A
REPORT
On
STUDY OF STOCK MARKET VOLATILITY USING TIME SERIES

Submitted to

MIT ADT UNIVERSITY
SCHOOL OF ENGINEERING & SCIENCE
DEPARTMENT OF APPLIED SCIENCE & HUMANITIES

In partial fulfilment of the requirements for the award of the degree of

MASTER OF SCIENCE
IN
APPLIED STATISTICS (DATA SCIENCE)

Submitted by

Mohan L. Bodake	**MITU21MSDS0006**
Namrata S. Thengal	**MITU21MSDS0007**

Under the Guidance of

Dr. Ashok Kumar
Prof. Sunil Thakre

June, 2023

SCHOOL OF ENGINEERING & SCIENCE
DEPARTMENT OF APPLIED SCIENCE & HUMANITIES
RAJBAUG, LONI KALBHOR, PUNE-412201

CERTIFICATE

This is to certify that the Capstone Major Project-22MSST422 entitled

Study of Stock Market Volatility Using Time Series

Submitted by

Mohan L. Bodake	**MITU21MSDS0006**
Namrata S. Thengal	**MITU21MSDS0007**

is a bonafide work carried out by them, under the supervision of guide and co-guide, it is submitted towards the partial fulfilment of the requirement of MIT Art, Design and Technology University, Pune for the award of the Master of Science in Applied Statistics (Data Science).

Dr. Ashok Kumar
Guide

Prof. Sunil Thakre
Co-guide

Dr. Pratibha Jadhav
Program Co-ordinator

Prof. Dr. Haribhau Bhapkar
HOD,
Applied Sciences & Humanities

Prof. Dr. Virendra Shete,
Director,
MIT School of Engineering & Science

External Examiner

Place: Pune
Date:

DECLARATION

We hereby declare that the Project entitled "**Study of Stock Market Volatility Using Time Series"** submitted towards the partial fulfilment of the requirement of MIT-ADT University, Pune for the award the Master of Science in Applied Statistics (Data Science) of the is a record of bonafide work carried out by us under the supervision of Dr. Ashok Kumar and Prof. Sunil Thakre, Department of Applied Science & Humanities, MIT School of Engineering & Science, Pune.

We further declare that the work reported in this report has been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

Place: Pune

Student Names and Sign

Mohan L. Bodake

Namrata S. Thengal

ACKNOWLEDGEMENT

We would like to thank **Prof. Dr. VIRENDRA SHETE sir**, Director, MIT School of Engineering and Science, & **Prof. Dr. HARIBHAU BHAPKAR Sir**, Head of Department of Applied Science & Humanities, **MIT ADT UNIVERSITY PUNE**, for providing me the necessary facilities. We would like to express our profound gratitude towards our guide **Dr. ASHOK KUMAR** for their support and suggestions. We would like to thank **Dr. PRATIBHA JADHAV**, **Dr. HARSH TRIPATHI**, and **Prof. M.S. KASTURE** for their timely suggestions and encouragement.

We would like to thank the non-teaching staff and all my colleagues for their support and valuable suggestions.

Also, we would like to thank the Almighty god who gave us faith and patience to complete this project.

Mohan L. Bodake

Namrata S. Thengal

ABSTRACT

Stock market forecasting is an important topic in academic studies, econometrics, and financial studies.it involves depth study on time series. The analysis of the top 10 NIFTY-50 performing companies on the National Stock Exchange is the main goal of this study. Many forecasting techniques can be implemented to this data to predict how future stock returns will behave.

ARIMA this models are mostly used in time series analysis and generally produce good results; however, different approaches must be studied in order to decide which one best matches the data. When taking into account the high data volatility and the economic situation of the country under study, other methods, especially the ARCH models, must be taken into consideration.

Certain techniques are used to predict stock market data globally. The top 10 Nifty 50 index companies, ranked by market capitalization, are the focus of this study. We can find out concerning overall market trends and the performance of important individuals on the Indian stock market through studying the stock price data of these companies. Time series models can be utilized for forecasting future stock prices, which is beneficial for analysts, investors, and other market participants. Accurate forecasts help investors make the best potential decisions regarding investments, and analysts can utilize these insights to give their customers helpful guidance. Also, the study offers a practical application and improves the body of knowledge in terms of finance. of time series analytic methods for predicting stock prices.

Table of Contents

1. INTRODUCTION

Several fields, such as economics, engineering, agriculture, and others, have developed time series analysis. However, the financial time series study has captured the attention of many academics in the field in recent years. If one analyses the present global situation, which included a serious recession in major markets like the US and Europe, it is simple to see the causes of this occurrences.

Classical Techniques for predicting have been used to forecast stock market since it is important for understanding and predicting future conditions. This method, called ARIMA, usually produces beneficial results in other domains. They must be carefully studied, however, because they usually not the most effective models to analyse the past and predicting the future in terms of financial time series. Financial sectors can be considered a particular case in the field of forecasting because it presents volatility, which arises since there is no pattern in the financial markets and show significant volatility throughout time due to a variety of factors. This is the reason for the invention of different time series methods, such as the ARCH model family. These models also known as heteroscedastic models. Heteroscedasticity is a statistical phenomenon that arises when data show different variances over time; it effectively explains the behaviour of financial time series. But the reality is that there is volatility The linear models will also be used in the financial sector. be put to the test because certain companies may have more stable time series, and heteroskedastic models may not perform improved for those. The best model will be selected and forecasts are going to be produced after throughout analysis.

In this project, we will be focusing on estimation and prediction of the stock prices. We will be concentrating on stock price forecasting in this project. The trend line is an important part of stock price analysis since it allows us to figure out if stocks' prices are generally rising or falling over time. We may start analysing the broad market patterns and the likely future trajectory of the stock prices by finding the trend line. We may begin fitting an appropriate time series model to the data after we have a good understanding of the trend line. Time series models are used to analyse data patterns that vary with time. We can spot any recurring patterns or trends that may exist by applying a time series model to the stock price data. This can help us to decide on possible investments with greater knowledge. We may analyse the data using lots of different time series models. The ARIMA model, the ARCH model, and the GARCH model are a few of the most widely used models. All of these models have benefits as well as of drawbacks, so we must carefully assess which model would work best with our data. Future values can be forecasted once our time series model has been fitted to the data. Understanding the expected future trajectory of the stock prices requires forecasting, which is an essential first step. We can predict future market trends using the information we collected and the insight we have gained from our analysis. Stock prices can be affected by a variety of factors, such as political developments, company performance, and economic conditions. We can start to grasp how these factors are affecting the market and improve our ability to make investing decisions by examining the data on the top 10 Nifty 50 companies by market capitalization.

Overall, this project could offer insightful information on the Indian stock market. We can identify some of the most important players in the company's sector by studying the top 10 Nifty 50 companies by market capitalization. We can decide on future investments by identifying the trend line, fitting a suitable time series model, and predicting future values. We are interested in analysing the data to determine what insights it contains.

Forecasting the stock market is a popular and significant topic for academic and business studies since there are numerous elements that could influence how a series behaves over time. High volatility, which can be caused by a variety of variables including economic conditions, political climate, financial crises, war, etc., is one of this type of time series' key characteristics. Due to these factors, it is difficult to forecast data for financial markets, and even while different techniques may be used, not all of them will produce the most favourable outcomes.

Usually, less volatile data has been chosen for ARIMA models. In order to handle time-correlated modelling and forecasting, this approach, which was initially proposed by Box and Jenkins, known as autoregressive integrated moving average (ARIMA) models. This approach is able to accurately describe occurrences and provide accurate forecasts in a number of areas. However, various methods must be tried for this study in order to compare outcomes and for the previously described reasons. Although other models generally predict the financial markets, ARIMA may have some validity. This class of models might produce useful results, especially if the data has significant volatility.

Forecasting the volatility of future stock prices is a challenge in financial time series analysis. To handle these issues, the ARCH/GARCH volatility models have been developed. The Engle (1982) and Bollerslev (1986) ARCH and

GARCH models are capable of capture volatility simultaneously. They are therefore the most recommended for anticipating certain kinds of time series.

