176/JAAS/71-04

**Publication date:** April 30<sup>th</sup> 2021

## 1. Introduction

Box & Jenkins (1970) popularized the ARIMA (p, d, q) model. The model holds that the past values of time series and their error terms contain information that can be used for forecasting. The p, d, q refers to Autoregressive (AR), integrated (I) and moving average (MA) parts of data set respectively. The models are applied in instances where data show evidence of none stationarity, the data is then transformed by differencing their natural logarithms. The model uses historical data and decomposes them to (AR) process, where there is memory of past events. An integrated process which accounts for stationarity , making it easier to forecast; and MA of the forecast error such that the larger the historical data, the more accurate the forecasts will be.

The ARIMA (p, d, q) model is arrived at by transforming the time series data to stationary. Cooper (1972) held that ARIMA (p, d, q) model could outperform the large and complex econometric models in many situations. If time series “d” is differentiated, then the model becomes stationary ( Ransang & Titida, 2006). When ARIMA (p,q) is applied, then the original data is transformed to stationarity and it is referred to us autoregressive moving average ( p, d, q). Where p is the number of autoregressive terms, d is the number of none- seasonal differences and q is the number of lagged forecasts errors in the prediction equation. Stationarity in the model is identified by the use of autocorrelation function graph (AFC), if the graph of the series values either cuts off or dies down fairly quickly, then the time series should be considered stationary. If the graph of ACF dies down extremely slowly, then the time series should be considered none stationary. If the original time series is stationary, d =0 and ARIMA model reduces to ARMA model.

In India, Reddy (2019) predicted the stock market Index using stochastic time series ARIMA (p, d , q). Reddy (2019) observed that stock market is volatile and that prediction of its movement will be more useful to the stock traders to design their trading strategies. He used the best fit ARIMA (0,1,0) model for forecasting the values of time series. Using the best fitted model, the predictions were made for the period ranging from 6<sup>th</sup> January, 2018 to 3<sup>rd</sup> June, 2018 using the weekly data ranging from 6<sup>th</sup> January, 2014 to 31<sup>st</sup> December 2017. The results obtained from the study confirmed the prospective of ARIMA model to forecast the future time series in short run and would assist the investing community in making the profitable investment decisions. The study however did not use quarterly data for a ten year period but instead used weekly data for a three year period. The study did not arrive at the best fit ARIMA (1, 1, 1) but instead he used the best fit ARIMA (0, 1, 0) for forecasting the value of time series.

Githingi (2014) studied modeling inflation in Kenya using ARIMA and VAR models. She asserted that inflation is important indicator of economic activity and it is used by decision makers to plan economic policies. She modeled inflation over the period 2005 to 2013 using two autoregressive models; Autoregressive Integrated Moving Averages (ARIMA) model and the Vector Autoregressive model. She used ARIMA model to fit historical CPI time series expressed in terms of the past values of itself plus current and lagged values of error term resulting to the model ARIMA (0,1,1). Data for the last six months was used to evaluate the performance of

the prediction. In conclusion, ARIMA model (1,1,0) resulted as the best model. Her study revealed that despite the exclusion of explanatory variables, there was evidence of substantial inflation inertia; this is also evidenced by the hypothesis that if the level of inflation is determined by inertia, then the parameters on lagged inflation should sum to unity. Finally, parameters of the selected lagged parsimonious model in her study sum approximately in unity. The study did not use only ARIMA (p, d, q) models but instead included VAR models. The study did not adopt the ten year study period as seen in the present study but instead used eight years study period. They did not fit the model with ARIMA (1, 1, 1) but instead fitted the model as ARIMA (0, 1, 1). He did not use quarterly data but instead used monthly data to fit the model. The nature of models is not known assuming the reviewed study could adopt the same methodology used in the present study.

