University of Fort Hare

Together in Excellence

HONOURS RESEARCH PROJECT

TOPIC: Stock market forecasting using time series analysis with ARIMA/SARIMA-ARCH models: A case study of the Johannesburg Stock Exchange

By

Mr. Banele Ngcethane

(202334512)

Submit in partial fulfilment of the requirements for the degree of

Bachelor of Science (Honours) in Applied Statistics

Department of Statistics

Faculty of Science and Agriculture

University of Fort Hare

Supervisor: Dr J Ndege

TABLE OF CONTENTS

DECLERATION

I, Banele Ngcethane, hereby declaring that this project is of my own hard working and with the assistance of my supervisor as far as I'm aware, all direct quotes have been all acknowledged.

I also affirm that I have never submitted this work for any other academic qualification at any institution apart from the University of Fort Hare for the Bachelor of Science Honours Degree in Applied Statistics

Name: Benele Ngcethane (202334512)

Signature B. Ngceth :

Date: 08/11/2023

Supervisor: Dr J Ndege

Signature:

Date

Head of Department: Dr A S Odeyemi

Signature:

Date:

DEDICATION

This project is dedicated to my family: my mother Gladys Ngcethane, my late father Lindile Ngcethane, my late brother Elliot Ngcethane, my siblings Aphiwe Ngcethane and Olwam Ngcethane, and my two daughters Sinothando and Hlaluminathi. I thank God for providing me with such a great family.

AKNOWLEDGEMENTS

I express my gratitude and thanksgiving to God once more for providing me with the chance to work on this project throughout difficult times. I also want to thank the following individuals for their unwavering encouragement and priceless support.

My University of Fort Hare supervisor, Dr. James Ndege, for everything he taught me, his support and advice during this project's writing, and for his encouraging words when I'm feeling low. I express my gratitude to him for looking after my mental well-being during the entire process.

Thank you, Dr. Azeez Adeboye, for all your advice, encouragement, and support during my project.

For allowing me to study and providing support during the academic year, I am grateful to Dr. A. Odeyemi, the department head.

The EMET bursary has improved my life in ways that words alone cannot express. I am incredibly appreciative of their presence and assistance in keeping me from starving to the point where I had to stop my education, and I owe the institution nothing. I sincerely hope they do this for the remaining diligent pupils.

My family, for being aware of the sacrifices and hard work I put in to attempt to improve my prospects at a better life.

I'm so grateful and in love with all of you. I pray that the Lord grants all your heart's desires.

List Of Tables

List Of Figures

ABSTARCT

The study was conducted to forecast stock market prices at the Johannesburg Stock Exchange (JSE) using forecasting models; the Seasonal Autoregressive Integrated Moving Average (SARIMA), ARIMA (Autoregressive Integrated Moving Average) models as well as the Autoregressive Conditional Heteroscedasticity (ARCH) model. The study aimed to combine the models and make the forecast that will handle both linear and non-linear pattern of the financial data effectively including volatility. The data used for this study was a monthly share price data from the JSE website from 1960-01 to 2022-12. Originally the data was not stationary but after the Box-Cox transformation and the first differencing the data was stationary. The seasonal variation was missing from the data and the study showed that the ARIMA (0,0,1) model was the best model in handling the linear dependence in the differenced data based on the Mean Error (ME), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). The residuals of the ARIMA (0,0,1) model were used to fit the ARCH model, and the ARCH (2) model being best in handling volatility changes based on the AIC and BIC. The combination of the ARMA (0,1)-ARCH (2) also showed some improvement in handling and forecasting the share price.

Keywords

ARMA model, ARIMA model, SARIMA model, ARCH model, JSE, ME, RMSE, MAE

CHAPTER ONE

INTRODUCTION

1.1 Background Information

The stock market is usually regarded as a collection of exchanges through which equity shares of public companies are issued, bought, and sold. The primary function of the stock market is to provide a way for companies to raise capital by selling ownership shares to public investors. It also allows private investors to buy shares of stock in public companies and become part owners of their businesses. The aggregate value of the entire stock market is often tracked and reported through market indexes, such as the Johannesburg Stock Exchange.

When a private company wishes to become public, it typically completes what is known as an Initial Public Offering, or IPO. During the IPO process, the company sells shares of stock to public investors to raise the money it needs to pay off debt or invest in its business. Once the IPO is complete, these shares of stock begin trading on one or more of the stock exchanges that make up the stock market.

Public investors can track the real-time market price of a stock by monitoring its ticker symbol, typically an abbreviation of no more than five letters representing a particular company's stock. Investors buy shares of stock in the hope that the company selling those shares will grow to become more valuable over time, thereby increasing the price of each share of stock. Stock prices are determined by the economic law of supply and demand, which fluctuate daily based on changes in investors' demand.

The base where traders, investors, buyers, and sellers congregate to establish the market price for each investment in the industry of interest is known as the stock market (Freddie, 2021) The presence of historical data for stock markets has been analyzed using forecasting time series

(FTS). Forecasting means estimating what can occur in the future given what is known in the past and what is happening currently (Shruti, 2019). The stock market is sometimes viewed by society as being either extremely dangerous for investments or unsuitable for trading. Despite the risk involved in the stock market, investors still have good trust in investing since it has great economic potential (Nashirah & Sofian, 2017).

The stock market is essential for obtaining funding from investors and facilitating access to firm shares in both emerging and developed countries. Financial markets have a considerable impact on society's economy, and governments' liberalized and globalized policies stimulate the rise of global business. Through financial networks, they enable the movement of savings, capital, and investments between the supply and demand sides of the economy (Ruipeng & Rangan , 2022).

The study will focus on the Johannesburg Stock Exchange (JSE), which has a market capitalization of over 850 billion USD and is the largest in Africa. It is followed by the Nigerian Stock Exchange, the Egyptian Exchange, and others (Gideon & Paul, 2014). These stock exchanges provide a platform for buyers and sellers to buy and sell shares of companies that are listed on the exchange. This platform benefits South Africa's real economy in several ways, including by giving businesses access to liquidity, facilitating savings and investments, and giving market participants transparency to make informed financial and economic decisions. However, it is worth noting that the rankings can change over time depending on several factors, including regulatory changes and market performances (Surendran, 2020).

Financial information is crucial in the stock market since it affects stock prices together with macroeconomic factors, monetary policy, international trade policy, and economic policy. Investors can use this information to decide whether to invest in a particular company's shares. Changes in government policies, rules, and norms both nationally and internationally, as well as economic crises, natural catastrophes, variations in the price of oil internationally, the impacts of inflation, foreign exchange rates, and inflation effects, all have an impact on stock prices and returns. Maximizing profits on investments is the primary goal of an investor. Many knowledgeable investors use a variety of approaches, including financial and technical analysis, to improve stock price prediction. Additionally, investors depend on financial and stock market professionals to research and provide suggestions made by analysts. Financial professionals and investors utilize these fundamentals or technical research and prediction algorithms to estimate future share values. (Makatjane & Moroke, 2021).

The tendency of markets to perform better or worse during specific times of the year is commonly referred to as seasonal trends in stock prices (Wei, Kee, & Musa, 2023). The stock markets are still significantly impacted by seasonal anomalies such as calendar anomalies, festival season, and weekend impacts (Wei, Kee, & Musa, 2023). Additionally, Volatility is a statistical indicator of the variation in stock market returns (Bakar & Rosbi, 2017). Highly volatile stocks come with more risk and potential losses for investors, but they can also occasionally provide significant returns.

Since the data is frequently non-stationary and typically displays non-linear patterns, predicting financial time series is a challenging task. The predominant linear techniques that estimate the movement of time series dependent on both the linearity of historical observations named the Autoregressive (AR) model, and error terms called the Moving Average (MA) model traditionally known as Autoregressive Integrated Moving Average (ARIMA) models in combination (Box, Jenkins, Reinsel, & Ljung, 1976)

Model identification, parameter estimation, and diagnostic testing are the three processes that make up the Box-Jenkins technique, which is used to fit the ARIMA model. Non-stationary data, either through the mean or variance, which are primarily visible in the stock market data, can be transformed and stabilized using several kinds of transformations that may be employed to the ARIMA model. The initial step is necessary to eliminate the trend and seasonal variations if the data show signs of being non-stationary (including visibility of trend or seasonality) and applying power transformations to the time series is the efficient technique, such as taking its logarithm, square root, or cube root, and some of the most straightforward and most efficient methods to make the variance constant over time.

The data can also be transformed by differencing it to get rid of trends and seasonality. When required, a researcher may employ the moving average approach or seasonal differencing to eliminate seasonality. Seasonal ARIMA (SARIMA), an extension of the ARIMA model, is capable of being used for seasonal and non-stationary data where there is a seasonal variation in the time series, following the approach of (Box, Jenkins, Reinsel, & Ljung, 1976). The level of precision prediction results might decrease if a linear condition is violated. In addition, there are instances where converting the data doesn't seem to resolve the issue when the time series displays non-constant variance in the errors or heteroscedasticity. Consequently, SARIMA does not constitute an acceptable approach for forecasting heteroscedastic time series. In that case, the Autoregressive Conditional Heteroscedasticity (ARCH) model and its expanded versions are among the best-suited time series forecasting models for handling changing variance (Shaik & Sejpal, 2020).

The global stock markets have been the subject of studies using time series models to forecast fluctuations. Many investigators have used several techniques, including several time series models. The study seeks an improvement in the comparison of time series models using Seasonal Autoregressive Integrated Moving Average (SARIMA)/ ARIMA and Autoregressive Conditional Heteroskedasticity (ARCH) models to forecast the stock market price.

1.2 Research problem statement/motivation

The modelling of time series data mean level is done using ARIMA/SARIMA (where seasonality is considered). The two models do not consider the variance or squared residuals (volatility) in the data. This study uses both the ARIMA/SARIMA and ARCH models where heteroscedasticity is considered. Utilizing time series model ARIMA/SARIMA-ARCH increases the accuracy, to reliably forecast stock market by considering the error variance of the residuals. The stock market exchange is complex to predict since it depends on several variables, including interest rates, inflation, conflicts, and many other volatility factors. A good model that can be able to handle some of these such as ARIMA/SARIMA-ARCH is quite useful.

Accurate stock market price predictions can aid traders, investors, and businesses in making wiser choices.

The research questions to be answered in this study are:

i) How will the addition of the ARCH model to the ARIMA/SARIMA model improve the accuracy
of stock market price predictions in comparison to the ARIMA/SARIMA model?

ii) Is there a difference in the performance of stock market price predictions before and after combining the time series models ARIMA/SARIMA and ARCH?

iii) Does taking Volatility into account assist in predicting stock market prices more accurately?

1.3 Research Aim, Objectives, and Hypotheses

1.3.1 Aim of the Study

The main aim of the study is to employ the time series ARIMA/SARIMA-ARCH models to examine and forecast stock market prices using Johannesburg Stock Exchange data from 1960-2022.

1.3.2 Objectives of the study

The objectives of the study are:

1. To assess how seasonality affects the prediction of stock market prices at the Johannesburg Stock Exchange (JSE).
2. To evaluate how the introduction of the ARCH model influences the ARIMA/SARIMA models capability in estimating stock market prices.
3. To assess predictive accuracy and performance of ARIMA/SARIMA-ARCH model in forecasting stock market price in comparison to the ARIMA/SARIMA model.

1.3.3 Research hypotheses

The research hypotheses are as follows:

H_0: Seasonality does not affect stock market prediction at the Johannesburg Stock Exchange (JSE).

H_1: Seasonality affects stock market prediction at the Johannesburg Stock Exchange.

H_0: The ARCH model has no influence over the ARIMA/SARIMA models in estimating the stock market prices at JSE.

H_1: The ARCH model has influence over the ARIMA/SARIMA models in estimating stock market prices at JSE.

H_0: The ARIMA/SARIMA-ARCH models are not better model than the ARIMA/SARIMA model in estimating stock market prices at JSE.

H_1: The combined ARIMA/SARIMA-ARCH models are better model than the ARIMA/SARIMA models in estimating stock market prices at JSE.

1.4 Significance of the study

This study focuses on enhancing stock market volatility prediction, which is crucial for risk management in financial markets. Various time series models are commonly used for this purpose, but they often overlook heteroskedasticity, a key factor in financial data. Specifically, the study aims to optimize the ARIMA/SARIMA-ARCH models, which combines ARIMA/SARIMA for season and trend analysis and ARCH model (Engle, 1982), for handling changing variance. The ARIMA/ SARIMA model (Box, Jenkins, Reinsel, & Ljung, 1976), is effective for forecasting non-seasonal and seasonal trends, while ARCH addresses issues like non-constant variance, making it valuable for predicting financial time series with volatility. By leveraging the ARIMA/SARIMA-ARCH models, this research seeks to improve the accuracy of JSE stock price forecasts while accounting for volatility, a critical aspect in risk management for traders and investors.