Stationarity is concept for modelling time series data. A stationary time series' mean and variance must not change over time. In general, If a stochastic process' mean, variance, and covariance between two time periods are all constant throughout time, it is said to be stationary. is determined only by the lag between the two time periods, instead of the actual time at which the covariance is computed. A time series is known to be constant If there isn't a consistent variation in the mean over time, no consistent variation in in the Period variations and variance have been predicted said to be stationary.

2. LITERATURE REVIEW

2.1 Stock price prediction using the ARIMA model

Ayodele Ariyo Adebiyi (2014) explores the application of the autoregressive integrated moving average model in predicting stock prices. The goal of the study is to improve stock price forecasting techniques. For this, literature is showing a lot of interest in a time series model called the ARIMA. The article offers instructions for constructing an ARIMA model. The New York Stock Exchange (NYSE) offered Nigeria Stock Exchange publicly available stock data that the authors utilized to create their forecasting model. The outcomes show that the ARIMA model is capable of competing with current approaches and has a lot of potential for predicting short-term stock prices. Which include the BIC, the standard error of regression, the, autocorrelation, partial autocorrelation and adjusted R2 functions, the researchers analyse several ARIMA models. In this paper two examples are given Nokia stock index and Zenith bank stock index. The authors involve visualizations for each index that include ACF and PACF plot and also ADF test for the selection the ARIMA model

2.2 Risk Model Validation: An Intraday VaR and ES Approach Using the Multiplicative Component GARCH

Ravi Summinga-Sonagadu (2019) gives the importance of properly analyzing risk in finance and the need to predict volatility in the market. It defines the way volatility examines market ability and is important for calculating risk. Expected Shortfall and Value-at-Risk. High-frequency data is used in forecasting. This risk metrics including the data that it gives more accurate forecasts of volatility. The role of the Component that multiplies Forecasting intraday VaR and ES using the generalized autoregressive heteroscedasticity model. for this they are using the MC-GARCH and GARCH model for the intraday forecasting the stock prices.

2.3 An Introductory Study on Time series Modelling and Forecasting

Ratnadip Adhikari (2013) provides of various time series forecasting models and their features. This paper covers the four models first is Box-Jenkins or ARIMA models, second Non-linear stochastic models third Neural network forecasting models and last is Support Vector Machine. This paper also gives the importance of selecting the appropriate model order. The research also covers ways to evaluate forecast accuracy. It deals with the chance of improving prediction accuracy with replace data collection processes when significant variations between the actual value and the predicted values.

3.OBJECTIVES

The object is to study the stock market with the help of time series modelling. Forecasting future stock value helps investors in making the best possible investment decisions, and analysts can use these insights to provide helpful advice to their customers. Also, the Research adds to the database of knowledge in the subject. of finance and provides a practical application of time series analytic methods for predicting stock prices. we aim to determine the best time series model which is appropriate in financial studies. By addressing following objectives, we aim to get desired outcomes.

- Analyzing the trends and fluctuations in share prices of top 10 companies from January 2011 to December 2022.
- Fitting of ARIMA, ARCH/GARCH time series model in a stock market.
- Forecasting of the ARIMA time series model.
- Comparison of ARIMA & ARCH time series model.

4.Model Descriptions

The time series is order to learn a concept behind the application of a models. A collection of random variables with a focus on the time at which they occur is termed as a time series. The development of mathematical models that might describe sample data is the main goal data-driven time series. In the context of time series analysis, a variety of methods might be used to estimate how the Data changed over time. Different kinds We're going to study and analyse the ARIMA and ARCH/GARCH models.

There are a few important steps that must be followed while working with time series analysis. plotting the data on a graph from right away is one of the crucial initial actions. The researcher can learn more about the type of data being 1analysed while also observing certain features of data. The trend in the data may be determined from the plot, thus it is essential to observe the data and identify any downward or increasing trends throughout time. The data's seasonality, that results in some patterns over time, is another important aspect. Since some forecasting techniques can't handle seasonality, it must be removed before the forecasting process, making this aspect essential to many different forecasting techniques. The analysis of outliers in a time series involves more issues. In regression models, outliers are particularly significant and may significantly change the model's performance. To determine variance is constant this can be difficult to determine when merely looking at the plot of the data, the variance must also be examined.

4.1. ACF AND PACF

Calculation of Autocorrelation Function(ACF) and the Partial Autocorrelation Function(PACF) is concept before defining the models themselves.

The ACF is written as:

$$\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}}$$

The ACF determines the linear predicated of the series t, say Xt, use this value Xs.

The partial autocorrelation function can be used to determine a stationary time series' partial correlation with its own lag values. It diminishes at all shorter legs of the time series values. It contrasts with the autocorrelation function, which eliminates further lags.

A conditional correlation in general is a partial correlation. because of that we know and, there is a relationship between those two variables. take into account the variables of another set of variables while assessing a regression where x_1, x_2, and x_3 are determiner variables. The relationship between y and x_3 in this case is the Determined is the correlation between the variables by how both are related to x_1 and x_2.

It may be represented mathematically as:

$$\frac{covariance(y, x_3|x_1, x_2)}{\sqrt{variance(y|x_1, x_2)variance(x_3|x_1, x_2)}}$$

To find the order of models, the ACF and the PACF analysed together.

4.2. STATIONARITY

Stationarity is concept for modelling time series data. A stationary time series' mean and variance must not change over time. In general, If a stochastic process' mean, variance, and covariance between two time periods are all constant throughout time, it is said to be stationary. is determined only by the lag between the two time periods, instead of the actual time at which the covariance is computed. A time series is known to be constant If there isn't a consistent variation in the mean over time, no consistent variation in in the Period variations and variance have been predicted said to be stationary.

One important thing to remember is that the Data from time series must be stationary. before to conduct any time series analysis. There are many tests for stationary diagnosis, such as the Augmented Dickey Fuller test. The data is stationary, we can continue.; otherwise, the data must be altered to make it stationary. The ARIMA model is the differenced series' ARMA model.

To find the stationarity, there are only a few processes that might be used. Along with them is the Augmented Dickey-Fuller ADF test. A Dickey-Fuller test is in the model equation below; the unit root test examines the null hypothesis that $\alpha = 1$. The initial lag on Y's coefficient is called alpha.

Null Hypothesis (H0): α=1

$$y_t = c + \beta_t + \alpha y_{t-1} + \phi \Delta Y_{t-1} + e_t$$

where,

- y_{t-1} = time series' first lag
- ΔY_{t-1} = first difference of the series at time (t-1)

The null hypothesis corresponds to that of the unit root test. Specifically, the coefficient of Y(t-1) is 1, showing that there is a unit root present. The series is assumed to be non-stationary if it is not rejected.

4.3 ARIMA MODELS

The ARIMA models is the most closely examined in this study. These models are often referred to as the Box-Jenkins methodology. ARIMA stands for 'Autoregressive Integrated Moving Average'.

The models were developed in 1976, and their results are based on the forecast of the lag of the y variable itself as a function of the dependent variable y, as expressed in p autoregressive terms, with q terms representing the past error. Due to the non-stationarity of many financial time series, variable transformation is necessary for using ARIMA models (p, d, and q). ACF and PACF must be analysed together to identify the model's orders, with the goal of minimising the AIC, a method to evaluating the quality of the model.

To better learn the models, we have to start with the AR and MA parts in the equation.

4.3.1. AR Models

AR stands for ‘Autoregressive’ model. They depend on the principle that a function of the previous values used to express the present value of the series, Xt. AR model can be defined by the equation:

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p} + \varepsilon_t$$

Where,

X_t: observation in time t

Φ_p: parameter of AR

p: order of model

ε: error term

4.3.2. MA Models

MA stands for 'Moving Average' model is the opposite of autoregressive models. It assumes that the equation's left-hand side looks as linear, MA model of order q, written as MA(q), implies that the observed data is created by a linear set of the white noise. This can be written as:

$$X_t = \omega_t + \theta_1\omega_{t-1} + \theta_2\omega_{t-2} + \cdots + \theta_q\omega_{t-q}$$

Where the MA q lags, and $\theta_1, \theta_2, \ldots, \theta_q$ are parameters

4.3.3. ARMA Models

The ARMA model is formed by combining AR and MA models. They can be expressed numerically as:

$$X_t = \Phi_1 X_{t-1} + \Phi_p X_{t-p} + \alpha_t - \theta_1\alpha_{t-1} - \cdots - \theta_q\alpha_{t-q}$$

ARMA models are not appropriate for modelling non-linear relationships, however, critical to the knowledge of stationary time series and AR models.