In United States of America, Jarrett (1989) studied forecasting monthly earnings per share; the study was a time series. First, he computed the Autocorrelation Function (ACF) and Partial Autocorrelation Functions; the functions indicated a slow decline to about lag 6 confirming that the model is AR. To determine the appropriate number of AR parameters, to estimate Partial Autocorrelation Function (PACF) is drawn and obtained. The Partial Autocorrelation Function contained a large value at lag 1 (approximately, 0.8) which is above 95% significant limit. The number of AR parameters to be estimated was 1, although AR (2) could also still be appropriate. In conclusion, the analysis of data indicated that ARIMA modeling and time series decomposition produced useful results. In examining the final estimates of the AR (2) model, the AR (2) parameter yielded a t value of 0.86 corresponding to a p- value of 0.391. A second round of ARIMA modeling without the AR (2) parameter estimate would yield better results with only the constant and AR(1) parameter included in the forecasting model. Jarrett (1989) however did not reveal the study period and she did the study in the United States hence predicting earnings per share in Kenya may reveal different results.

Murrithi *et al.*, (2014) did a study titled analysis of the fluctuation of Gross Domestic Product in Kenya using Autoregressive Integrated Moving Average ARIMA (p,d,q). Using the famous Box-Jenkins (1976) methodology, to determine the most efficient and adequate model for analyzing the Kenyan GDP. The methodology involves stages of identification, estimation, diagnostic, checking forecasting of a univariate time series. Their study used an exploratory research design and a sample of 51 observations was used. The annual data was obtained from the World Bank National Accounts Data. They noted in their study that the Kenyan GDP has been fluctuating in the past years. They used statistical methods to analyze, explain and present the time series data of Kenya's GDP growth from 1960-2011. Comparing with other models, they asserted that ARIMA (2,1,2) was selected as the final model for forecasting the Kenya's GDP growth for 2012-2030. All the model assumptions were satisfied (randomness, independent and normality of the residuals). In conclusion, the study found out that Kenya's GDP will continue to rise if other factors remain constants. The study however did not model Gross Domestic Product using quarterly data, they did not use ten years to model Gross Domestic Product but instead used 51 years to model Gross Domestic product and they did not fit ARIMA (1,1, 1) but instead fitted ARIMA (2, 1, 2).

In India, Shankar & Abhishek (2019) did a study on predictive efficiency of ARIMA (p, d, q) and ANN models; analysis of Nifty Fifty in Indian Stock market. They applied simultaneous use of AAE, RMSE, MAPE and MSPE statistical tools. They predicted the accuracy of the two models. The results indicated that ARIMA (2,1,2) and ANN (4-0-1) with both train functions GDX and BFG were the best predictors with ANN dominating over ARIMA (p, d, q) model in the final outcome. In conclusion, neural network had better predictive capability in case of long span of time and non linear volatile series like Indian Stock Market Index. Even though neural networks had advantage over ARIMA (p, d, q) models, they are still capable to provide a right and validated direction for the investors to make better decisions. This study did not use annual or quarterly data; instead they used daily data for a period of ten years. The study did also not focus solely on ARIMA (p, d, q) models but instead focused on comparing ARIMA (p, d, q) with neural networks. Furthermore, the study was not done in Kenya; instead the study was done in India. The methodological differences and the study area could lead to mixed results.

Reddy (2019), Githingi (2014), Jarrett (1989) and Murrithi *et al.*, (2014), Shankar & Abhishek (2019) are in agreement that modeling ARIMA (p, d, q) time series model for purposes of forecasting is essential for decision making. For instance Reddy (2019) asserted that stock market is volatile and that prediction of its movement is more useful to the stock traders to design their trading strategies. However these studies used different study periods and data points in their time series analysis. Some studies used daily data, some used monthly data; some studies used weekly data while others used annual data. None of the studies reviewed used quarterly data for ten years time series period. Furthermore, the studies were done mostly in developed countries and they used different variables for purposes of modeling; this may halt fitting models and subsequent forecasting given lack of uniformity in regulations, general economic environment and efficiency in securities exchange systems. Lastly, the studies used different research philosophies and designs.