Risk management in financial markets depends on being able to predict stock market volatility accurately. Volatility projections are used by traders and investors to manage risk, modify their investment plans, and protect themselves from potentially dangerous market swings. The goal of this study is to analyze the data and come up with the best parameters of the SARIMA-ARCH model for forecasting stock market prices at the JSE. ARIMA/SARIMA was developed by (Box, Jenkins, Reinsel, & Ljung, 1976), to offer a robust analytical tool for evaluating and forecasting time series data that exhibit both non-seasonal and seasonal trends. However, it has some shortcomings. The ARCH model (Engle, 1982), solves some of these issues, one of them being the issue of non-constant variance (heteroscedasticity).

The ARCH model is essential in analyzing and predicting financial time series data in the presence of volatility. Given how difficult it is to predict stock market prices in the presence of volatility, this study will use the ARIMA/SARIMA-ARCH models to examine the problem of volatility in forecasting stock market prices. This study uses time series models

(ARIMA/SARIMA and ARIMA/SARIMA-ARCH) to compare and predict the stock market price of the Johannesburg Stock Exchange. When used to analyze stock market data, time series models (MA, AR, ARMA, ARIMA, exponential smoothing, SARIMA, etc.) can occasionally improve accuracy depending on the characteristics of the data, but they do not consider heteroscedasticity (unequal variance or non-constant variance) into consideration. This study will use the ARIMA/SARIMA-ARCH models and contrast them to their baseline model (ARIMA/SARIMA) to predict stock market prices on the Johannesburg Stock Exchange (JSE).

1.5 Limitations

The study is limited to:

i. Historical data available from the Johannesburg Stock Exchange (JSE).
ii. Univariate modeling.
iii. Volatility when using the ARCH model requires a large value of the lag q, and this results in estimating many parameters, which may cause difficulty.
iv. There are other analytical methods that could be used, but this study is based only on the ARIMA/SARIMA-ARCH models.

CHAPTER TWO

LITERATURE REVIEW

This section deals with a brief literature review of studies carried out using time series models in estimating /predicting stock market prices:

The study by Zahangir et. al., (2013) in Dhaka, Bangladesh, using data from Dhaka Stock Exchange (DSE20 and DSE general index), from 1 December 2001 to 10 September 2011, examined the Autoregressive Conditional Heteroscedasticity (ARCH) models on predicting volatility using daily data. The results of the Power ARCH (PARCH) model demonstrated how strongly previous volatility affects current volatility. The DSE20 and DSE general indexes return series showed a considerable influence of last volatility on recent volatility according to the Thresh ARCH (TARCH) model, the metrics, Root Mean Absolute Error (RMAE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The Theil's Inequality Coefficient were used to evaluate the model, for DSE20 index returns, and the ARCH and PARCH models were chosen as the top-performing models, while ARCH model outperformed all the other ARCH models for the DSE general index return series.

Ahmed et al., (2017) conducted a study using ARCH models on Stock market Indexes of Pakistan, India, Bangladesh, and China respectively using data from 1 January 1996 to 31 December 2015. The stationarity was tested using the Augmented Dicky-Faller (ADF) test to check the presence of the unit root in all Stock Exchange and it was found that the data suffers from a unit root, and the first difference was taken to make the data stationary and tested again using the ADF test and the data was found to be stationary, the normality test was carried using the Jarque-Bera (JB) test and the data was found to be not normal since all the stock indexes were skewed to the right. The conclusion was according to the Schwarz Information Criterion (SIC), the Generalized ARCH (GARCH) outperformed all the other four ARCH models, on the Bombay Stock Exchange (BSE)-SENSEX, and Shanghai Stock Exchange (SSE) composite. The Exponential GARCH

(EGARCH) outperformed all the other four ARCH models on the Dhaka Stock Exchange (DSE 20), while the PARCH and EGARCH outperformed all the other ARCH models on the Karachi Stock Exchange (KSE) index.

Ganbold et al., (2017) proposed a heightened approach to modeling and forecasting exchange rate volatility in Turkey and addressing the challenges of political turbulence and uncertainty. The methodology used daily exchange rate data from 2005-2017. Autoregressive Conditional heteroskedasticity (ARCH) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models were used to forecast the volatility. The study used monthly data forecasting using ARIMA, Seasonal ARIMA (SARIMA), and Structural Vector Autoregression (SVAR) methodologies. According to the effects of SVAR and SARIMA, the currency rate depreciated from March 2017 to April 2017 before appreciating again from May 2017 to September 2017 and further declining again against the Turkish lira. The SARIMA model forecast was predicted to be more accurate than the other models based on the forecast comparison by Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).

A study from Algeria by (Boudrioua & Boudrioua, 2020) adopted the Box-Jenkins technique and Seasonal ARIMA to predict monthly DZAIRINDEX returns from June 2010 to May 2020 on the Algiers Stock Exchange (ASE). The data was found to be stationary without taking any difference using the Augmented Dickey-Fuller (ADF) test. The Seasonal Autoregressive Integrated Moving Average SARIMA (2,0,0) $(0,0,1)_{12}$ model with zero mean was determined to be the best model for predicting monthly DZAIRINDEX returns in Algeria based on Akaike's Information Criterion (AIC), as compared with other Seasonal ARIMA models, and the residuals were found to be normally distributed using the Ljung-Box test and other normality tests.

Kemalbay & Korkmazoglu (2021) carried out a study in Turkey on a comparative approach of SARIMA-ARCH and GP (Genetic Programming) models on stock market prediction at the Istanbul Stock Exchange National 100 (XU100) index using data from 2 January 2017 to 22 February 2019, and compared the obtained results, with the help of evaluating the metrics; Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), symmetric MAPE (sMAPE), Mean Error (ME), and Theli's U. The SARIMA-ARCH model was found to vary by a minimal number of 0.00022 on RMSE with GP model but having the same Mean Square Error (MSE) and Thuli's U with little or no difference on the MAPE and SMAPE. The Genetic Programming model outperformed the SARIMA-ARCH model based on the characteristics of the data.

Wei et al., (2023) conducted a study to examine seasonal versus non-seasonal trends in the stock market in Malaysia with ARIMA and SARIMA models using weekly data from Yahoo Finance, from 2016 to 2021. The study used six different industries/companies to analyze seasonality and non-seasonality. First a decomposition technique was applied to divide the data into seasonal, cycle components, and trends to develop a more accurate model, and then used forecast errors and AIC to measure the difference between actual demand and the forecast demand. After using both the ARIMA and SARIMA models, they compared and analyzed both the model residuals by plotting histogram graphs and checked for normality without de-seasonalized method of ARIMA (0, 0, 0) and de-seasonalized method of the SARIMA (0, 0, 5) $(1, 0, 0)_{52}$ model. In conclusion, the study analyzed Sum of Squares Error (SSE) values and proved that the seasonality produced better fits than the model without seasonality, and out of the six industries, four of them fitted best in the presence of seasonality.

Dikko et al., (2015) used the ARCH models to model abrupt shifts in the Nigerian Insurance Stock using data from the Nigerian Stock Exchange from 2 January 2000 to 14 May 2014, focused on ten Insurance companies. The data was tested for stationarity using the ADF test for all ten insurance companies, and it was found that it was stationary without taking any difference at 0.01 level of significance. Further tested for normality, the Jarque-Bera Test was used, and the data was found to be not normally distributed. The presence of the ARCH effect was tested using the language Multipliers for ARCH test, and the null hypothesis that there is no ARCH effect was rejected in favor of the alternative hypothesis that there is an ARCH effect. The conclusion was that the ARCH (1) model outperformed all the Twelve ARCH models, being suitable and competitive in considering volatility.

A study carried out by Aminu et. al., (2021) used the Box-Jenkins approach to develop a model with seasonality to forecast Nigeria's Exchange rate using the monthly average exchange rate from January 1981 to December 2018. The study used the ADF and Phillips-Perron (PP) tests to check for stationarity of the data, and the original data was found to be non-stationary, and the differencing approach was done to make the data stationary and used the two tests, i.e., ADF and PP test to check if the data was stationary and found to be stationary after the first difference at 1%, 5%, and 10% level of significance. The Auto-Correlation Function (ACF) and Partial ACF (PACF) were plotted, and found that the SARIMA (0, 1, 1) $(1, 1, 1)_{12}$ model performed better according to the Akaike Information Criterion (AIC) and Bayesian IC (BIC) values under the rule of the lower AIC and BIC values the better the model. The forecast was done both on the in-sample and out-off sample to check the accuracy of the model. The out-of-sample used was for 2019 data, and the SARIMA (0, 1, 1) $(1, 1, 1)_{12}$ model was chosen to be the best model to make decisions about Nigeria's Exchange rate.

Nwokike & Okereke (2021) conducted a study to compare the performance of Seasonal Artificial Neural Networks (SANN), SARIMA, and ARIMA models to forecast the quarterly GDP of Nigeria from the first quarter of 1960 to the fourth quarter of 2014. For the Box-Jenkins techniques, the data was transformed as it showed changing variances that was confirmed using the ADF test. The conclusion was that the SANN (4-10-4) model outperformed the Box-Jenkins methods with MSE and RMSE of 0.041 and 0.20, respectively, and the SARIMA (0, 1, 0) (0, 1, $1)_4$ model outperformed the ARIMA (2, 1, 2) model based on the MSE of 0.527, and RMSE of 0.73 while the MSE and RMSE of ARIMA (2, 1, 2) model were 0.705 and 0.84 respectively.

Another study by Masimba & Shadreck (2021) in Zimbabwe compared the ARCH models in forecasting the Volatility of the Zimbabwean Stock Exchange from 2009 to 2015. The study first tested the stationary using the ADF test, and the results clearly indicated that there was no unit root, and the data was stationary. Further, the ARCH effects were tested using the ARCH-LM (Langrage Multiplier) test, and the null hypothesis that no ARCH effect was rejected in favor of the Alternative hypothesis that there were ARCH effects, and the normality was checked using Jarque Bera test and found that the residuals were not normally distributed. The conclusion was the use of the ARCH (5) model in forecasting Zimbabwean Stock Exchange volatility compared to other ARCH models.

A study by Shannon & Samouilhan (2008) used the ARCH model with its extensions and GARCH models to forecast volatility on the Johannesburg Stock Exchange using data from the McGregor database and SAFEX/Cadiz from 01-Febuary-2004 to 28-September-2006. The study used in sample model selection and out-of-sample forecast to determine the best-performing model, and from the in-sample and out-of-sample parameter and model selections, the conclusion was based on the MSE, RMSE, Mean Absolute Error (MAE), Mean Error (ME), then the Thresh ARCH (2, 2). The (TARCH) model outperformed all the other ARCH models, providing the best volatility forecasting, specification, and more accurate and unbiased on one week and one day ahead.

Sepato (2017) conducted a study in South Africa and compared the GARCH and ARCH models in the presence and absence of outliers using the data from the Johannesburg Stock Exchange database, from 3 January 2011 to 21 April 2016. The data was tested for stationarity using the ADF and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) and found to be not stationary. The log transformations were done to make the non-stationary data to be stationary and tested using the stationary tests. The study used ACF and PACF to determine the parameters of the models. The RMSE, MAE, and MAPE were evaluated for performing ARCH and GARCH models, and Ten models were found to be competitive in the presence of outliers based on the AIC, BIC, Hannan-Quinn Information Criterion (HQIC). The results showed that the (Autoregressive Moving average) ARMA (0, 2)-EGARCH (1, 1) was the best model based on the MSE, MAE, AIC, Shwarz's Bayesian Criterion (SBC), and MAPE.

The study by Pillay (2020) proposed an ARIMA model to forecast the share price index of the Johannesburg Stock Exchange (JSE) from 1 August 2019 to 31 July 2020. The ADF test was used for stationarity test, and the log transformation was applied to stabilize the variance, and the first difference was taken to make the data stationary. Model selection, and parameter estimation was done using ACF, PACF, AIC, BIC, ME, RMAE, MAE, and MAPE, and the proposed model was found to be ARIMA (4, 1, 4) model, which proved to be a more stable and suitable model than other ARIMA models. The normality of the residuals of the ARIMA (4, 1, 4) model were checked and found to be normally distributed using the Quantile-Quantile plot (Q-Q plot), ACF, and the Ljung-Box tests. The conclusion was that the use of the ARIMA (4, 1, 4) model was the best on forecasting the share price index of the JSE for that period.

Makatjane, K & Moroke N (2021) carried a study to compare SARIMA, Markov-Switching Exponential Generalized Autoregressive Conditional Heteroscedasticity (MS-EGARCH), Generalized Extreme Value Distribution (GEVD) and SARIMA-MS-EGARCH-GEVD to Predict

Extreme Daily Regime Shifts in Financial Time Series in South Africa from the Johannesburg Stock Exchange (JSE) from a period beginning 4 January 2010 to 31 July 2020. The SARIMA model was done first, and the first difference was applied on both non-seasonal and seasonal components to remove the trend and make the data stationary. The study used the residuals of the SARIMA model to perform model selection for MS-EGARH-GEVD. The study concluded that SARIMA (2, 1, 0) $(2, 1, 0)_{240}$-MS (2)-EGARCH (1, 1)-GEVD is the best model based on the AIC, and BIC, with a performance, sensitivity, and specificity of 98%, 79.89%, and 98.40% respectively.