As we take differencing, the ARMA model becomes p is the AR term, The order of differencing is d., and q is the MA term in the ARIMA (p, d, q) model.

Steps for modelling with ARIMA model:

I. Check the stationary of the series and Perform ADF test after analysing ACF and PACF
II. If it isn't stationary, make it thus by modifying or take difference it once or more.
III. Check whether the given series is AR, MA, or ARIMA.
IV. Evaluate the model.
V. Evaluate residuals.
VI. forecast using the model.

It is important to note that in order to estimate the orders of the model, the ACF and the PACF must be analysed with regard to the modelling steps. Theoretical ACF and PACF patterns are shown in the table below:

Table 4.1: Theoretical Patterns of ACF and PACF

THEORY OF ACF AND PACF PATTERNS		
Model	**Pattern of ACF**	**Pattern of the PACF**
AR (p)	Exponentially Decays or damped sine wave pattern, or both.	Significant spikes caused by lags p
MA (q)	Significant spikes caused by lags q	Exponentially declines
ARMA (p, q)	Exponentially decays	Exponentially decays

Akaike Information Criterion:

The Japanese statistician Hirotugu Akaike created the Akaike information criteria in 1974, and it bears his name. It now forms the basis of a paradigm for the statistics and is also widely used for statistics inference.

The Akaike Information is one way for evaluating the model, which may be mentioned mathematically as:

$$AIC = 2k - 2\ln(\hat{L})$$

Where,

K: number of model's evaluated parameters

$\hat{L}$: Maximum likelihood function value for the model

Bayesian Information Criterion:

In the statistical process of choosing a model, the Bayesian Information Criterion (BIC) is employed. It was first introduced by Gideon Schwarz in 1978 as an extension of the Akaike Information Criterion (AIC).

The BIC is formally defined as,

$$BIC = -2\ln(L) + kln(n)$$

Where,

L: Maximum likelihood function value for the model

K: number of model's evaluated parameters

n: Number of observations

To compare two or more models, the model is selected which has lowest AIC value. A term associated with the number of parameters in the model affects the variance error.

The Ljung-box test is used to analyse the residuals. The residuals must be independent of each other in the context of the model. The p-value of test indicate that Whether or not the residuals are independent. The test's null hypothesis is that there is no correlation between the residuals.

As determined by the Ljung − Box test statistic:

$$Q' = T(T+2)\sum_{h=1}^{N} \frac{\rho^2\,(h)}{T-h} \sim \chi^2{}_{\mathrm{N}}$$

Where,

T: sample size

ρ^2 : sample correlation at lag h

N: number of lags tested.

The root mean square error (RMSE) is a used to measure difference between observed and predicted values. It is a highly efficient way to assess model correctness. RMSE is defined as:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(Observed - Predicted)^2}{N}}$$

Where,

N: number of observations

Predicted: predicted value

Actual: actual value

4.4 ARCH/GARCH MODELS

In situations with high volatility, ARCH models are used. Because of its characteristics, such as high volatility and variance. So it is commonly used in financial time series. ARCH is the short form for autoregressive conditional heteroscedasticity. In the sense that heteroscedasticity observed over various time periods may not be associated, heteroscedasticity may have an autoregressive structure. The ARCH model serves as a time series volatility model.

It is important to test that the time series have enough features to support the use of heteroscedastic models to apply the ARCH model. If the mean of the time series is 0then the ARCH model can be defined as:

$y_t = \sigma_t \varepsilon_t$

Where σ_t:

$$\sigma t = \sqrt{\alpha_0 + \alpha_1 y_{t-1}^2}$$

According to the $\mathrm{ARCH}(1,0)\mathrm{model}$, the variance of the model $\mathrm{y_t}$is as follows when

y_{t-1} is taken into account:

$$\mathrm{Var}(y_t|y_{t-1}) = \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$$

To understand the order of model, it is important to analyse the ACF and PACF plots. After getting the order of model, it should be fitted on data to estimate the parameters.

ARCH models have gone through many changes since they were first introduced. GARCH models, or generalised autoregressive conditional heteroscedasticity models, are among these variations The model time t is often expressed as follows: GARCH (1,1), which may be expressed as:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

The variants of the ARCH models, referred to as GARCH models, were given.

5. METHODOLOGY

As was said before, the study's major goal is to conduct time series analysis in depth in order to explain the past stock values in the future. These tasks are performed in works of this kind for better the results

- Data Collection and Data Description
- Data Visualization
- Analysis Using Software

5.1Data Collection and Data Description

Every data pertaining to financial markets has to be publicly open due to constitutional demands and open. In this case, all the data necessary to carry out the analysis appears on the National Stock Exchange and Yahoo Finance website. The collected data is on daily bases from January 2011 to December 2022.

Our Dataset is as follows:

	Date	Symbol_1	Open_1	High_1	Low_1	Close_1	Adj Close_1	Volume_1	Symbol_2	Open_2	...	Close_9	Adj Close_9	Volur
0	2011-01-03	RELIANCE	527.499817	528.242798	521.358032	522.843933	477.112671	4804792	TCS	583.500000	...	282.209991	249.201828	1106
1	2011-01-04	RELIANCE	525.023315	534.879883	523.685974	533.493042	486.830414	10163093	TCS	581.950012	...	273.559998	241.563553	3103
2	2011-01-05	RELIANCE	534.929382	539.882446	529.976379	532.849121	486.242706	11972802	TCS	572.625000	...	269.320007	237.819550	2521
3	2011-01-06	RELIANCE	533.938782	540.575867	532.229980	537.703125	490.672211	10043704	TCS	584.000000	...	262.239990	231.567627	3716
4	2011-01-07	RELIANCE	535.573303	538.768005	524.032715	527.697937	481.542114	8249359	TCS	588.000000	...	260.095001	229.673508	4443
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
2953	2022-12-26	RELIANCE	2514.750000	2542.000000	2492.399902	2524.050049	2524.050049	2764496	TCS	3228.350098	...	597.099976	597.099976	1320
2954	2022-12-27	RELIANCE	2530.000000	2548.800049	2515.250000	2544.699951	2544.699951	2659749	TCS	3269.199951	...	601.900024	601.900024	963
2955	2022-12-28	RELIANCE	2538.000000	2549.800049	2521.500000	2544.449951	2544.449951	3442509	TCS	3249.800049	...	601.049988	601.049988	798
2956	2022-12-29	RELIANCE	2527.000000	2548.899902	2525.500000	2543.300049	2543.300049	3198493	TCS	3231.100098	...	611.799988	611.799988	2013
2957	2022-12-30	RELIANCE	2545.100098	2577.000000	2541.100098	2547.199951	2547.199951	3364092	TCS	3286.050049	...	613.700012	613.700012	1305

2958 rows × 71 columns

Table No 5.1: Data Review

Dataset containing 10 companies stock prices over the period of January 2011 to December 2022. The companies are Reliance, TCS, HDFCBANK, Infosys, Hindustani Unilever, ICICIBANK, HDFC, ITC, SBIN and Bharti Airtel.

Each company having their symbol, open price, high price, low price, close price, adjacent close price and volume.

The variable description is as follows:

Variable Name	Description
Symbol	This means that the name of that company in the form of symbol.
Close Price	The closing price of the share market before the market closes.
Open Price	The open price of share market.
High Price	High price of a particular stock.
Low Price	Low price of a particular stock.
Adjacent Close	Closing price of a stock on previous day.
Volume	This means that the how many shares are buy and sell during the day.