## 2. Methods

A model developed to establish forecasting models for stock return in Nairobi Securities Exchange. Autoregressive process AR expresses a dependent variable as a function of past values of the dependent variable. A  $p$ th-order process is of the form of:

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad \text{.....1}$$

Where,  $y_t$  is the stationary depended variable being forecasted at time  $t$ .  $y_{t-1}$  and  $y_{t-2}, \dots, y_{t-p}$  is the response variable at time lags  $t-1, t-2, \dots, t-p$  respectively.  $\alpha = \mu(1 - \phi_1 - \phi_2 - \dots - \phi_p)$ .  $\phi_1, \phi_2, \dots, \phi_p$  are the coefficient to be estimated,  $\varepsilon_t$  is the error term at time  $t$  with mean zero and a constant variance. Using the backshift operator we can write the AR(P) model as:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) y_t = \phi(B) y_t = \varepsilon_t \quad \text{.....2}$$

The moving average model of order MA(q) is defined as:

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad \text{.....3}$$

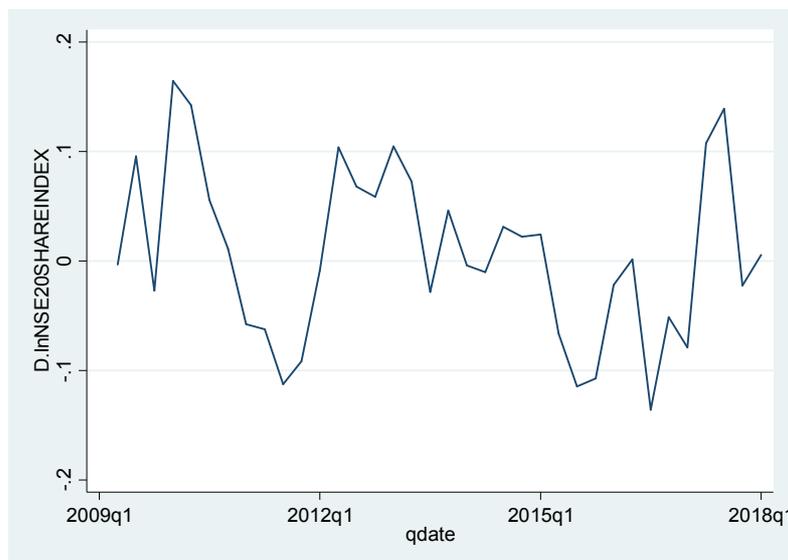
Where;  $q$  is the number of lags in the moving average and  $\theta_1, \theta_2, \dots, \theta_q$  are parameters to be estimated. The moving average operator is given by:

$$\phi(B) = \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q \quad \text{.....4}$$

### 2.1 Stationarity Test

Data was transformed to stationary by differencing their natural logarithms (Granger & Newbold, 1974, Gujarati, 2006). Most researchers employ the first difference approach, mainly as a result of Nelson and Plossers' (1982) work which they argued that many macroeconomic time series use difference stationarity and not trend stationary. Granger and Newbold (1974) refer to results obtained from regressions that contain non-stationary data as spurious regression results which are empirical results with high- R squared estimates and very low Durbin-Watson statistics, which can be regarded as excellent results but are of no use (Gujarati, 2006).