Kaseke (2015) conducted a study to model Volatility on the Stock Exchange using ARCH and GARCH models from the Johannesburg Stock Exchange in South Africa using daily data from January 2003 to December 2013 and monthly data from January 1990 to December 2013. The study analyzed three companies from the Johannesburg Stock Exchange. Under the daily data, the closing price, returns, squared returns, ACF, and PACF of the squared returns and returns of these companies were analyzed to understand the data and develop models. The GARCH (1, 1) models and its extensions AR (4)-GARCH (1, 1), AR (2)-GARCH, AR (1)-GARCH (1, 1) and IGARCH (1, 1) models outperformed the ARCH (1) model.

The literature reviews above have discussed some brief outlines of some statistical methods that have been used in many instances to analyze and predict financial time series data with various preferable outcomes (models) based on the available data from diverse financial markets.
The next chapter discusses the methodology adopted specifically for this study.

CHAPTER THREE

METHODOLOGY

3.1 Introduction

The SARIMA stands for Seasonal Autoregressive Integrated Moving Average, and ARIMA for (Autoregressive Integrated Moving Average), thus they forecast mainly the stock market prices based on past values and past forecast errors and including seasonality to help in forecasting the stock market price more accurately. The ARCH model (Autoregressive Conditional Heteroscedasticity) on the other hand, is designed to capture the volatility clustering observed in financial markets, where periods of high Volatility tend to be followed by more periods of high Volatility. The model allows for time-varying variance in the data, making it more flexible than traditional linear models. ARCH model focuses on the variance so that the model can handle the non-constant variance (heteroskedasticity) and volatility clustering by combining both ARIMA/SARIMA and ARCH models (ARIMA/SARIMA-ARCH), making the model more accurate in forecasting stock market prices and helps in understanding future prices of the stock markets where both seasonality and Volatility are involved (Kemalbay & Korkmazoglu, 2021).

3.2 ARIMA/SARIMA-ARCH Models

3.2.1. The ARIMA/SARIMA models

The SARIMA model is an expansion of the ARIMA model that includes seasonal components. It was developed to incorporate both seasonal and non-seasonal trends in time series data.

The SARIMA models are made up of the following components:

3.2.1.1 non-seasonal component

AR (P) denotes Autoregressive Process of order P:

It represents the relationship between the current value of the time series and its past values.

The model is given as:

$$\mathbf{y_t = \alpha_0 + \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + \varepsilon_t \qquad \varepsilon_t \sim iid\left(0, \sigma_t^2\ \right)}$$

Where t denotes the time, p is the order of the AR model, representing the number of lagged terms of the time series that are used as predictors for the current value, y_t is the value of the time series at time t, $\boldsymbol{\alpha_0}$ is the constant term also known as intercept, $(\boldsymbol{\alpha_2}, \boldsymbol{\alpha_2}, \cdots, \boldsymbol{\alpha_p})$ are the Autoregressive coefficients, these coefficients determine the impact of the previous p-value on the current value, $(\mathbf{y_{t-1}}, \mathbf{y_{t-2}}, \cdots, \mathbf{y_{t-p}})$ are the past observations of the time series, $\boldsymbol{\varepsilon_t}$ is the error term at time t, representing the random noise or unpredicted component of the time series at that time t.

MA (q) denotes (The Moving Average process of order q):

It represents the relationship between the time series and past forecast errors.

With the model given as:

$$\mathbf{y_t = \boldsymbol{\beta}_0 \boldsymbol{\varepsilon}_t + \boldsymbol{\beta}_1 \boldsymbol{\varepsilon}_{t-1} + \cdots + \boldsymbol{\beta}_q \boldsymbol{\varepsilon}_{t-q}} \qquad \boldsymbol{\beta}_0 = \mathbf{1}$$

Where the q parameter represents the number of lagged past forecast errors that are used in the model as predictors for the current value, $\mathbf{y_t}$ is the value of the time series at time t. $(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_q)$ are the moving average coefficients. These coefficients determine the impact of the past q forecast errors on the current value.

I (Integrated):

Represents the number of non-seasonal differences needed to make the time series stationary (i.e., constant mean and variance)

The general form of differencing in a time series is represented as follows:

The first order differencing is $\boldsymbol{\nabla} = \boldsymbol{y_t - y_{t-1}}$

3.2.1.2 Seasonal component

This captures the seasonal trend consisting of:

Seasonal Autoregressive (SAR), Seasonal Integrated (SI), Seasonal Moving average (SMA).

SARIMA$(\mathbf{p}, \mathbf{d}, \mathbf{q})(\mathbf{P}, \mathbf{D}, \mathbf{Q})_{\mathbf{S}}$ is given by:

$$\boldsymbol{\Phi_P(B^S)\phi_p(B)(1 - B^S)^D(1 - B)^d y_t = \theta_q(B)\Theta_Q(B^S)\varepsilon_t}$$

Where:

$\emptyset_{\mathbf{p}}(\mathbf{B}) = \mathbf{1} - \emptyset_{\mathbf{1}}\mathbf{B} - \cdots - \emptyset_{\mathbf{p}}\mathbf{B}^{\mathbf{p}}$

$\boldsymbol{\theta}_{\mathbf{q}}(\mathbf{B}) = \mathbf{1} - \boldsymbol{\theta}_{\mathbf{1}}\mathbf{B} - \cdots - \boldsymbol{\theta}_{\mathbf{q}}\mathbf{B}^{\mathbf{q}}$

$\boldsymbol{\Theta}_{\mathbf{P}}(\mathbf{B}) = \mathbf{1} - \boldsymbol{\Theta}_{\mathbf{1}}\mathbf{B}^{\mathbf{S}} - \cdots - \boldsymbol{\Theta}_{\mathbf{P}}\mathbf{B}^{\mathbf{PS}}$

$\boldsymbol{\Theta}_{\mathbf{Q}}(\mathbf{B}) = \mathbf{1} - \boldsymbol{\Theta}_{\mathbf{1}}\mathbf{B}^{\mathbf{S}} - \cdots - \boldsymbol{\Theta}_{\mathbf{Q}}\mathbf{B}^{\mathbf{QS}}$

B represents the backward shift operator, y_t is the observed value at time t, ε_t is for the estimated residual error at time t, with a mean zero and constant variance σ^2, i.e iid $(0,\ \sigma_\varepsilon^2)$

Non-seasonal components:

$\emptyset_{\mathbf{p}}(\mathbf{B}) = \mathbf{1} - \emptyset_{\mathbf{1}}\mathbf{B} - \cdots - \emptyset_{\mathbf{p}}\mathbf{B}^{\mathbf{p}}$

$\boldsymbol{\theta}_{\mathbf{q}}(\mathbf{B}) = \mathbf{1} - \boldsymbol{\theta}_{\mathbf{1}}\mathbf{B} - \cdots - \boldsymbol{\theta}_{\mathbf{q}}\mathbf{B}^{\mathbf{q}}$

Seasonal components:

$\boldsymbol{\Theta_P(B) = 1 - \Theta_1 B^S - \cdots - \Theta_P B^{PS}}$

$\boldsymbol{\Theta_Q(B) = 1 - \Theta_1 B^S - \cdots - \Theta_Q B^{QS}}$

3.3 ARCH Models

The ARCH models are statistical approaches that characterize the variance of the present error term. They are frequently used in the modeling of financial time series that exhibit time-varying volatility and volatility clustering. They are helpful when the error variance in a time series follows an AR model. It is the first conditional autoregressive model that takes error variation (heteroskedasticity) into account. The mean of this model is equal to zero, with a constant variance. According to (Dinardi, 2020), the order of squared error values for earlier periods may have an impact on the conditional variance. The ARCH model is described as follows:

Assume ε_t is a time series with mean equal to zero and variance equal to σ_t^2.

Then the model for ARCH (q) model can be written as:

$$\boldsymbol{\sigma_t^2 = \Omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2}$$

$$\boldsymbol{\varepsilon_t = \sigma_t z_t \qquad z_t \sim iid\left(0, \sigma_t^2\right)}$$

$\boldsymbol{\sigma_t^2}$ is the estimated conditional variance, $\boldsymbol{\varepsilon_t}$ is the residual return, $\boldsymbol{\varepsilon_{t-i}^2}$ are squared residuals at time $\boldsymbol{t-i}$, $\boldsymbol{\Omega > 0}$, $\boldsymbol{\alpha_1 \geq 0}$, for, $\boldsymbol{i = 1, 2, \cdots, q}$.

3.4 Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

The p and q variables for the ARIMA model are chosen using the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF), which are important techniques for studying linear time series. A correlogram demonstrates the two sets of observations are dependent on one another by illustrating how correlated they are at various time lags. The ACF and PACF may also be used to locate periodicities, outliers, and category time series, as well as to fit autoregressive models and identify models (Alharbi & Csala, 2022).

The outliers will be the most significant when the ACF values are near -1. A few consecutive outliers can improve the forecast when it has high positive values by compensating for and overcoming the bias of the limited sample. Although various adjustments can reduce the ACF impact of outliers, their biases still exist asymptotically (Alharbi & Csala, 2022). On the other hand, the PACF offers an interesting perspective for viewing time series structure and offers sufficient criteria for a list of real numbers to characterize weak stationary time series.

3.5 The Augmented Dickey-Fuller (ADF) Test and the Null Hypothesis

According to (Virginia, Ginting, & Elfaki, 2018), the augmented Dickey-Fuller (ADF) test is used to evaluate if a collection of data is stationary or non-stationary. High-order differencing is a technique used to make non-stationary data stationary.

ADF test can be determined according to the two values named below:

1. The Statistic value.

The unit root hypothesis is rejected by a significant negative statistic value, which shows that the time series has no unit root and is thus stationary. The time series has a unit root and is thus non-stationary if the ADF test statistic result is positive, showing that the null hypothesis of a unit root is not rejected. Furthermore, the ADF test often entails the application of the null hypothesis to evaluate the data as stationary or non-stationary by labeling them as H_1or H_0.

H_1 denotes a time series without a unit root; in this situation, the null hypothesis is rejected, and the data is stationary. Alternatively, H_0 denotes a time series with a unit root; in this situation, the null hypothesis is not disproved, and the data are not stationary.

2. The p-value.

The p-values that are used to decide between H_1 and H_0 are shown below.

1. If f $\boldsymbol{p \leq 0.05}$, no unit root, the null hypothesis is rejected, and the data is stationary.
2. If $\boldsymbol{p > 0.05}$ has a unit root, the null hypothesis is not rejected, and the data is non-stationary (Fan, Xiao, & Wang, 2014).

The data to be used in the analysis is from the Johannesburg Stock Exchange and the website is

https://data.oecd.org/price/share-prices.htm

sharepricedata
(2023).xlsx

CHAPTER FOUR

DATA ANALYSIS AND RESULTS

4.1 Introduction

The data used in the analysis is from the Johannesburg Stock Exchange (JSE), a monthly data from January 1960 to December 2022 as indicated on the website given on the methodology section. The statistical methods applied in the analysis are based on the Box, Jenkins, Reinsel, & Ljung (1976) and (Engle, 1982), with the ARIMA/SARIMA and ARCH models using R programming.

Figure 4.1 shows the data components which are the observed, trend, seasonality, and errors. The second figure (Figure 4.2) indicates the Share price data for the same period from January 1960 to December 2022 and Figure 4.3 shows the ACF and PACF of the data indicating that the data is not stationary as the ACF decays slowly and exponentially with more significant spikes.