The summary statistics of all the data is mentioned in table below:

Company	Mean	SD	Skewness	Kurtosis
RELIANCE	1021.5242	743.6225	0.9851	-0.4805
TCS	1655.8151	939.1529	0.8237	-0.4265
HDFCBANK	786.8839	446.4073	0.3507	-1.2620
INFOSYS	702.4715	427.8989	1.3662	0.6322
HINDUNILVR	1268.3777	752.9964	0.4121	-1.2889
ICICIBANK	349.5897	197.5050	1.3385	0.8019
HDFC	1521.5959	663.2114	0.2866	-1.1604
ITC	226.1818	51.0193	-0.1650	-0.1160
SBIN	282.6437	99.5572	1.3589	1.4124
BHARTIARTL	400.7080	143.4151	1.4047	0.9432

5.2Data Visualisation

An exploratory data in order to utilize the model properly, analysis must be done. Some steps must be considered and followed in the context of this study, is as follow:

- The original series' plot
- ACF and PACF Plot
- Perform autoarima function order of model
- Split data into train and test data set
- Fit the model using train data set and estimates the parameters
- Execute all essential tests to check that the models will produce the optimal outcomes.

It is important to plot original series in order to determine several important time series characteristics like trend, and outliers. ARIMA models will not fit the data efficiently. The series' stationarity is another important part Depending on this first plot, what kind of information we need to ascertain. The latter is required in order to create various forecasting models.

The ACF and PACF will be data is stationary. It will provide details about the order of the models. This rule applies to both homoskedastic and heteroskedastic models. Our data is non-stationary in this context, and we are unsure on the order of model from ACF and PACF. As a result, we use the autoarima function to determine the model's order.

Autoarima function creates a different combinations of orders and the order which has minimum AIC value will be the best fit model for our data. Autoarima gives the appropriate order of model then fit the ARIMA model on train dataset and estimate the parameter. After model fitting, forecast the test data set to check the accuracy of model. Also check the accuracy of model to understand how much our model is significant to predict future events.

Regarding exploratory data analysis, each company will be assessed independently. Each company followed the same process.

Initially all companies stock prices were plotted on a single graph. The figure below shows the trend lines for all the 10 companies.

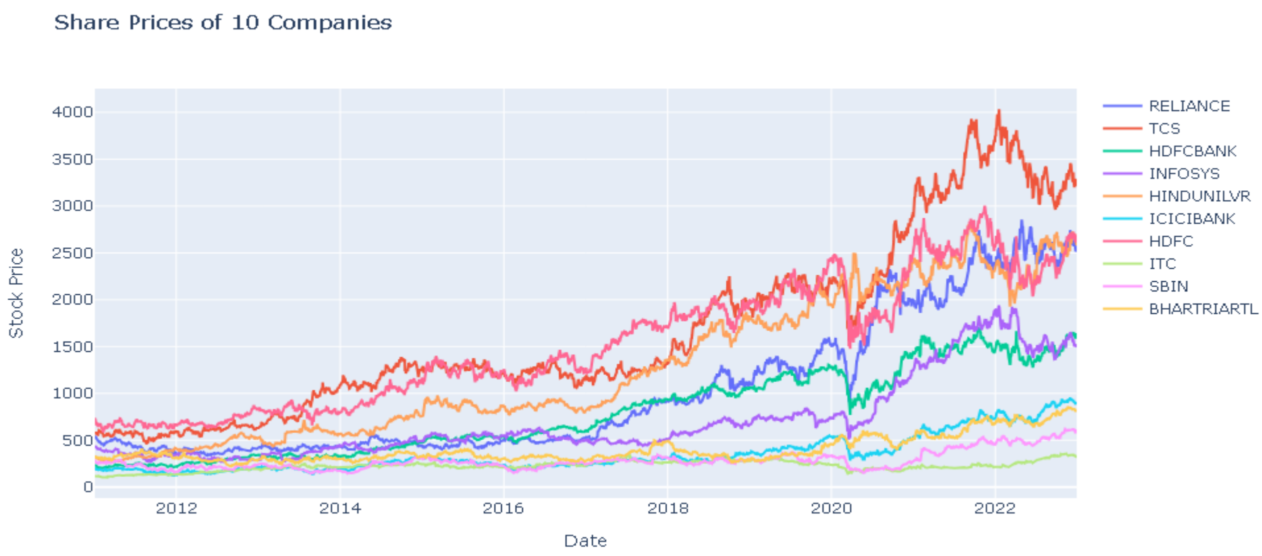

Fig 5.1: Stock prices plot

With respect to stock prices plot, the conclusion is that share prices has increasing trend. In the period of pandemic they show decreasing trend, however; there is a huge growth in share prices after the pandemic.

These 10 companies are from different sectors such as IT Services and Consulting, Financial services, Consumer Goods, Oil-Gas & Consumable Fuels and others. In which TCS and Infosys belongs to IT services and Consulting sector, HDFC bank, ICICI bank and SBI bank belongs to Financial services, Hindustani Unilever and ITC belongs to Consumer Goods sector and Reliance, HDFC and Bharti Airtel belongs to Oil-Gas, Housing Development and Telecom sector respectively.

The sector wise share prices are shown in the figures below:

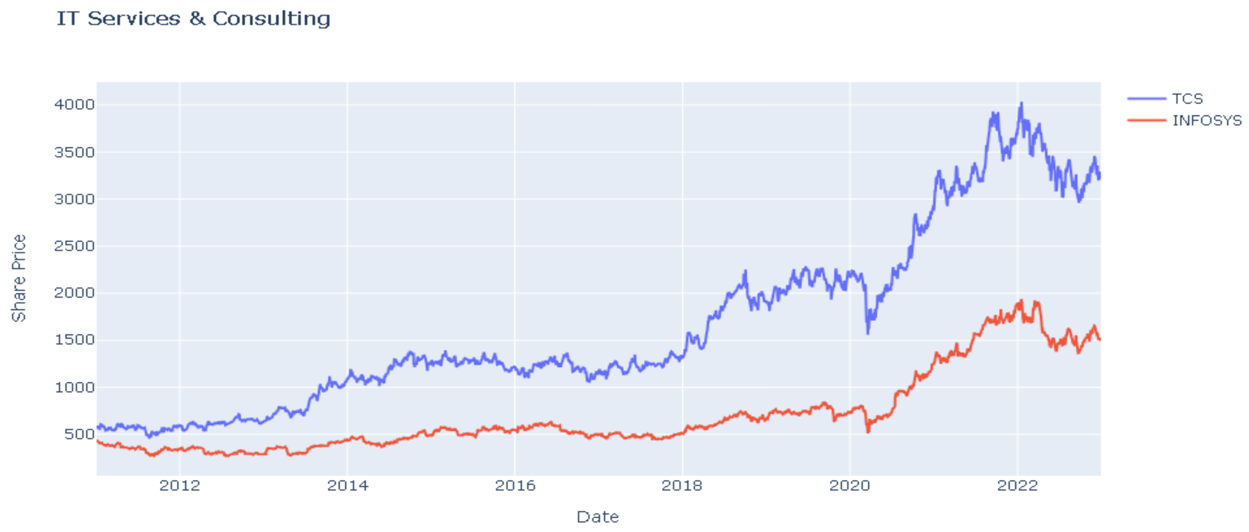

Fig 5.2: IT services & consulting stock prices

From fig 5.2 we conclude that stock price of TCS is relatively high than that of Infosys stock price.

Fig 5.3: Financial Services stock prices

From fig 5.3, we conclude that stock price of HDFC Bank is relatively very high than other two.