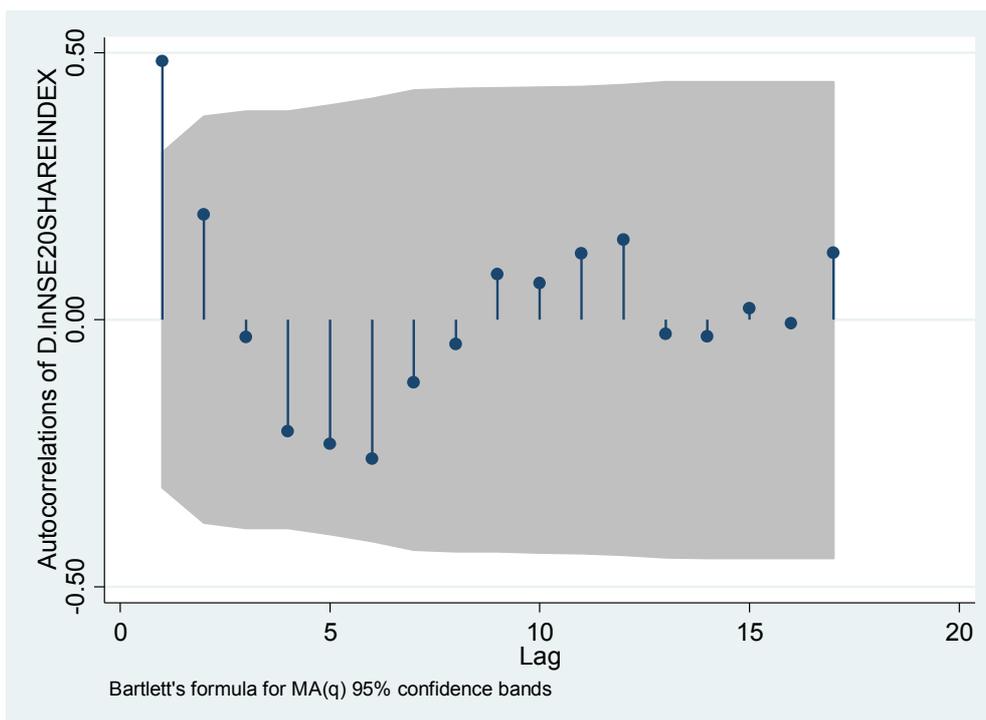
The difference of the natural logarithm of NSE-20 share index is stationary as shown in figure 1, the difference of the natural logarithm is stationary around 0 and with increased variance over time and a constant mean.



**Figure 1:** Stationarity of NSE 20 Share Index  
 Source: Field Data, 2019

### 2.2 Autocorrelation

Many parametric statistical procedures assume that the errors of the models used in the analysis are independent of one another, that is, the errors are not correlated. When this assumption is not met in the context of time-series research designs, the errors are said to be auto correlated or dependent. Because time-series designs involve the collection of data from a single participant at many points in time rather than from many participants at one point in time, the assumption of independent errors inherent in many parametric statistical analyses may not be met. When this occurs, the outcome of these analyses and the conclusions drawn from them are likely to be misleading unless some corrective action is taken (Bradley & Sean, 2006). Figure 2 below shows that there are no issues with autocorrelation of NSE-20 Share Index.