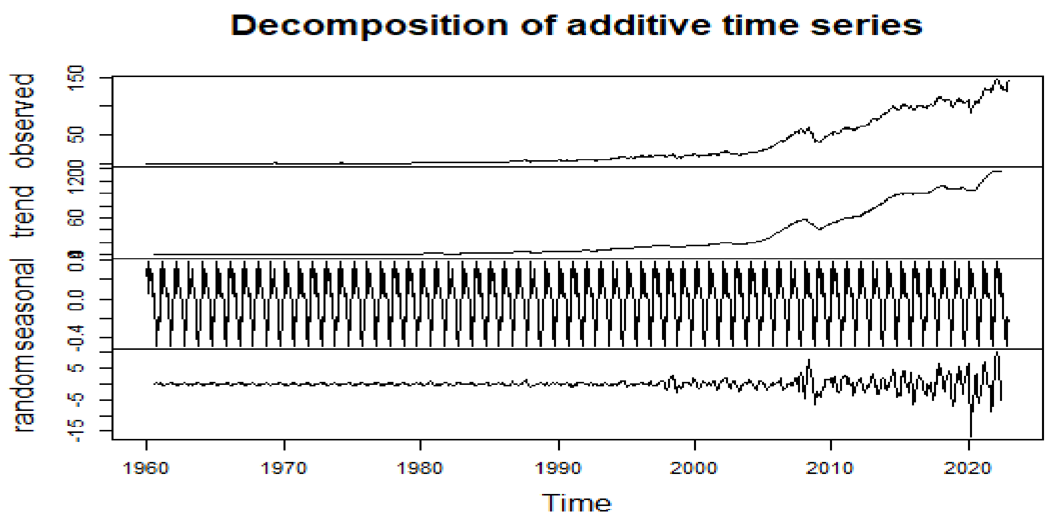


Figure 4.1. Decomposition of the time series data

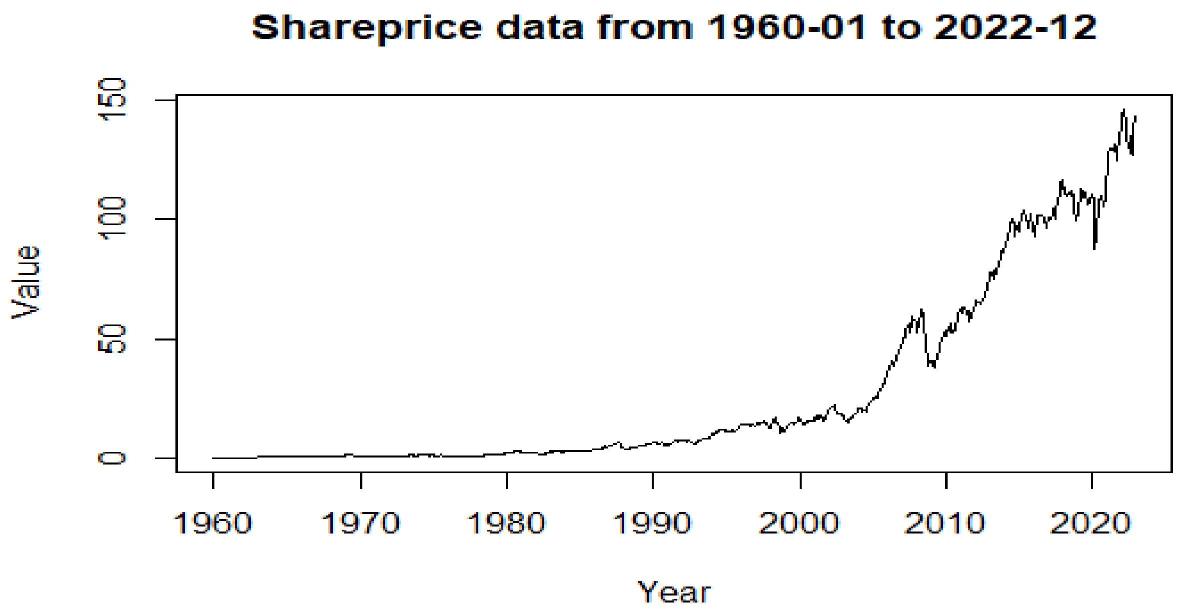


Figure 4.2 Share prices from Jan 1960 to Dec 2022

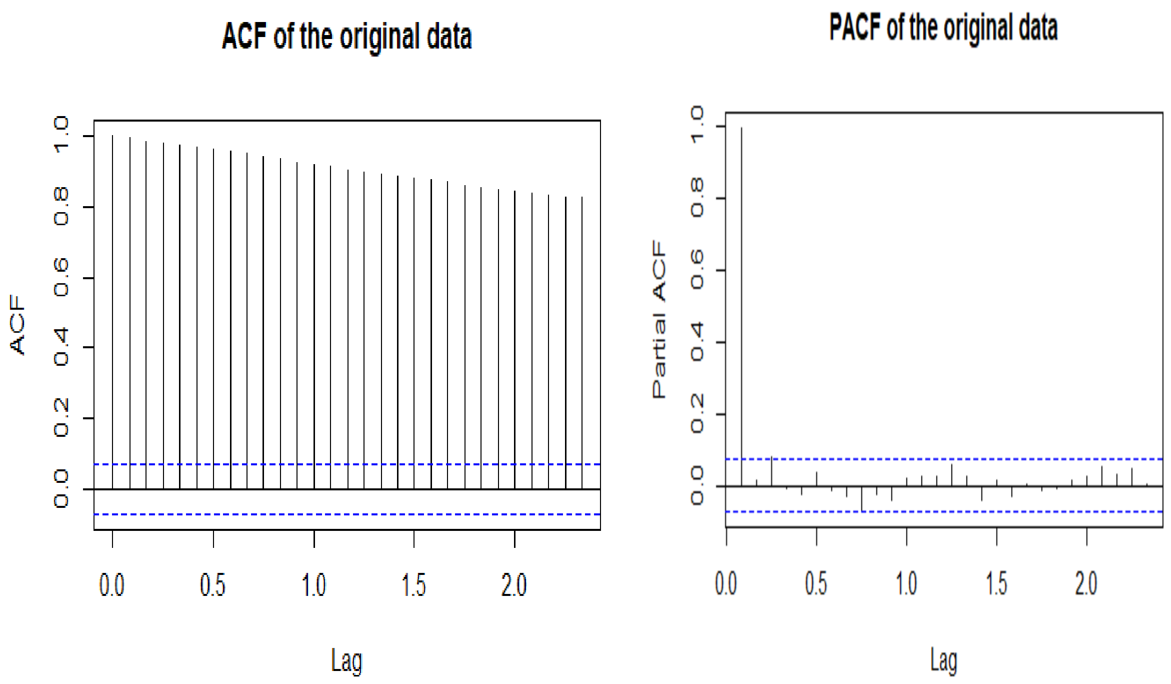


Figure 4. 3 Plot of time series data with ACF and PACF

4.2 Stationarity test

The plots above were done to explore the nature of the data. The data showed some trends and non-constant variance. The stationarity check was done using the ADF and PP tests.

The null hypothesis under the ADF and PP tests, stated that, the data contained a unit root and is not stationary, and the alternative hypothesis stated that the data contained no unit root and was thus stationary. Since the p-values of both tests in table 4.1 and table 4.2 are greater than 0.05, and thus the data is not stationary on both the original data and the transformed data.

Augmented Dickey Fuller (ADF) Test			Phillips-Perron Unit Root Test		
statistic	lag	p-value	statistics	lag	p-value
0.5003	6	0.99	0.26285	6	0.99

Table 4.1 Test for stationarity with ADF and PP tests

Augmented Dickey Fuller (ADF) Test			Phillips-Perron Unit Root Test		
statistic	lag	p-value	statistics	lag	p-value
-2.941516	6	0.17936178	-3.2512	6	0.0791

Table 4.2. ADF and PP tests on a transformed data

4.3 Transformation

Figure 4.4 shows the transformed data using the Box-cox transformation to make the variance constant and improve the trend from the original data. The ACF of the transformed data indicates that the data is not stationary, and the mean is constantly showing an increasing trend. The

ADF and PP tests also confirm that the data is still not stationary as it contains a trend (figure 4.5).