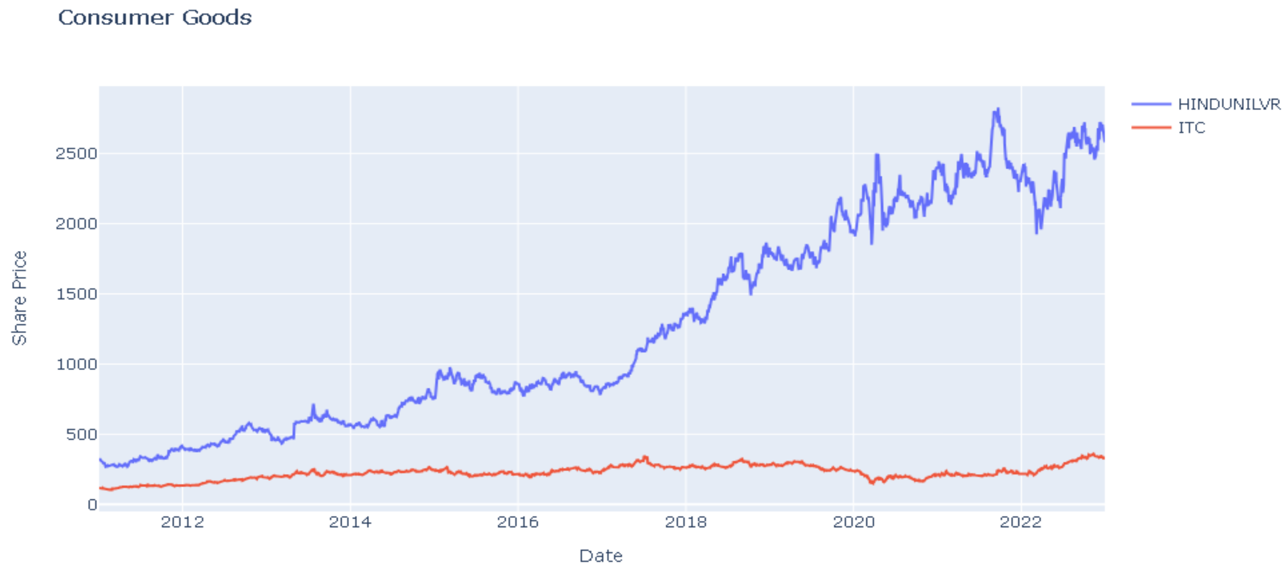

Fig 5.4: Consumer Goods sectors stock prices

From fig 5.4, we conclude that stock price of Hindustani Unilever is relatively high than that of ITC stock price.

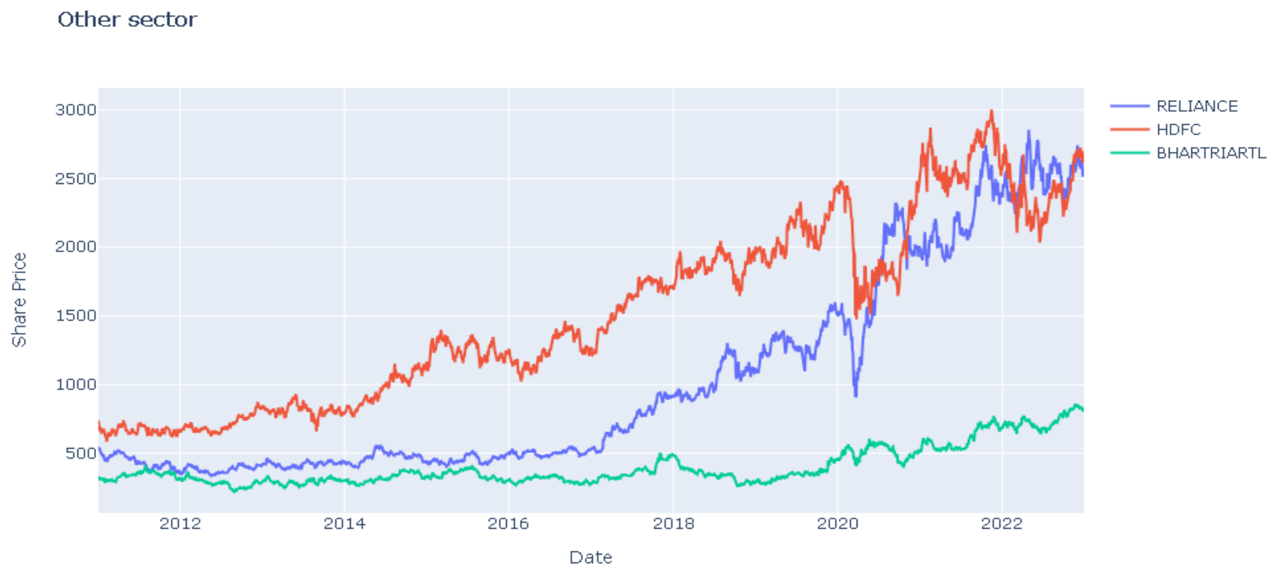

Fig 5.5: Other sectors stock prices

From fig 5.5, we conclude that stock prices of Reliance and HDFC are relatively high than that of Bharti Airtel stock price.

5.3 Analysis Using Software

The ADF test should be done to assess the series' stationarity. A unit root, or non-stationarity of the series, is the null hypothesis in this test. There may not be a unit root, is an alternate hypothesis. The result of ADF test for 'Reliance' is given in table 5.3.

Table 5.2: Reliance Unit-Root test

ADF Test		
	t-statistic	Prob
Test Statistic	0.9749	0
Test Critical Values		
1%	-3.4330	
5%	-2.8630	
10%	-2.5670	

A unit root, or non-stationarity of the series, is the null hypothesis in this test. There may not be a unit root, is an alternate hypothesis.

The table below summarises the stationarity results for the other companies analysed in the study:

Company	**ADF Stat**	**p-value**	**Observation**	**Conclusion**
TCS	-0.3268	0.9216	p-value > Critical value	Non-stationary
HDFCBANK	-0.1356	0.9457	p-value > Critical value	Non-stationary
INFOSYS	0.2186	0.9732	p-value > Critical value	Non-stationary
HINDUNILVR	-0.0634	0.9529	p-value > Critical value	Non-stationary
ICICIBANK	0.8895	0.9929	p-value > Critical value	Non-stationary
HDFC	-0.6989	0.8470	p-value > Critical value	Non-stationary
ITC	-1.7325	0.4144	p-value > Critical value	Non-stationary
SBIN	0.0006	0.9585	p-value > Critical value	Non-stationary
BHARTIARTL	0.2135	0.9730	p-value > Critical value	Non-stationary

Table 5.3: Result of All Companies

After the stationarity test, we plot ACF & PACF to determine the order of series. The ACF measures the correlation between a time series and its lag values. It helps in determining the existence of any correlation or pattern in time series data. The Partial Autocorrelation Function accounts for the intermediate delays to determine the correlation between a time series and its lagged values. It provides insights into the direct relationship between observations at different lags, excluding the indirect relationships through the intermediate lags. Following plots show the ACF & PACF of each of the company at different lags.