**Figure 2 Autocorrelation of NSE-20 Share Index**

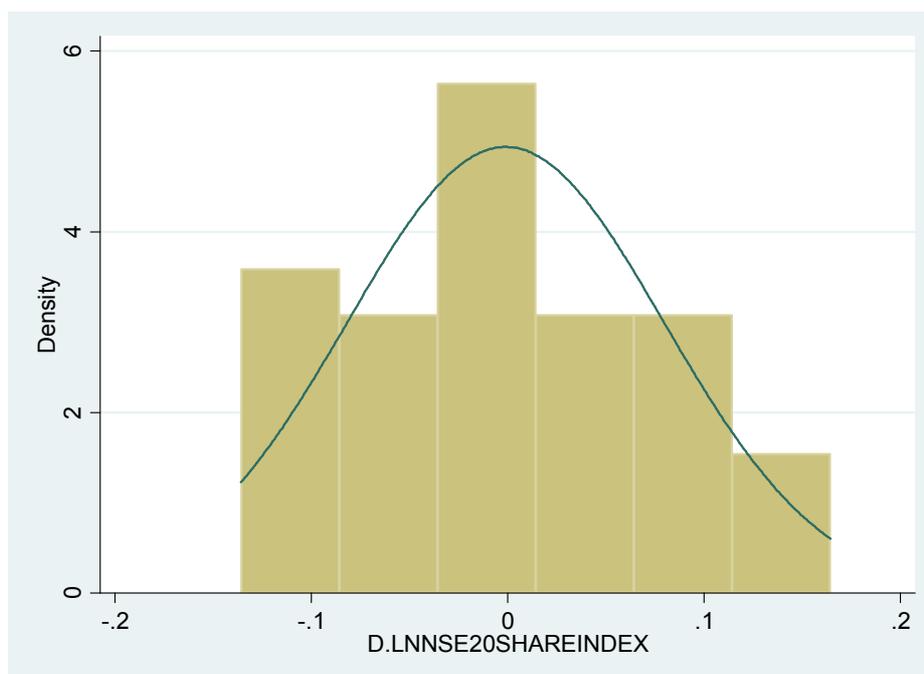
*Source: Field Data, 2019*

### 3. Results

#### Descriptive Statistics

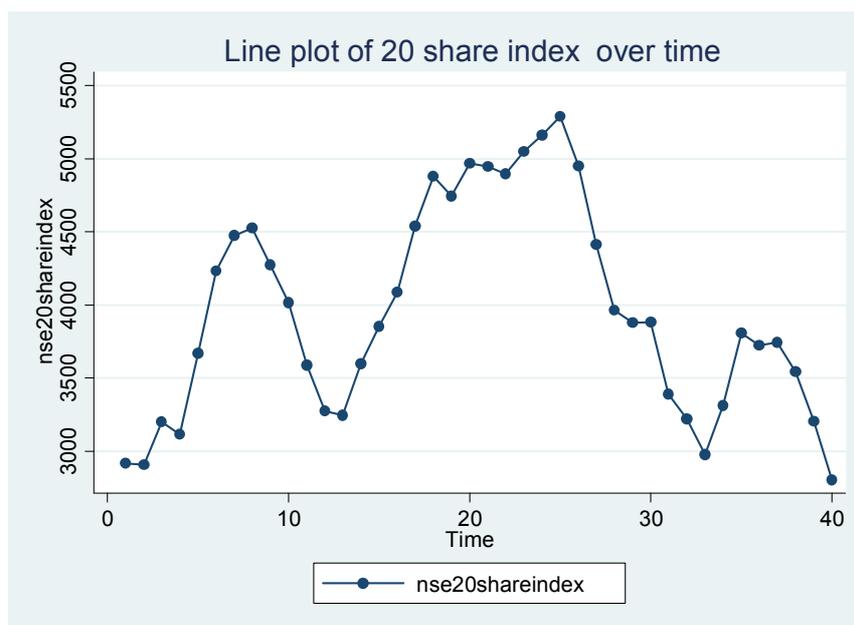
Variable	Obs	Mean	Std. Dev	Min	Max	Skewness	Kurtosis
NSE-20 Share Index	40	8.266954	0.1828903	7.939608	8.573505	-0.0040621	1.861072

NSE-20 Share Index had a mean of 8.266954 and a standard deviation of 0.1828903 with minimum values of 7.939608 and maximum value of 8.573505.