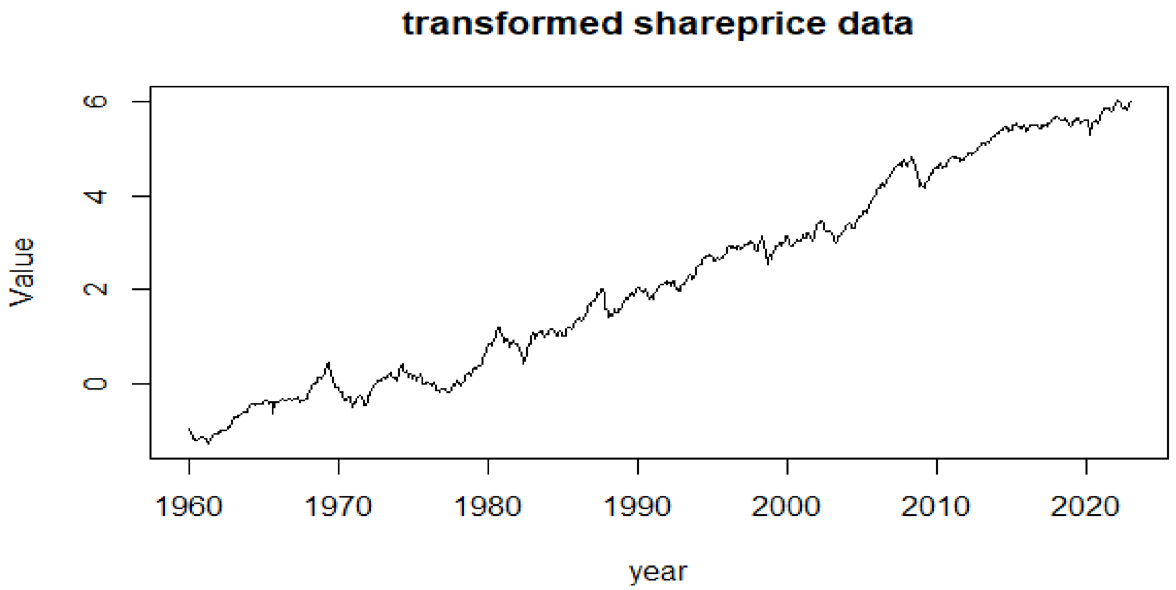


Figure 4.4 Transformed time series data using Box-Cox transformation

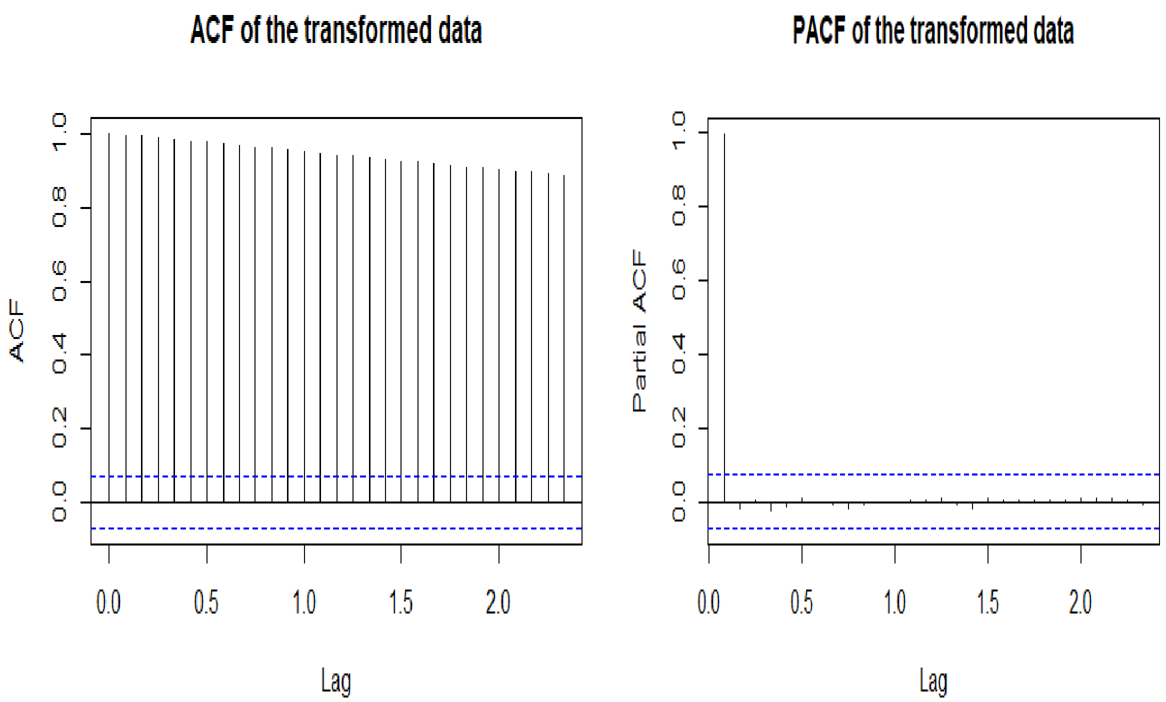


Figure 4.5 ACF and PACF of the transformed data

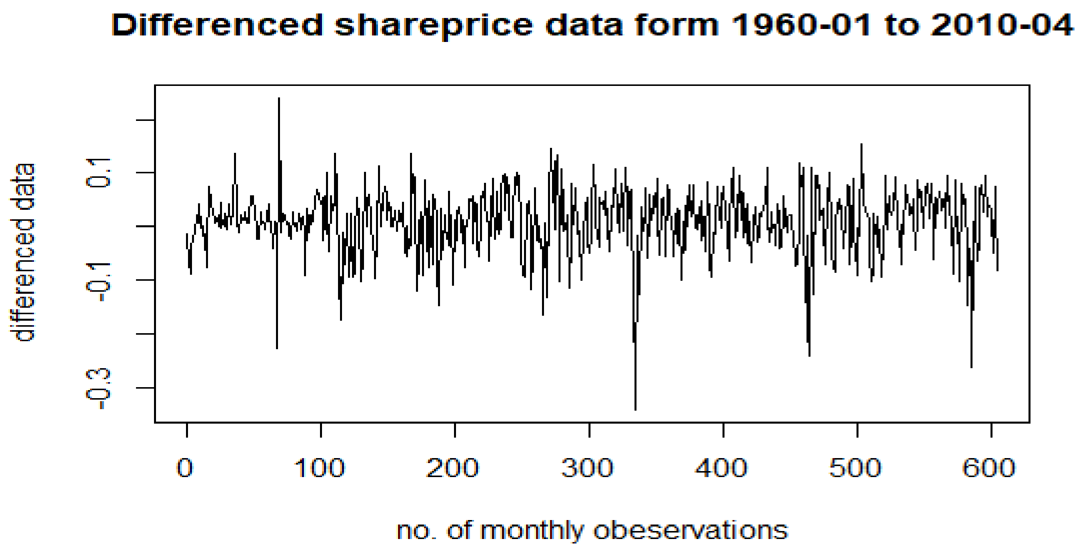


Figure 4.6 Frist Differenced transformed data

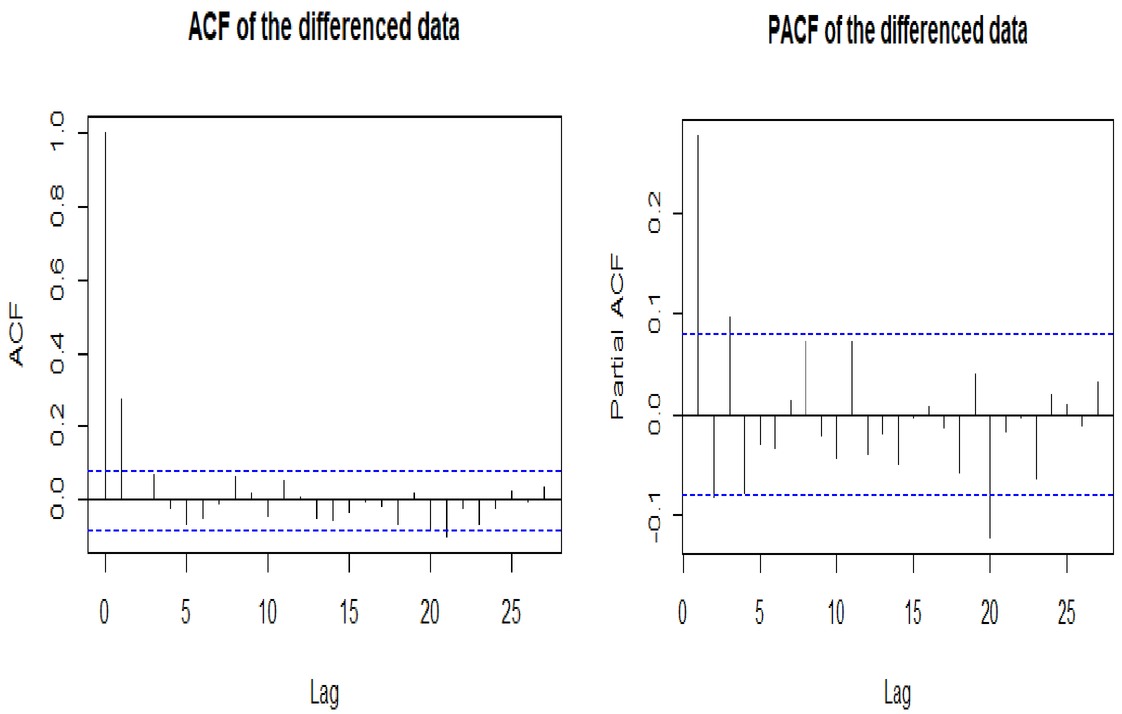


Figure 4.7 ACF and PACF of the first differenced data

Augmented Dickey Fuller (ADF) Test			Phillips-Perron Unit Root Test		
statistic	lag	p-value	statistics	lag	p-value
-10.48330	6	0.01	-21.171	6	0.01

Table 4.3 ADF and PP test on a differenced data

4.4 ARIMA/ SARIMA model

The transformed data was “differenced” once to make the data stationary (figure 4.6), with the ACF showing only one significant spike (figure 4.7) suggesting that the data is stationary and this was confirmed by the ADF and PP tests (p-values = 0.01, table 4.3). Thus, concluding that the data is stationary after the first difference since the p-value is less than 0.05. The model selection and parameters were done based on the ACF and PACF. AIC and BIC to select the best model with the lowest AIC and BIC values, and found that the ARIMA (0,0,1) and ARIMA (1,0,1) models seemed to have similar values based on the AIC (table 4.4). The models are arranged in ascending order based on the AIC. Further analysis was done on these two models to compare their ME, RMSE and MAE and it was found that the model ARIMA (0,0,1) has a better ME than ARIMA (1,0,1) model shown in table 4.5.

ARIMA model	AIC	BIC
ARIMA (1,0,1) with non-zero mean	-2214.526	-2196.019
ARIMA (0,0,1) with non-zero mean	-2214.259	-2200.378
ARIMA (1,0,2) with zero mean	-2213.798	-2190.665
ARIMA (0,0,2) with non-zero mean	-2213.541	-2195.034
ARIMA (2,0,2) with non-zero mean	-2212.665	-2184.905
ARIMA (0,0,1) (0,0,1) [12] with non-zero mean	-2212.266	-2193.759
ARIMA (0,0,1) (1,0,0) [12] with non-zero mean	-2212.266	-2193.759

Table 4.4 Selected models based on the AIC and BIC

ARIMA model	AIC	BIC	ME	RMSE	MAE
ARIMA (1,0,1)	-2214.526	-2196.019	1.075e-05	0.056	0.041
ARIMA (0,0,1)	-2214.259	-2200.378	4.0075e-0 6	0.056	0.041

Table 4.5 Accuracy comparison of the two models

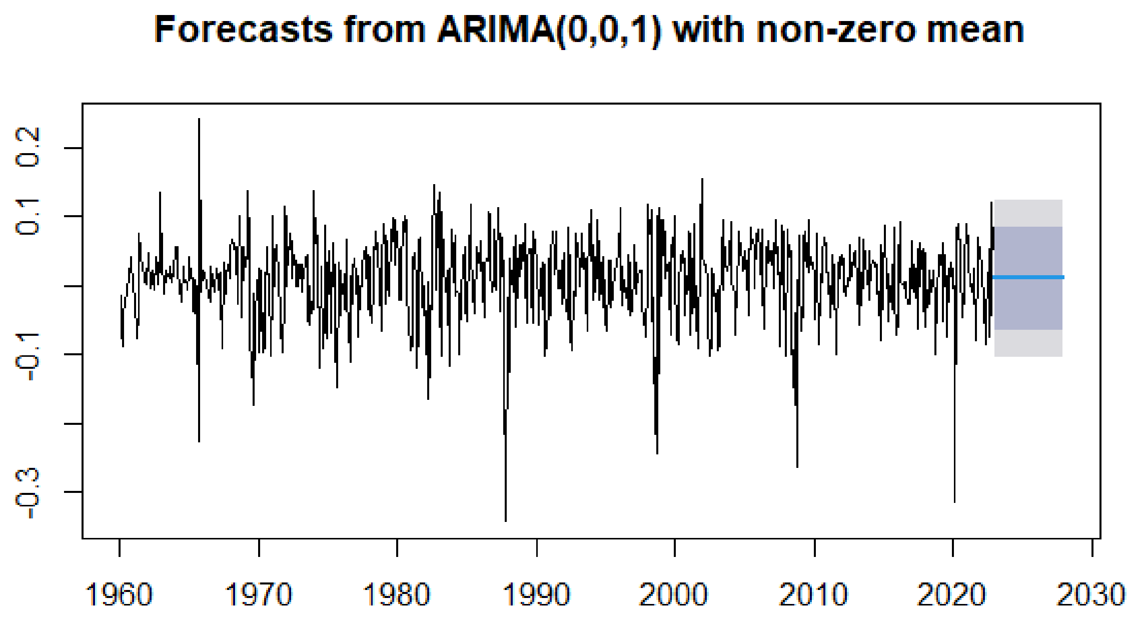


Figure 4.8 Forecast of the ARIMA model for the next 5 years

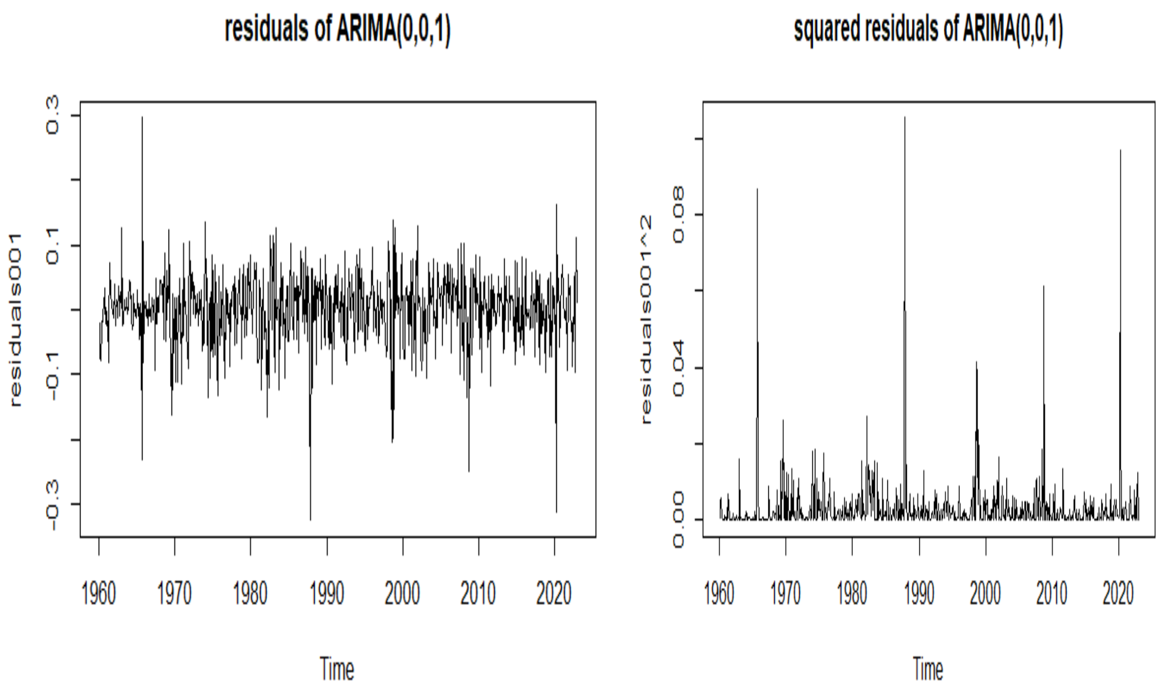


Figure 4.9 Residuals and squared residual plots of the ARIMA model

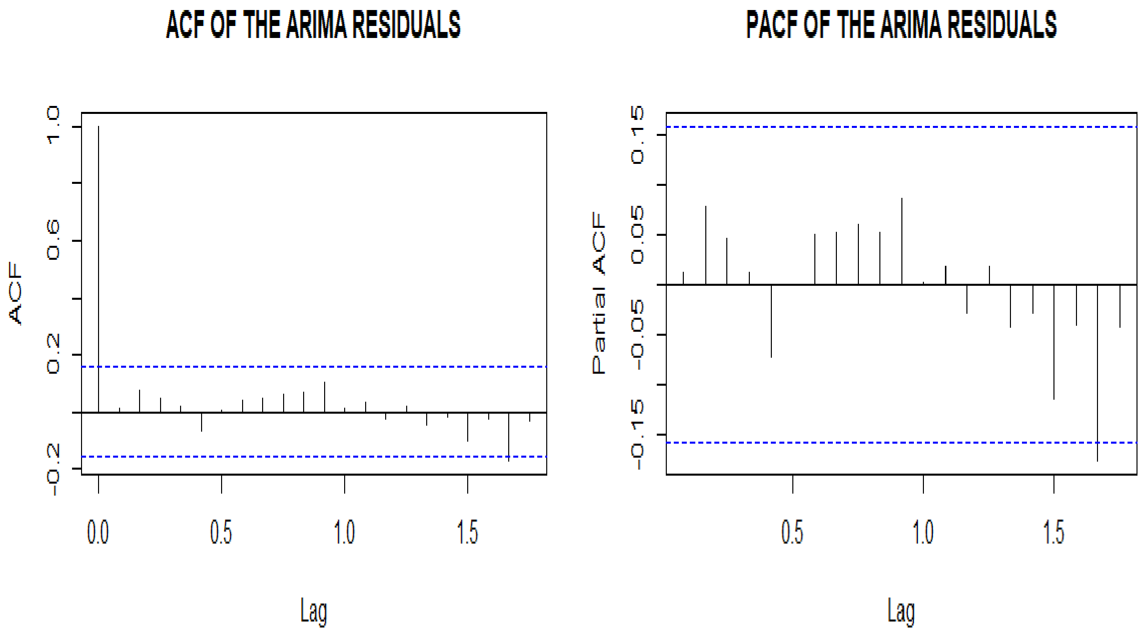


Figure 4.10 ACF and PACF of the ARIMA residuals

Lag	10	15	20	25	100
Ljung-Box Q Statistic	11.51	15.135	19.385	27.473	110.15
p-value	0.3192	0.4417	0.4969	0.3327	0.2291

Table 4.6 Ljung-Box test for ARIMA residuals

4.5 ARCH model

Figure 4.8 shows the forecast carried out using the ARIMA(0,0,1) model as the best performing model for the next 5 years at 95% confidence interval. Figure 4.9 shows the residuals and squared residual plots, which appear to be normal with the ACF and PACF graphs showing no significant spikes (figure 4.10) suggesting that the data is normally distributed, but the variance is still not constant overtime. Table 4.6 shows the Ljung-Box test to test the autocorrelation/independence of the residuals, since the p-value is greater than 0.05 at different lags, the residuals are independent.