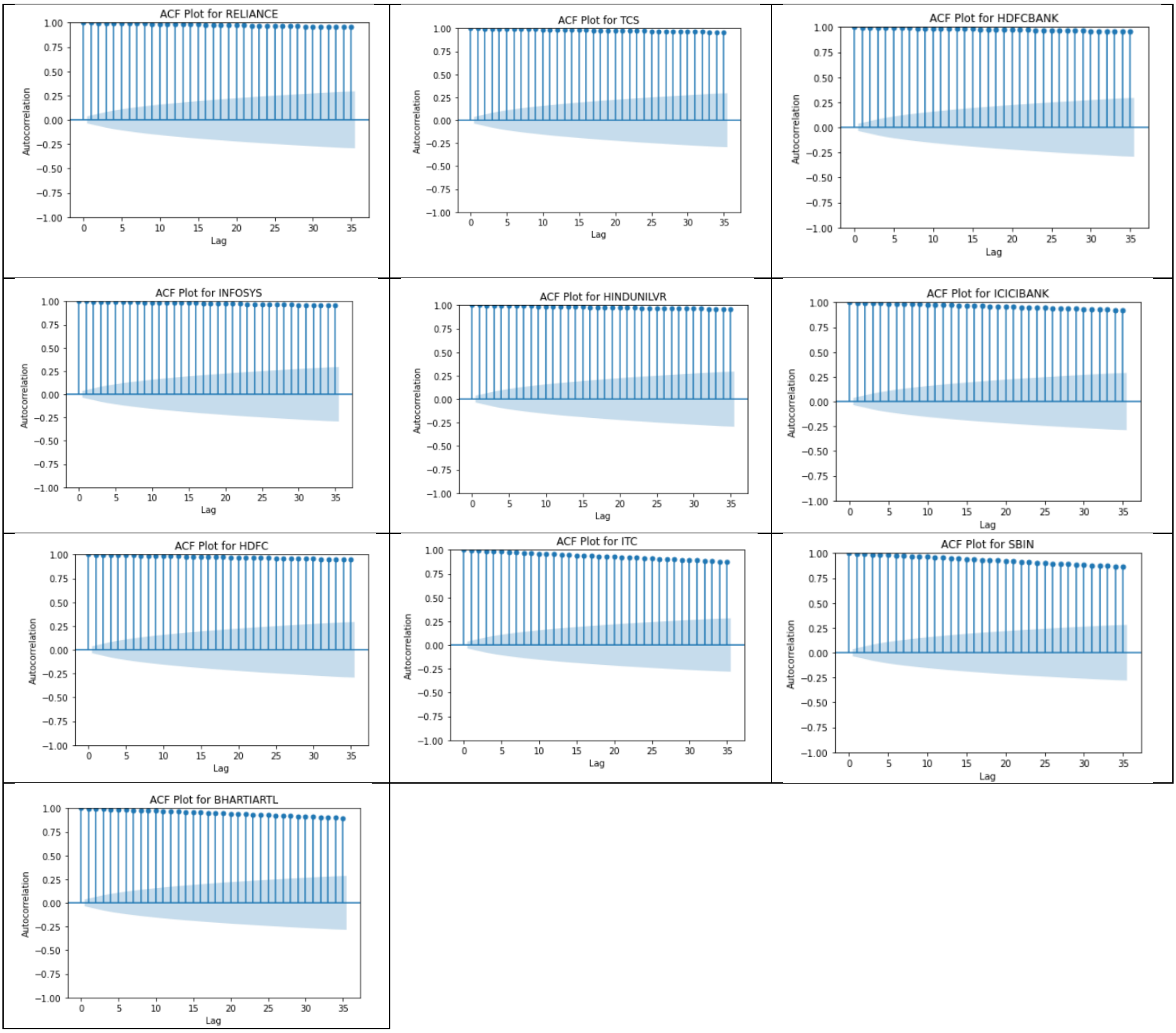

Fig 5.6: ACF plots

The autocorrelation plot for various delays is shown above. For significance level 0.05, we can observe that there is some autocorrelation in this case.

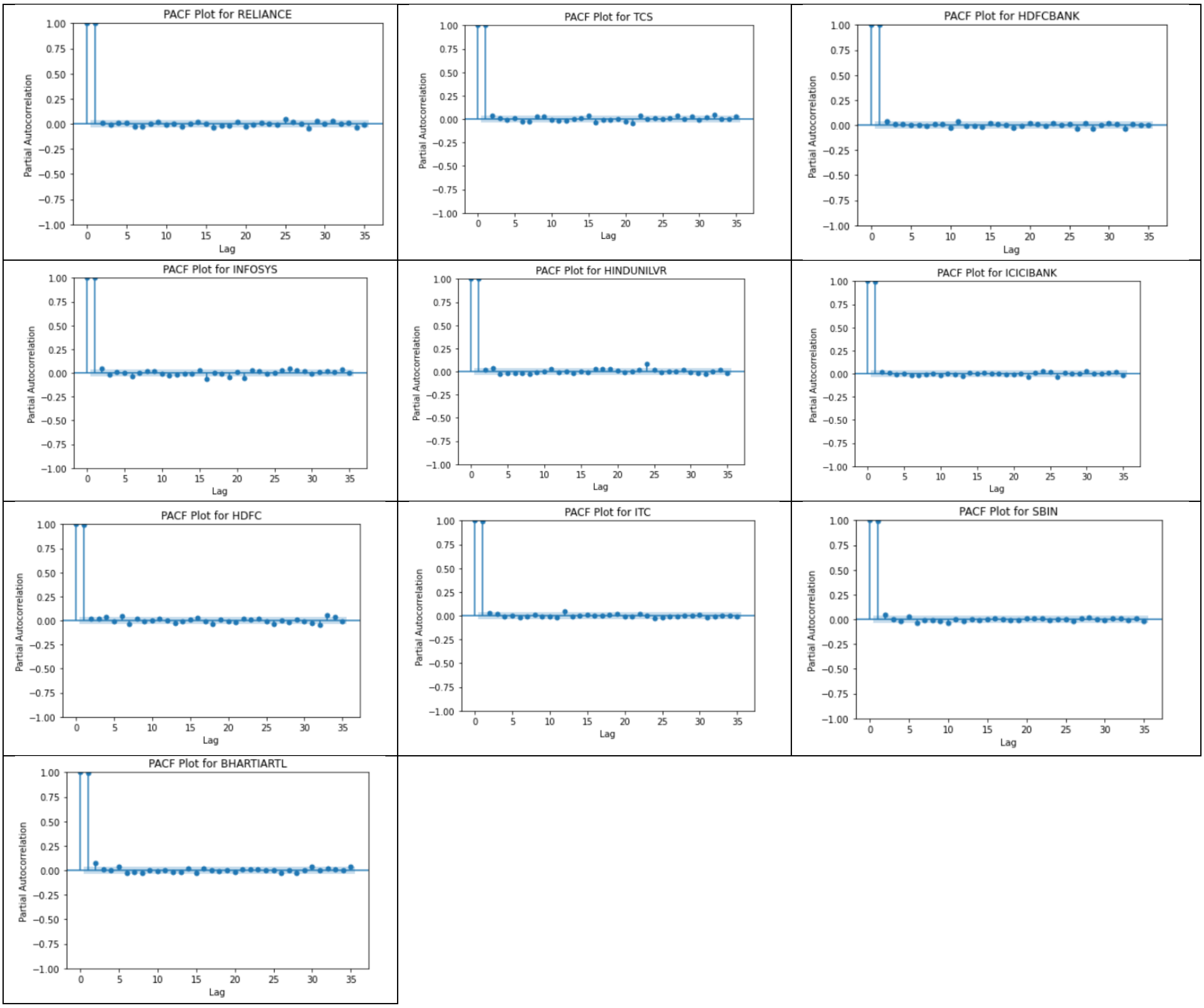

Fig 5.7: PACF plots

We can see from the partial autocorrelation that there is some partial autocorrelation for the various values of delays at a 0.05 level of significance. Although the 100% partial autocorrelation for lag 0 is clear, the partial autocorrelation for lag 1 is likewise quite high.

5.4 ARIMA MODEL

A systematic family of time-correlated modelling and forecasting models known as the autoregressive integrated moving average model is introduced in Box and Jenkins. when using this model, the first step is to plot the time series while looking for trends, outliers, and seasonality. For stationarity, the ADF test was already performed.

All time series are non-stationary; apply first or second order differencing to make the series stationary. It could be very difficult to define the order of the models only based on the ACF and PACF; therefore, with the help of auto Arima function The optimal model will be chosen based on the findings of this study after multiple models have been tested and evaluated.

ARIMA (5,1,0) model for company Reliance according to the AIC, is the best, as the AIC have minimum values in model. Using ARIMA (5,1,0) model Reliance:

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                  close   No. Observations:                 2958
Model:                 ARIMA(5, 1, 0)   Log Likelihood              -13403.449
Date:                Wed, 14 Jun 2023   AIC                          26818.897
Time:                        15:52:09   BIC                          26854.849
Sample:                             0   HQIC                         26831.838
                               - 2958
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.0218      0.011      1.939      0.053      -0.000       0.044
ar.L2          0.0237      0.010      2.321      0.020       0.004       0.044
ar.L3         -0.0438      0.010     -4.324      0.000      -0.064      -0.024
ar.L4         -0.0338      0.011     -3.071      0.002      -0.055      -0.012
ar.L5          0.0344      0.012      2.988      0.003       0.012       0.057
sigma2       507.2793      5.205     97.456      0.000     497.077     517.481
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):             15589.22
Prob(Q):                              0.96   Prob(JB):                         0.00
Heteroskedasticity (H):              25.49   Skew:                             0.08
Prob(H) (two-sided):                  0.00   Kurtosis:                        14.25
===================================================================================
```

Fig 5.8: Reliance ARIMA Model

Using AIC, this is best model. In terms of residuals, they are plotted below:

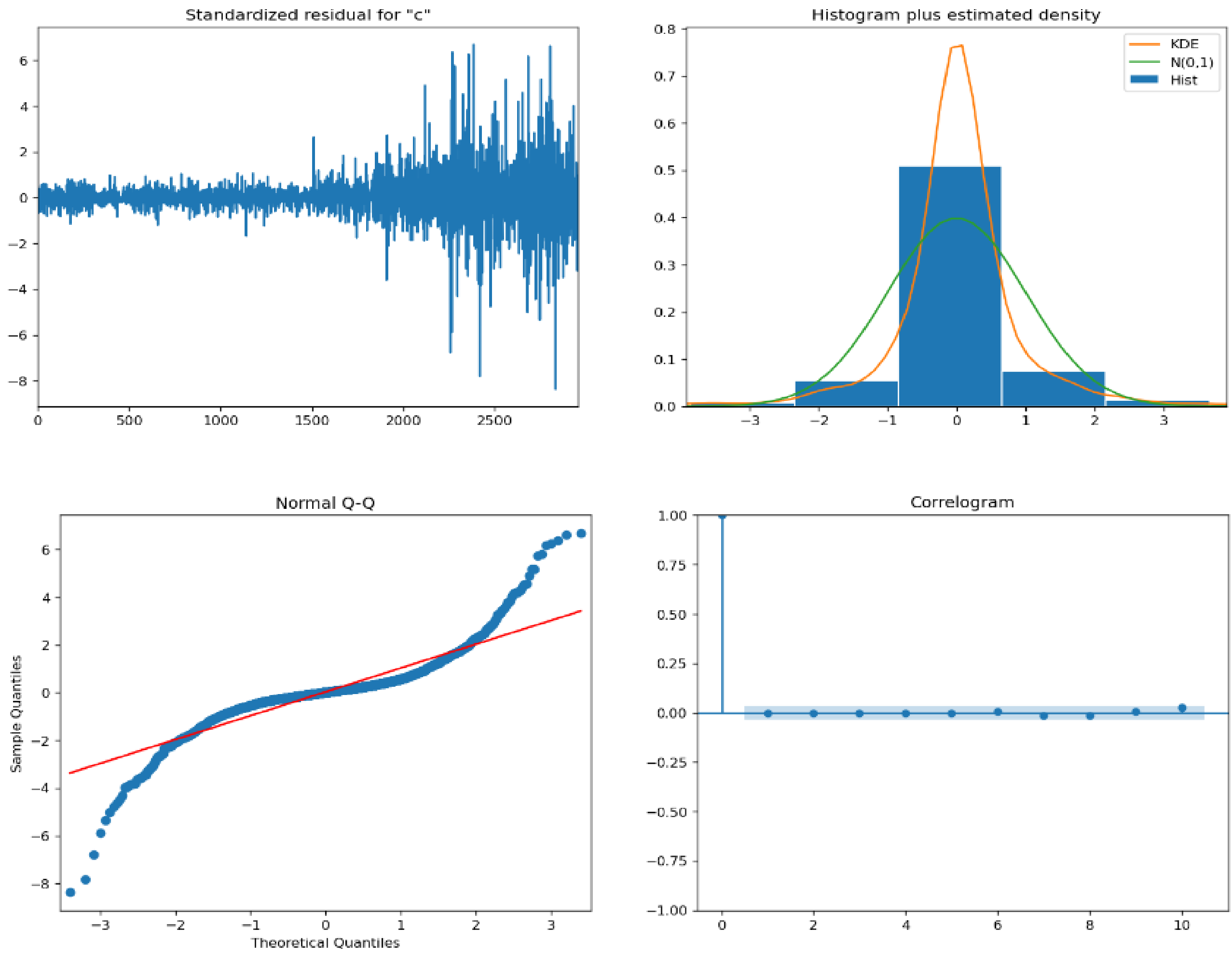

Fig 5.9: Residual Plots

Top left: The variance and variability of the residual errors seem to be uniform, with a mean zero.

Top Right: Residuals are normally distributed with mean zero and standard deviation one.

Bottom left: All of the dots must line up exactly with the red line. A skewed distribution would be shown by any large deviations.

Bottom Right: The residual errors are not autocorrelated, as shown by the Correlogram, also known as the ACF plot.

The results for all the companies are displayed in the table below:

Name Of Company	Parameters	Standard Error	AIC	BIC
Reliance ARIMA(5,1,0)	ϕ1= 0.0218	0.011	26818.897	26854.849
	ϕ2= 0.0237	0.010		
	ϕ3= -0.0438	0.010		
	ϕ4= -0.0338	0.011		
	ϕ5= 0.0344	0.012		
TCS ARIMA (1,1,0)	ϕ1= -0.0039	0.012	28283.884	28295.868
HDFCBANK ARIMA(4,1,0)	ϕ1= 0.0490	0.009	23886.373	23916.333
	ϕ2= -0.0712	0.008		
	ϕ3= -0.0309	0.009		
	ϕ4= 0.0281	0.010		
INFOSYS ARIMA(5,2,0)	ϕ1= -0.8038	0.010	24275.882	24311.832
	ϕ2= -0.6647	0.013		
	ϕ3= -0.4897	0.015		
	ϕ4= -0.3231	0.014		
	ϕ5= -0.1132	0.011		
HINDUNILVR ARIMA(4,1,0)	ϕ1= -0.0460	0.009	26780.459	26810.419
	ϕ2= -0.0530	0.009		
	ϕ3= 0.0298	0.010		
	ϕ4= 0.0450	0.008		
ICICUBANK ARIMA (1,1,0)	ϕ1= 0.0066	0.012	20333.149	20345.132
HDFC ARIMA(5,1,0)	ϕ1= 0.0120	0.010	28477.927	28513.879
	ϕ2= -0.0486	0.010		
	ϕ3= -0.0847	0.010		
	ϕ4= -0.0094	0.012		
	ϕ5= 0.0485	0.010		
ITC ARIMA (2,1,2)	ϕ1= 0.5527	6.789	16054.185	16084.144
	ϕ2= 0.3292	6.234		