**Figure 3: Histogram for NSE-20 Share Index**

*Source: Field Data, 2019*

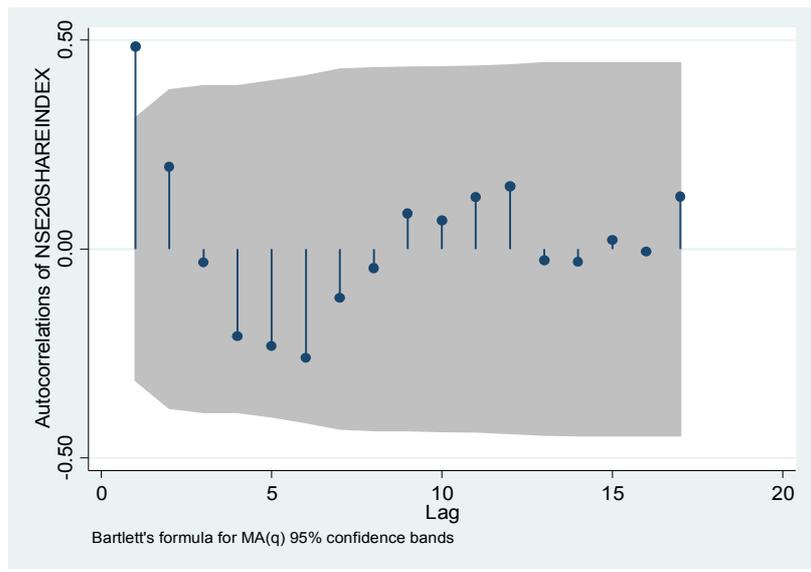


**Figure 4: Line Plot of 20-Share Index over Time**  
 Source: Field Data 2019

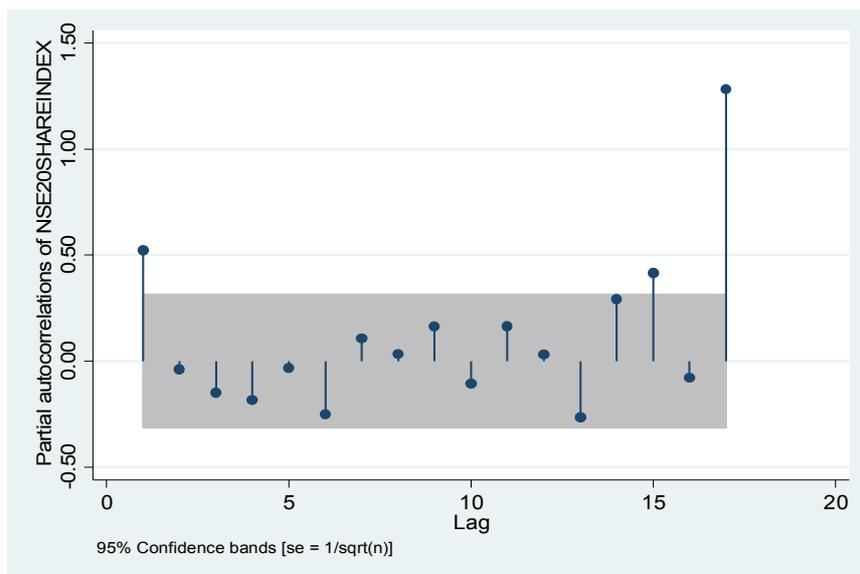
Table 1: Autocorrelation and partial Autocorrelation of NSE-20 Share Index

LAG	AC	PAC	Q	Prob>Q	-1	0	1	-1	0	1
					[Autocorrelation]			[Partial Autocor]		
1	0.4847	0.5223	9.8874	0.0017						
2	0.1967	-0.0378	11.559	0.0031						
3	-0.0317	-0.1490	11.604	0.0089						
4	-0.2094	-0.1848	13.606	0.0087						
5	-0.2328	-0.0321	16.155	0.0064						
6	-0.2607	-0.2518	19.448	0.0035						
7	-0.1174	0.1065	20.138	0.0053						
8	-0.0458	0.0323	20.246	0.0094						
9	0.0855	0.1639	20.635	0.0144						
10	0.0689	-0.1053	20.897	0.0218						
11	0.1250	0.1659	21.789	0.0261						
12	0.1507	0.0300	23.133	0.0266						
13	-0.0268	-0.2640	23.178	0.0396						
14	-0.0304	0.2937	23.237	0.0565						
15	0.0219	0.4146	23.269	0.0786						
16	-0.0068	-0.0781	23.272	0.1067						
17	0.1255	1.2814	24.417	0.1085						