Table 4.7 tests the ARCH effects in the residuals of the ARIMA model. Since the p-values are all less than 0.05, the presence of heteroscedasticity in the residuals is confirmed.

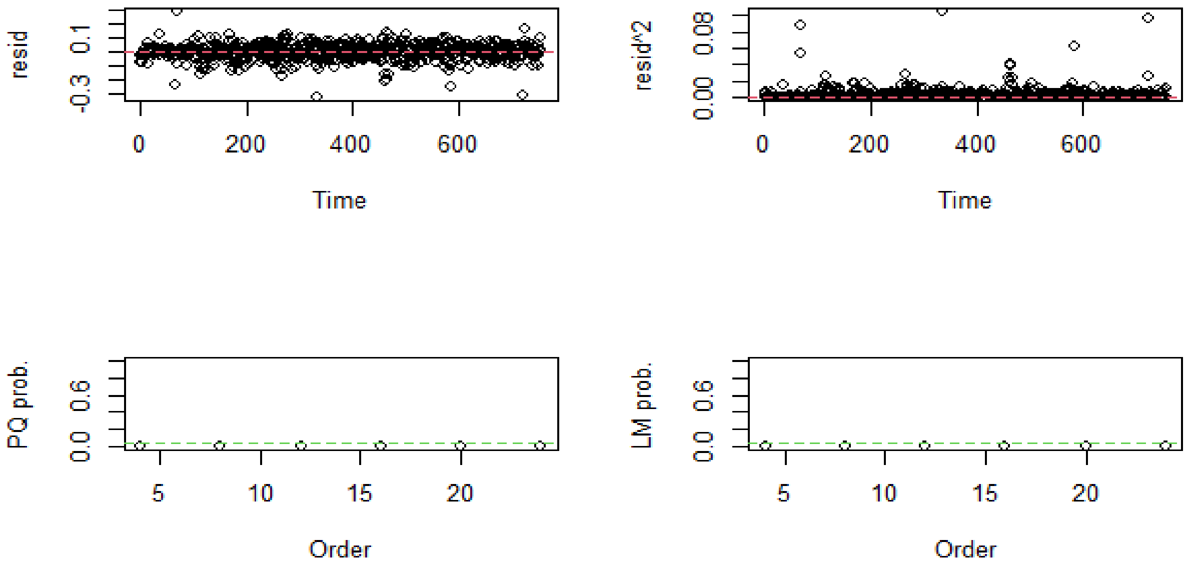


Figure 4.11 ARCH test

Portmanteau-Q test			Lagrange-Multiplier test		
order	PQ	p-value	order	LM	p-value
4	40.8	2.95e-08	4	731	0.00e+00
8	41.2	1.90e-06	8	355	0.00e+00
12	44.0	1.53e-05	12	230	0.00e+00
16	45.3	1.26e-04	16	169	0.00e+00

Table 4.7 Test for the arch effects on the residuals

	ARCH(1)	ARCH(2)	ARCH(3)	ARCH(4)
AIC	-7.132251	-7.144152	-7.143130	-7.139402
BIC	-7.113866	-7.119640	-7.112489	-7.102634
Ljung-Box test residuals	0.07073839	0.2126482	0.3651242	0.364612
Ljung-Box test squared residuals	0.9999999	0.9999999	0.9999999	0.9999999
LM ARCH test	1	1	1	1
Log Likehood	2695.425	2700.918	2701.532	2701.124

Table 4.8 ARCH model selection

The selection of the ARCH model is based on the AIC and BIC values, the ARCH(2) model has a lower AIC and BIC value than the other ARCH models with the p-value of the Ljung-Box test greater than 0.05 suggesting that the residuals of the ARCH(2) model are independent and with the ARCH-LM test, the p-value is greater than 0.05 which indicates that there is no ARCH effect in the ARCH(2) model.

The ARCH (2) model parameters are all significant and the model is given by:

Coefficients	Estimate	Standard error	p-value
mu	2.298e-03	2.113e-04	2e-16
omega	3.140e-05	2.354e-06	2e-16
alpha1	1.000e+00	1.608e-01	5.05e-10
alpha2	1.408e-01	6.140e-02	0.0219

Table 4.9 Estimates of the ARCH (2) model

$$\sigma_t^2 = 3.140\mathrm{e}-05 + \varepsilon_{t-1}^2 + 1.408\mathrm{e}-01\varepsilon_{t-2}^2$$

4.6 ARMA (0,1)-ARCH (2) model

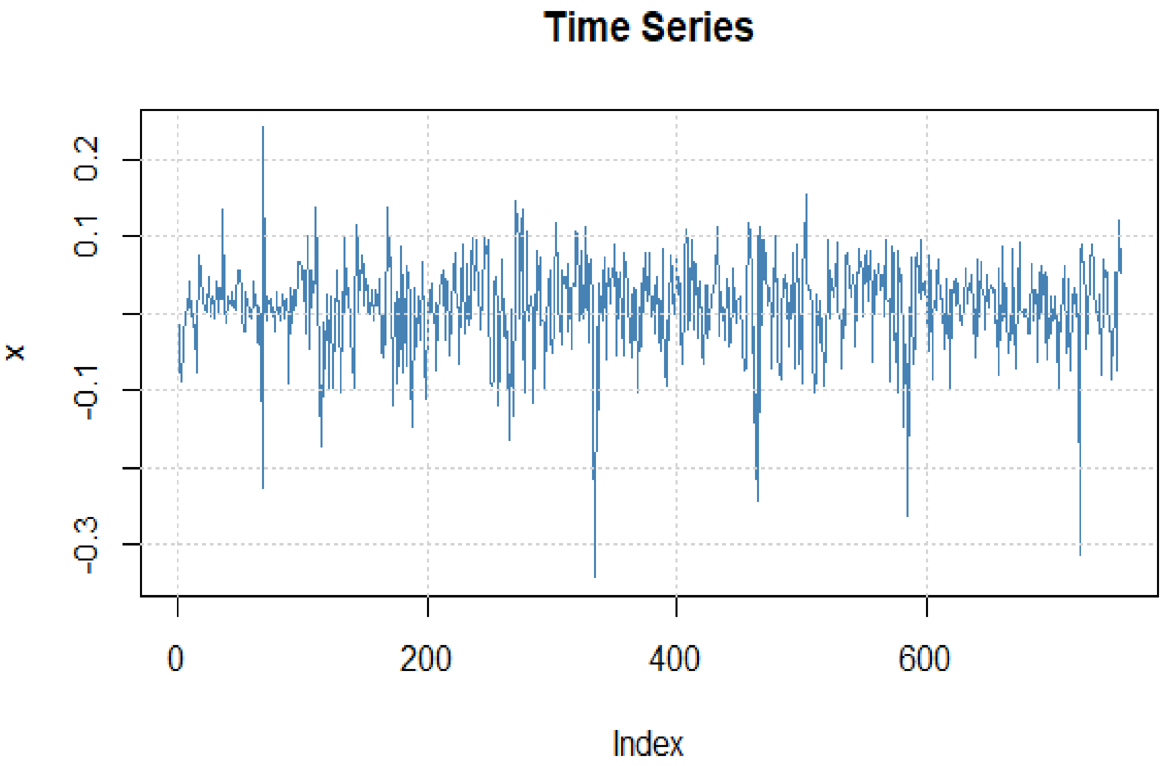


Figure 4.12 Fitted values of ARMA (0,1)-ARCH (2) model

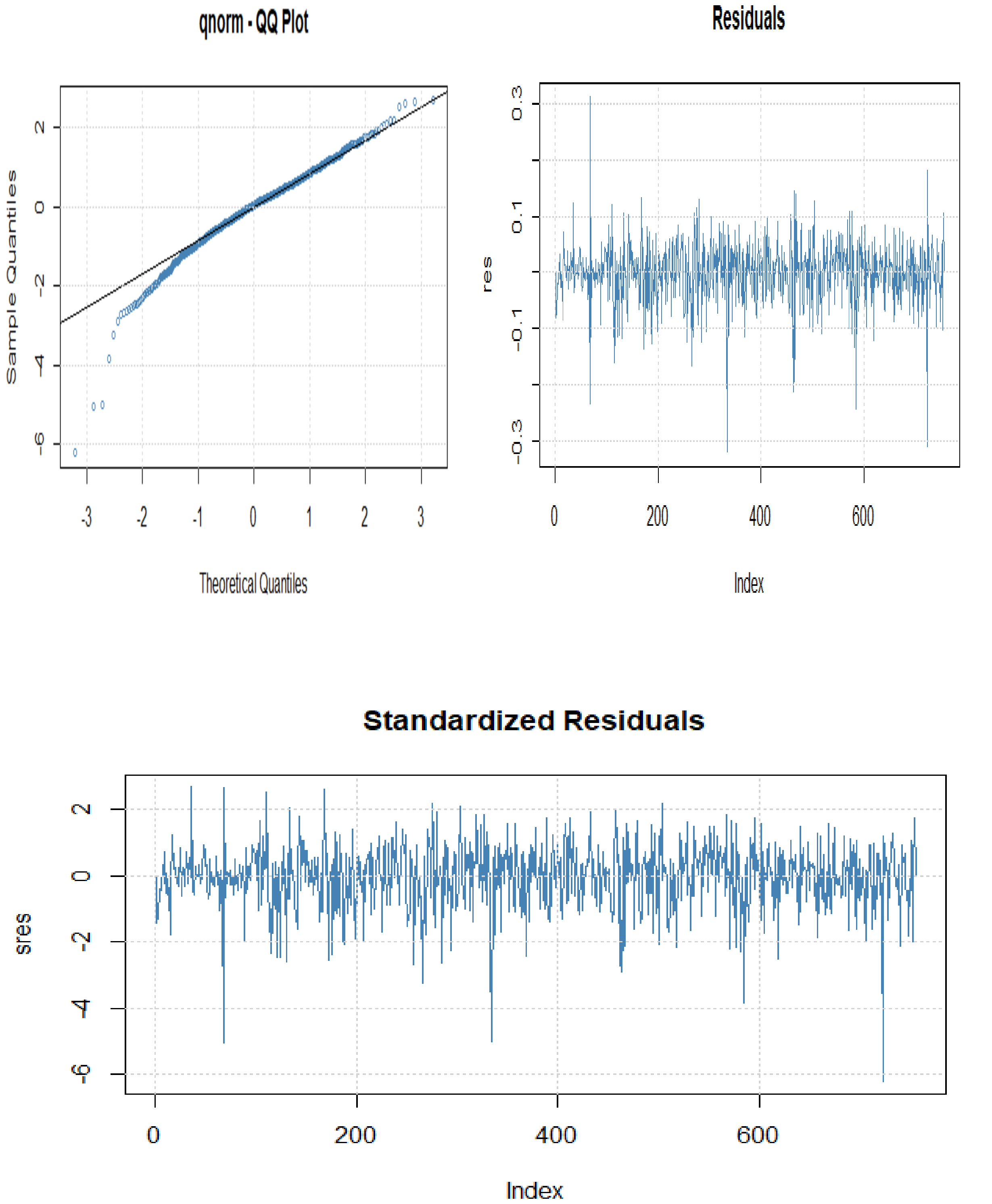
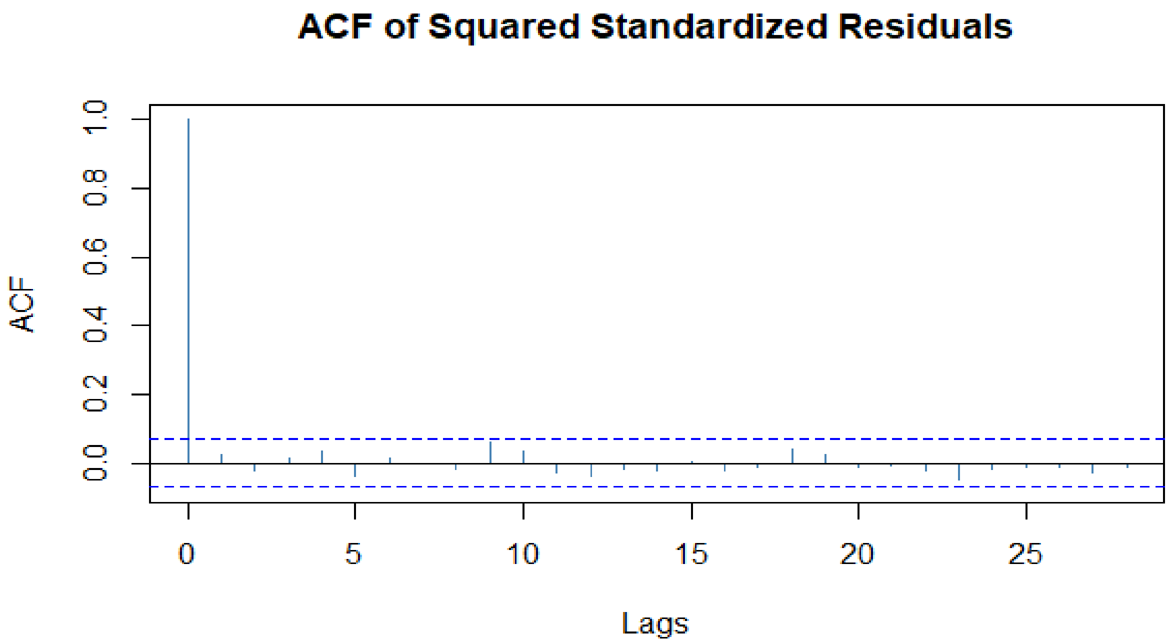


Figure 4.13 QQ-plot residuals, residuals, and standardized residuals of ARMA (0,1)-ARCH (2)



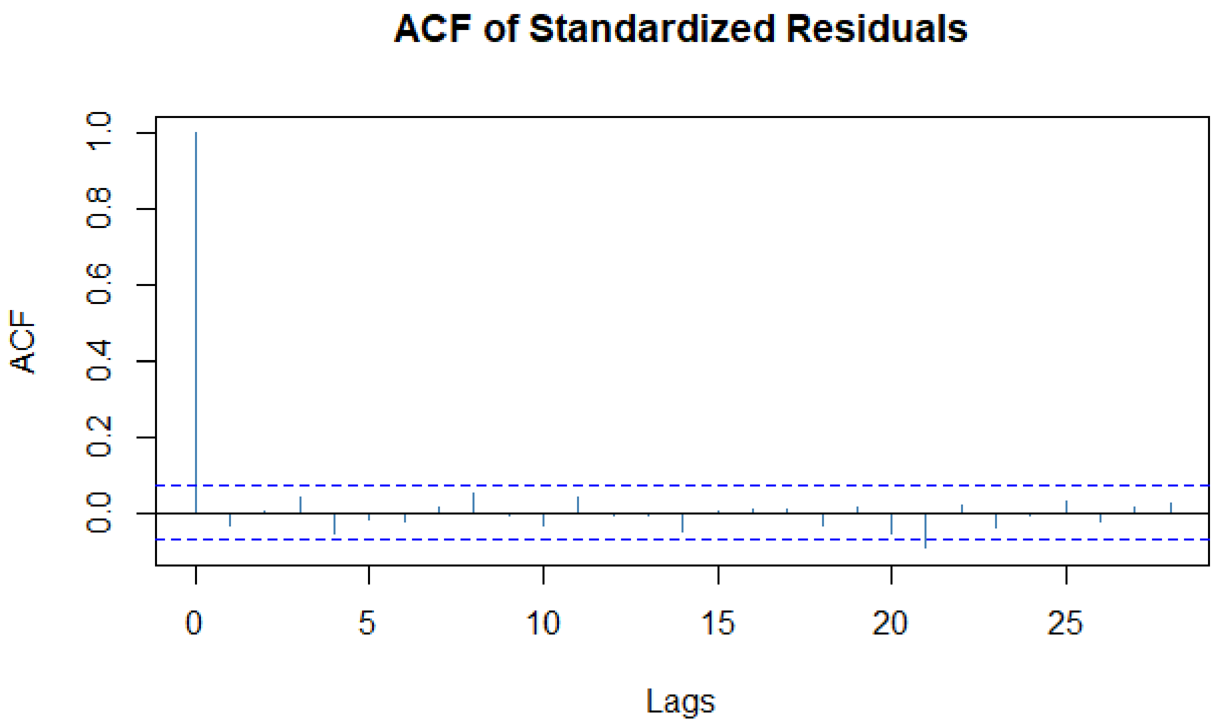
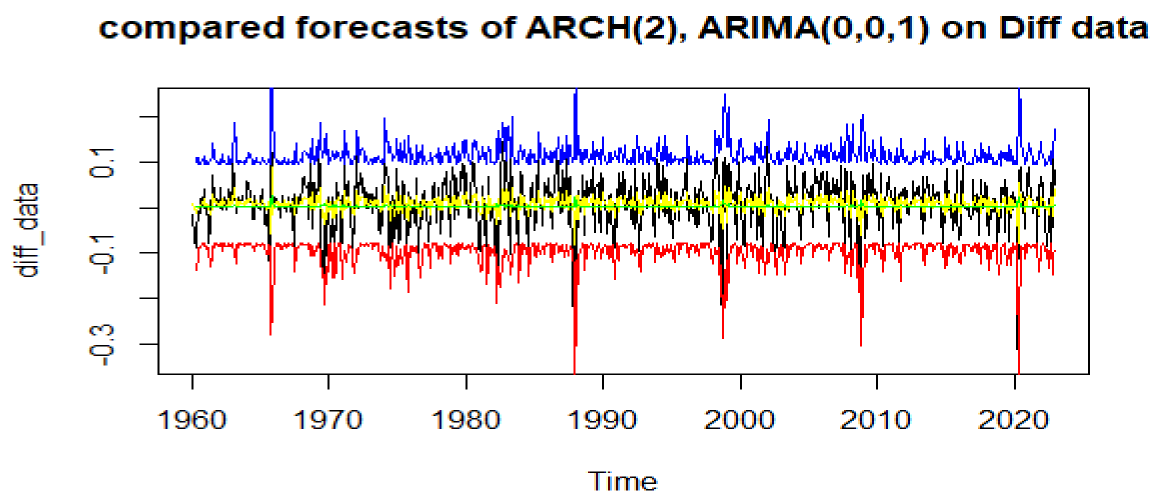


Figure 4.14 ACF and PACF of the standardized residuals of the ARMA (0,1)-ARCH (2) model.

The above figure 4.12, is the fitted values from the ARMA (0,1)-ARCH (2) model which accurately fits the differenced data showing that the model fits the data well. Figure 4.13 shows that the residuals of the ARMA (0,1)-ARCH (2) model appear to be normal. The ACF of the standardized residuals and squared residuals on figure 4.14 of the ARMA (0,1)-ARCH (2) model shows no significant spikes, hence the model fits the data well.



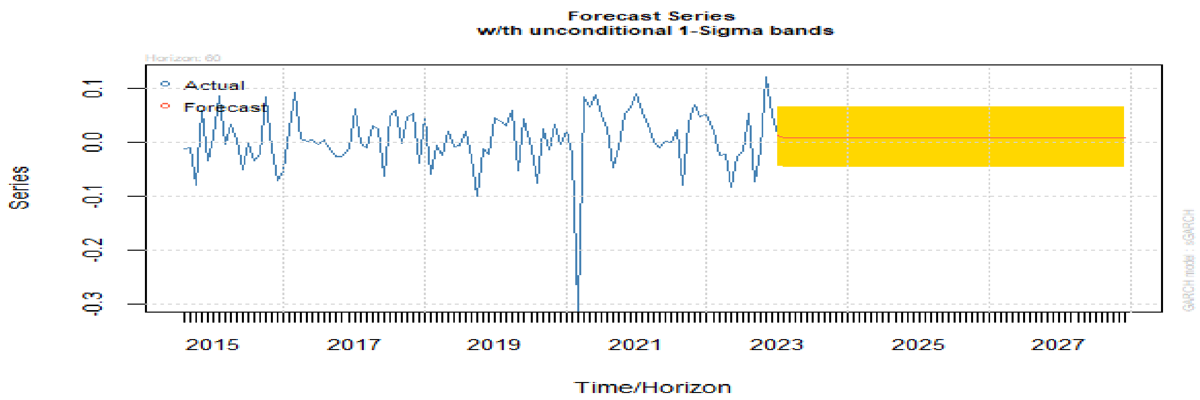


Figure 4.15 Forecasted values of the ARCH (2), ARIMA (0,0,1) models and forecast of ARMA (0,1)-ARCH (2)

4.7 Diagnostic check

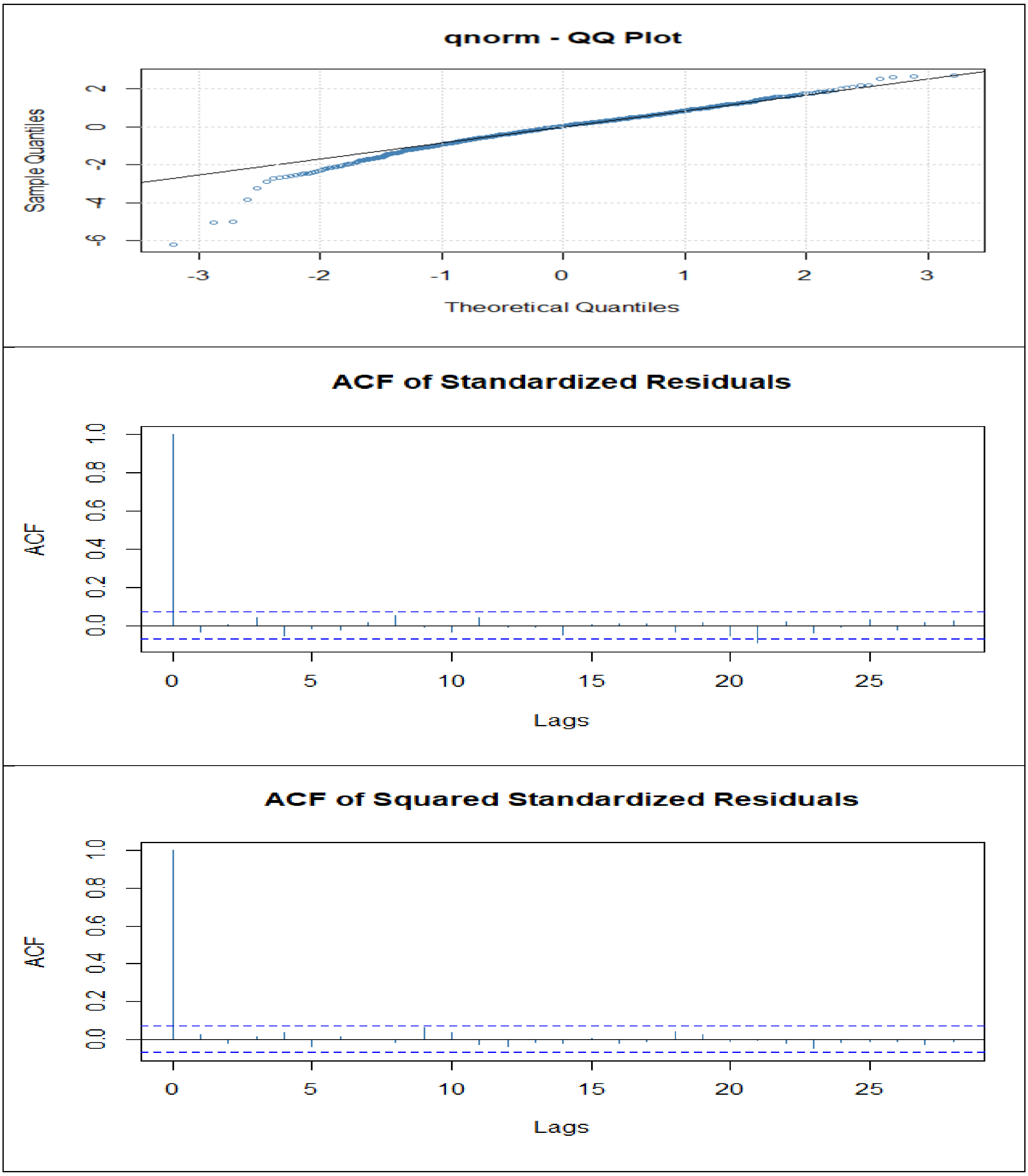


Figure 4.16 QQ-plot, ACF of the standardized and squared residuals

The standardized residuals of the ARMA (0,1)-ARCH (2) shows no significant spikes and the QQ-plot shows that the standardized residuals are approximately normal as more points lie on the line on figure 4.16. The forecast of the ARIMA (0,0,1) fit the data well and better than the ARCH (2), and the forecast of the ARMA (0,1)-ARCH (2) model, indicate that the share price will decay for one month and remain constant for the next 5 years.

CHAPTER FIVE

DISCUSSION, CONCLUSION AND RECOMMENDATIONS

5.1 Introduction

This chapter discusses the main findings of the study with respect to the research objectives and hypotheses. Conclusion and recommendations are also presented in this chapter.

5.2 Discussion

Stock market forecasting is a challenging study which depends on the characteristics of the data, and using the time series models like the SARIMA model offers a wide range of how to handle the characteristics of the data such as seasonal and non-seasonal components. Including the ARCH model to the SARIMA model, offers more efficient ways to handle the data characteristics including volatility changing i.e., non-constant variance, as the SARIMA model assumes constant variance over time, but the stock market mostly exhibits non-constant variance as the graph of figure 4.2 indicates, showing an increasing trend with exponential growth with the ACF and PACF tests confirming it to be non-stationary.