Source: Field Data 2019



**Figure 5: Autocorrelation of Nairobi Securities Exchange 20 Share Index**  
 Source: Field Data 2019



**Figure 6: Partial Autocorrelation of Nairobi Securities Exchange 20 Share Index**  
 Source: Field Data 2019

Results in table 1 reveal that the ARIMA (p, d, q) model for Stock Return is ARIMA (1,1,1) where with AR (1), I(1) and MA (1). **Table 1** used time series data for a single variable to ascertain autocorrelation. ARIMA (1, p, d) model includes one predictor, the observation immediately preceding the current value and an ARIMA (2, d, q) model includes two predictors, the first and the second preceding observations. Given that the data became stationary after the first differencing, I(d) was also 1; this is consistent with observations by jebb *et al.*, (2015) who asserted that the order of differencing rarely needs to be greater than two in order for the data to be stationary and that in practice, the choice comes down to whether the series is transformed into either its first or second differences, the optimal choice should be the order of differencing that results in the lowest series variance and does not result in an increase in variance that characterizes over differencing. As seen in figure 6, the PACF is none zero and it is gradually decaying.

#### 4. Conclusion

ARIMA (p, d, q) model was used to identify the appropriate model that could be used for forecasting stock return in Nairobi Securities exchange, Upon assessing the series and transformation of the data to attain stationarity, ARIMA (1, 1, 1) was identified as the appropriate model that could be used for forecasting. AR (p) with p= 1, I(d) with d=1 and MA (p) with p=1 was adopted. According to reviewed studies, to asses if there are

AR (p) terms, we observe the autocorrelation function, if it drops sharply after p lags, the AR term exists and should be included. The I(d) term exists when the series is differentiated to attain stationarity; this is referred to as data transformation. Traditionally, the I(d) will always not exceed two lags or t-2. The MA (p) term was identified by use of partial autocorrelation which related the series at time t with previous random shocks which always interact with complexity (white noise). Just like the AR (p) term, when the partial autocorrelation function gradually decays as it approach zero, then MA (p) terms exists. Alternatively, when there is a sharp drop of partial autocorrelation function after p lags then MA term exists.

## References

- Cooper, RL 1972, *The Predictive Performance of Quarterly Econometric Models of the United States, Econometric Models of Cyclical Behaviour*. New York: Bert G. Hickman,ed.
- Box, GE 1970, *Time Series Analysis, forecasting and control*. SanFrancisco: Holden Day.
- Bradley, H & Sean, L 2006, *Encyclopedia of Measurement and Statistics*. SAGE publications
- Gujarati, DN 2006, *Essentials of Econometrics, 3rd ed*. New York: McGraw-Hill, New York .
- Githingi, VW 2014, Modeling inflation in Kenya using ARIMMA and VAR models. *Unpublished masters of Science in Social Statistics at The niversity of Nairobi School of Statistics* .
- Granger, CW & Newbold, P 1974, Spurious Regression in Econometrics. *Journal of Econometrics*, vol.2 (2) , 111- 120.
- Jarrett, JE 1989, Forecasting monthly earnings per share - Time series model. *International Journal of economics and Management Sciences* .
- Jebb, AT, Tay, L Wang, WW & Huang, Q 2015, Time series analysis for psychological research: Examining and forecasting change, 6:727. *Frontiers in Psychology* .
- Pandey, VS & Abhishek, B 2019, Predictive Efficiency of ARIMA and ANN models: Analysis of Nifty Fifty in Indian Stock Market. *International Journal of Applied Engineering Research, ISSN 0973-4562 Volume 14, No2* , 237-244.
- Rosenberg, B Reid, K. & Lanstein, R 1984, Persuasive evidence of market inefficiency. *Journal of Portfolio Management* 11 , 9-17
- Rangsan, N & Titida, N 2006, *ARIMA model for forecasting oil palm price: Proceedings of the 2nd IMT-GT* . Universiti Sains Malaysia, Penang: Regional Conference on Mathematics, Statistics and Applications.
- Reddy, CV 2019, Predicting the Stock Market Index using Stochastic time series ARIMA modeling: The sample of BSE and NSE. *Indian Journal of Finance* .
- Nelson, CR & Plosser, CI 1982, Trends and Random Walks in Macroeconomic Time Series: Some evidence and implications. *Journal of Monetary Economics* 10 , 139-162.