As the data shows less seasonality this also point that according to the first null hypothesis that seasonality does not affect stock market forecasting is accepted as the models in table 4.4 suggest that there exist less seasonal effects from the JSE share price data with the seasonal models appearing with larger AIC and BIC, than the non-seasonal models. After the transformation and differencing are done, the data follows all the assumptions for the models to be applied as shown in figure 4.6, where the data is stationary around the mean of zero and have a constant variance, with the ACF and PACF having no spikes outside the thresh hold meaning no significant spikes but showing some volatility on the residuals of the ARIMA model in figure 4.9, which requires the ARCH model to be incorporated into the model as the residuals of the ARIMA model in figure 4.9 shows some volatilities. However, it has no influence in the share prices as the standardized residuals of the ARMA(0,1)-ARCH(2) model in figure 4.13 are the same as that of the SARIMA(0,0,1) model in figure 4.9, although the ARCH-LM test shows that there exist volatility/ ARCH effects in table 4.7, since all the p-values are less than 0.05, but according to the analysis the volatility, this is really not significant. Thus, the ARCH model has no influence over the SARIMA model.

Also, the analysis shows that the data gives the models less information to interpret the data and that simply means that the simple models like moving average models can be effective to make forecast from the data that as the data shows few characteristics that can be handled by more complex models based on figure 4.8. The forecast of the ARIMA(0,0,1) model with the forecast of the ARMA(0,1)-ARCH(2) model in figure 4.15 are very similar therefore, the ARMA(0,1)-ARCH(2) model is not a better model than the ARIMA(0,0,1) model.

The comparison of the ARCH(2) model and the ARIMA(0,0,1) model in forecasting the differenced data in figure 4.15, shows that the models fits the data very well and the combination of both models shows some improvement as the ARCH model nearly produces a straight line in predicting the differenced share price data than the ARIMA(0,0,1) model separately since it captures some changing movements such as volatility as there exist less volatility from the data.

5.3 Conclusion

The study was conducted to explore the use of SARIMA and ARCH models to make forecast on the share price data from the Johannesburg Stock Exchange (JSE) from 1960-2022, with the purpose to develop a suitable model to make future predictions by accounting for volatility in the JSE. The SARIMA model is a model with linearity, stationarity, white noise, and constant variance assumptions, if violated transformations and differencing are performed to get the best model to handle the data. The ARCH model is only done if the residuals are white noise and contains some volatility clustering, meaning the model could suit the data well but the volatility is always present in the financial stock markets data which can be handled by the ARCH models. The Box-Cox transformation was done and then the data differenced to remove trends. Following the application of Box-Jenkins method, the SARIMA(0,0,1) model was found to be the best model with the lowest AIC values, using the conditional heteroscedasticity approach by (Engle, 1982), the ARCH(2) model was found from the residuals of the SARIMA model to be the best model with all the coefficients having a p-value less than 0.05, showing significance, Then combined with the SARIMA model to make forecast, forming the SARIMA-ARCH model. According to the study of Kemalbay, G., & Korkmazoglu, O. (2021), used the SARIMA-ARCH model and found that the SARIMA-ARCH model performed better than the SARIMA model which is the same as in this study as it used the SARINA/ARIMA-ARCH model but differ from the results as this study shows less difference between ARIMA/SARIMA-ARCH model with ARIMA model where both shows the same forecast of a straight line. Another study by Brida, J and Risso W. (2011) used the SARIMA-ARCH model and the conclusion was the SARIMA-ARCH model forecast best which is the similar to the study by Kemalbay, G., & Korkmazoglu, O. (2021). And different to this study as the data shows less seasonality and less volatility that the preveous studies

Further tests were performed from the ARIMA/SARIMA-ARCH model to test for ARCH effect and autocorrelation of the residuals, the results shown that the is no ARCH effect on the ARIMA/SARIMA model and no autocorrelation from the residuals. The forecast was done using the ARIMA (0,0,1)-ARCH (2) model and the forecast was very similar to that of ARIMA (0,0,1) model as the data only gave unstable trend from 2010 to 2022 and not from 1960. The AIC and BIC were used to select the best SARIMA/ARIMA and ARCH model, the ME, RMSE, and the MAE were also used to compare the best two SARIMA/ARIMA models for model evaluation regarding the data. The study further adds knowledge to analyzing the Johannesburg Stock Exchange data using time series data.

5.4 Recommendations

The study focused on ARIMA/SARIMA-ARCH model in analyzing time series data form JSE. It is recommended that other methods can also be applied to analyze future stock market prices for investors to feel confident and have knowledge to invest in the stock market. It is also advisable that the data should be reduced to a manageable period that incorporates more changes from 2000 to 2020. It would also be beneficial to split the data to create a sample to train the model and make future forecast from the same data using the testing data, and perhaps, advisable to use 70/80% training and 30/20% for testing to select the best model. Incorporating the ARCH model from this research was to understand the use of the ARCH models, The study also recommends that other ARCH models like GARCH, TGARCH, TACH, and other hybrid ARCH models like NAGARCH models could also be applied to the financial time series data to explore where more volatility is likely to happen to give more meaningful interpretation to the data.

Abbreviations

ARMA Autoregressive Moving Average

ARIMA Autoregressive Integrated Moving Average

SARIMA Seasonal Autoregressive Integrated Moving Average

GARCH Generalized Autoregressive Conditional Heteroscedasticity

TGARCH Threshold Generalized Autoregressive Conditional Heteroscedasticity

TACH Threshold Autoregressive Conditional Heteroscedasticity

NAGARCH Nonlinear Asymmetric Generalized Autoregressive Conditional Heteroscedasticity

AIC Akaike Information Criterion

BIC Bayesian Information Criterion

Q-Q plot Quantile-Quantile plot

ARCH Autoregressive Conditional Heteroscedasticity

LM Langrage Multiplier

ACF Autocorrelation Function

PACF Patial Autocorrelation Function

ADF Augmented Dickey-fuller test

PP Phillips-Perroni

QP Portmanteau-Q test

ME Mean Error

RMSE Root Mean Square Error

MAE Mean Absolute Erro

References

1. Alharbi, F., & Csala, D. (2022). ASeasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX) Forecasting Model-Based Time Series Approach. 21.

2. Ahmed, F., Umer, S., & Ali, R. (2017). Modelling Stock Indexes Volatility of Emerging Markets. *Management Sciences*, 12.

3. Aminu, M., Abdul-Nasir, T., Nadira, I., & Usman, S. (2021). Forecasting Nigeria's Exchange Rate Using SARIMA Modeling. *Social and Economic Research*, 12.

4. Bakar, N., & Rosbi, S. (2017). Autoregressive Integrated Moving Average (ARIMA) Model for Forecasting Cryptocurrency Exchange Rate in High Volatility Environment: A New Insight of Bitcoin Transaction . *Advanced Engineering Research and Science (IJAERS)* , 8.

5. Boudrioua, M., & Boudrioua, A. (2020). Modeling and Forecasting the Algerian Stock Exchange Using the Box-Jenkins Methodology. *Economics, Finance and Accounting Studies (JEFAS)*, 15.

6. Box, G., Jenkins, G., Reinsel, G., & Ljung, G. (1976). *Time series analysis, control, and forecasting.* Canada: John Wiley & Sons.

7. Brida, J.G. and Risso, W.A., 2011. A SARIMA-ARCH model for the overnight-stays tourism in South Tyrol. Tourism Economics, 17(1), pp.209-222.

8. Dikko, H., Osebi, A., & Agboola, S. (2015). Modelling Abrupt Shift in Time Series Using Indicator Variable: Evidence of Nigerian Insurance Stock. *Finance and Accounting*, 12.

9. Dinardi, F. (2020). Forecasting the Stock Market Using ARIMA and ARCH/GARCH Approaches . 50.

10. Engle, R. (1982). Autoregressive Conditional Heteroscedasticity With Estimates Of The Variance Of United Kingdom Inflation. *econometric society*, 22.

11. Fan, C., Xiao, F., & Wang, S. (2914). Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. *Applied Energy*, 10.

12. Freddie, P. (2021). Artificial Neural Networks To Predict Share Prices On The Johannesburg Stock Exchange. *Computer Science*, 99.

13. Ganbold, B., Akram, I., & Fahrozi , R. (2017). Exchange rate volatility: A forecasting approach of using the ARCH family along with ARIMA SARIMA and semi-structural-SVAR in Turkey. *ECONOMICS*, 37.

14. Gideon, B., & Paul, A. (2014). Co-movement of Africa's equity markets: Regional and global analysis in the frequency–time domains. *Physica A*, 22.

15. Kaseke, F. (2015). Modelling Volatility in Stock Exchange Data: A Case Study Of Three Johannesburg Stock Exchange(JSE) Companies. 138.

16. Kemalbay, G., & Korkmazoglu, O. (2021). Sarima-arch versus genetic programming in stock price prediction. *Engineering and Natural Sciences – Sigma Mühendislik ve Fen Bilimleri Dergisi*, 13.

17. Makatjane, K., & Moroke, N. (2021). Predicting Extreme Daily Regime Shifts in Financial Time Series Exchange/Johannesburg Stock Exchange—All Share Index. *Financial Studies*, 18.

18. Masimba, F., & Shadreck, M. (2021). Comparison of Arch Family Models Using E-Views Computer Software Packages: A Case of Zimbabwe Stock Exchange . *Multidisciplinary Research*, 10.

19. Nashirah, A., & Sofian, R. (2017). Autoregressive Integrated Moving Average (ARIMA) Model for Forecasting Cryptocurrency Exchange Rate in High Volatility Environment: A New Insight of Bitcoin Transaction . *Advanced Engineering Research and Science (IJAERS)* , 8.

20. Nwokike, C., & Okereke, E. (2021). Comparison of the Performance of the SANN, SARIMA and ARIMA Models for Forecasting Quarterly GDP of Nigeria. *Mathematics* , 20.

21. Pillay, S. (2020). Determining The Optimal Arima Model For Forecasting The Share Price Index Of The Johannesburg Stock Exchange. *Management Information and Decision Sciences* , 11.

22. Ruipeng, L., & Rangan , G. (2022). Investors' Uncertainty and Forecasting Stock Market Volatility. *Behavioral Finance*, 12.

23. Sepato , L. (2017). A comparative analysis of ARCH and GARCH type models before and after the removal of outliers . 119.

24. Shaik, M., & Sejpal, A. (2020). The Comparison of GARCH and ANN Model for Forecasting Volatility: Evidence based on Indian Stock Markets: Predicting Volatility using GARCH and ANN Models. *Prediction Markets*, 18.

25. Shannon, G., & Samouilhan, N. (2008). Forecasting volatility on the JSE. *Investment Analysts*, 11.

26. Shruti, C. (2019). Stock Price Prediction Using Time Series Models . 43.

27. Surendran, P. (2020). Determining The Optimal Arima Model For Forecasting The Share Price Index Of The Johannesburg Stock Exchange. *Management Information and Decision Sciences*, 11.

28. Virginia, E., Ginting, J., & Elfaki, F. (2018). Application of GARCH model to forecast data and volatility of share price of energy(Study on Adaro Energy TBK, LQ45). *Energy, Economics and policy*, 9.

29. Wei, M., Kee, O., & Musa, S. (2023). Seasonal versus non-seasonal trends in stock market Malaysia. *Mathematics, Actuarial and Quantitative Studies*, 19.

30. Zahangir, A., Siddikee, N., & Masukujjaman, M. (2013). Forecasting Volatility of Stock Indices with ARCH Model. *Financial Research* , 18.

University of Fort Hare
Together in Excellence

ETHICS CLEARANCE

REC-270710-028-RA Level 01

Project Number:	202334512-BN-JN
Project title:	Stock market forecasting using time series analysis with ARIMA/SARIMA-ARCH model: A case study of the Johannesburg Stock Exchange.
Qualification:	BSc (Hons) Applied Statistics
Student Name:	Ngcethane B
Student No.:	202334512
Supervisor:	Dr J Ndege
Dept, Faculty/ Centre:	Department of Statistics Faculty of Science and Agriculture
Co-supervisor:	N/A

On behalf of the University of Fort Hare's Research Ethics Committee (UREC), I hereby grant ethics approval for 202334512-BN-JN. This approval is valid for 12 months from the date of approval. Renewal of approval must be applied for BEFORE termination of this approval period. Renewal is subject to receipt of a satisfactory progress report. The approval covers the undertakings contained in the above-mentioned project and research instrument(s). The research may commence as from the 11 October 2023, using the reference number indicated above.

Note that should any other instruments be required, or amendments become necessary, these require separate authorization.

Please note that UREC must be informed immediately of

- Any material changes in the conditions or undertakings mentioned in the document.
- Any material breaches of ethical undertakings or events that impact upon the ethical conduct of the research.

The Principal Researcher/Student must report to UREC in the prescribed format, where applicable, annually, and at the end of the project, in respect of ethical compliance.

UREC retains the right to

- Withdraw or amend this approval if
    - Any unethical principal or practices are revealed or suspected.
    - Relevant information has been withheld or misrepresented.
    - Regulatory changes of whatsoever nature so require.
    - The conditions contained in the Certificate have not been adhered to.

- Request access to any information or data at any time during the course or after completion of the project.

Your compliance with Department of Health 2015 guidelines and any other applicable regulatory instruments and with UREC ethics requirements as contained in UREC policies and standard operating procedures, is implied.

UREC wishes you well in your research.

Yours sincerely

Professor Michael Aliber
Representative: Faculty of Science and Agriculture Research Ethics Committee
11 October 2023