	θ1= -0.5680	6.787		
	θ2= -0.3373	6.344		
SBIN ARIMA(2,1,2)	ϕ1= 0.1072	0.307	18992.947	19022.906
	ϕ2= 0.6556	0.181		
	θ1= -0.0745	0.305		
	θ2= -0.6636	0.180		
BHARTIARTL ARIMA(2,1,3)	ϕ1= 0.8039	0.014	20774.887	20810.839
	ϕ2= -0.9778	0.130		
	θ1= -0.8619	0.021		
	θ2= 1.0138	0.020		
	θ2= -0.0581	0.014		

Table No 5.3: ARIMA Model Results

5.5 FORECASTING USING ARIMA MODELS

Making future predictions is known as extrapolation in the traditional statistical approach to time series data. In more modern fields, the phrase "time series forecasting" is used to refer to the subject. Forecasting is the process of utilising models fit on previous data to forecast future observations. In this study, forecasting is done using ARIMA models as ARIMA models has minimum standard error and AIC as compare to ARCH family.

Forecasting using ARIMA models involves fitting the model to historical data and then using it to make predictions for future values. Here's a steps of forecast using ARIMA models:

1. Data Preparation: Collected the historical time series data that will be forecasted. Checking missing values and outliers and ensure the data is in the right format.
2. Model Selection: Determined the best order for the ARIMA model. The ACF and PACF graphs are analysed to determine potential values for the Moving average (q), auto-regressive (p), and differencing (d) parameters or by autoarima function.
3. Model Fitting: ARIMA model is fitted on train data with appropriate values for the (p, d, q) parameters when initializing the model and estimated the coefficients.
4. Diagnostic Checking: Assessed the goodness of fit by examining diagnostic plots and statistics. Plotted the residuals to check for any patterns or unusual behaviour.
5. Forecasting: Use the fitted ARIMA model on testing dataset to make forecasts for future values.
6. Evaluate the Forecast: Compared the forecasted values to the actual values in the validation or test dataset.

The following figures show the forecasted stock prices for all the companies.

Fig 5.11 shows the forecasted prices for reliance company. Here, blue line is for training the model, red line is actual prices of stock and dotted green line is forecasted stock prices.

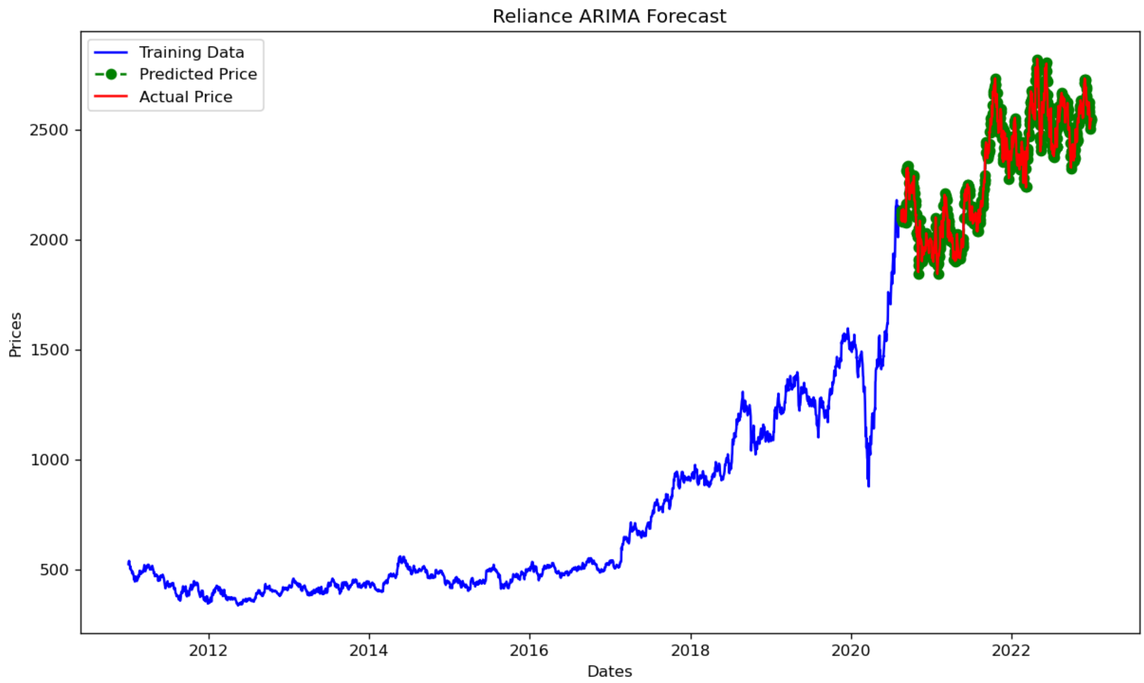

Fig 5.11: Reliance ARIMA Forecast

Fig 5.12 shows the forecasted prices for TCS company. Here, blue line is for training the model, red line is actual prices of stock and dotted green line is forecasted stock prices.

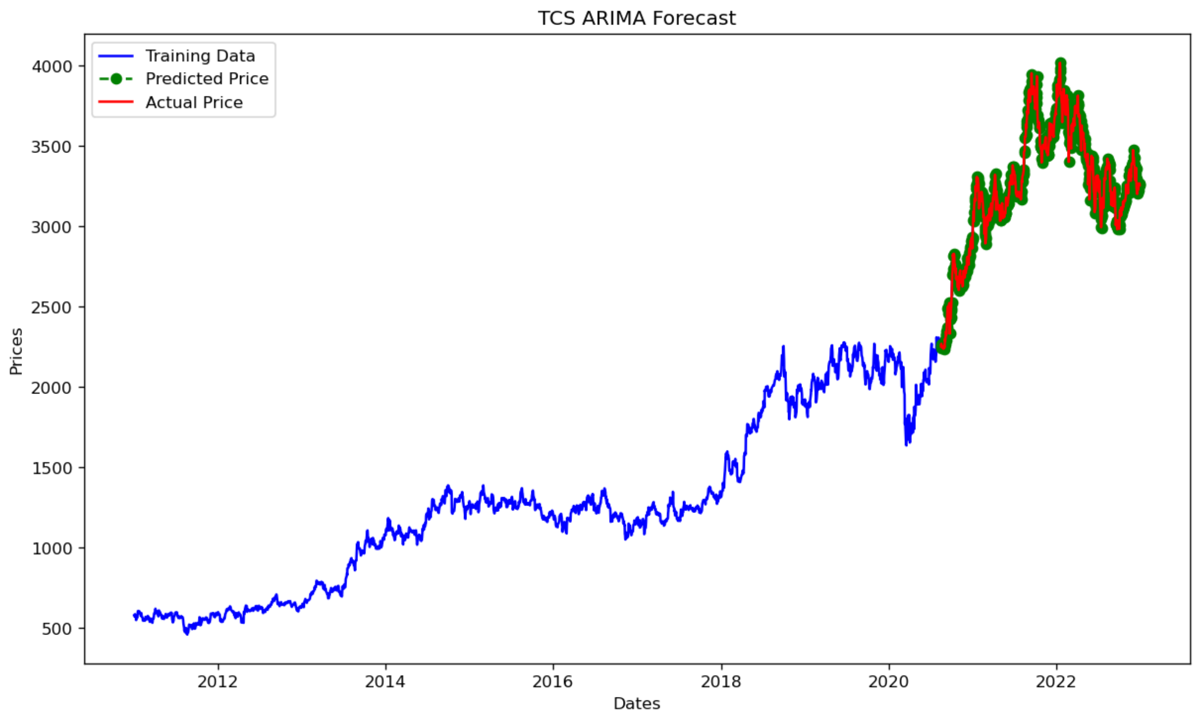

Fig 5.12: TCS ARIMA Forecast

Fig 5.13 shows the forecasted prices for HDFC Bank. Here, blue line is for training the model, red line is actual prices of stock and dotted green line is forecasted stock prices.

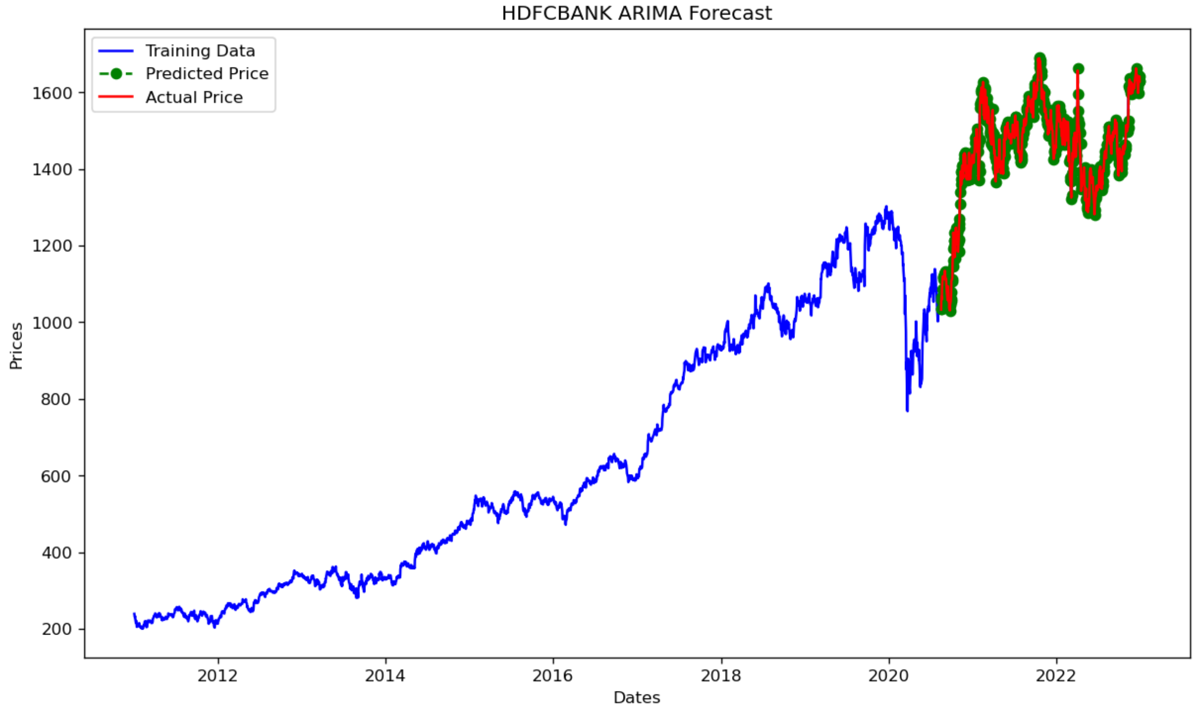

Fig 5.13: HDFC Bank ARIMA Forecast

Fig 5.14 shows the forecasted prices for Infosys company. Here, blue line is for training the model, red line is actual prices of stock and dotted green line is forecasted stock prices.

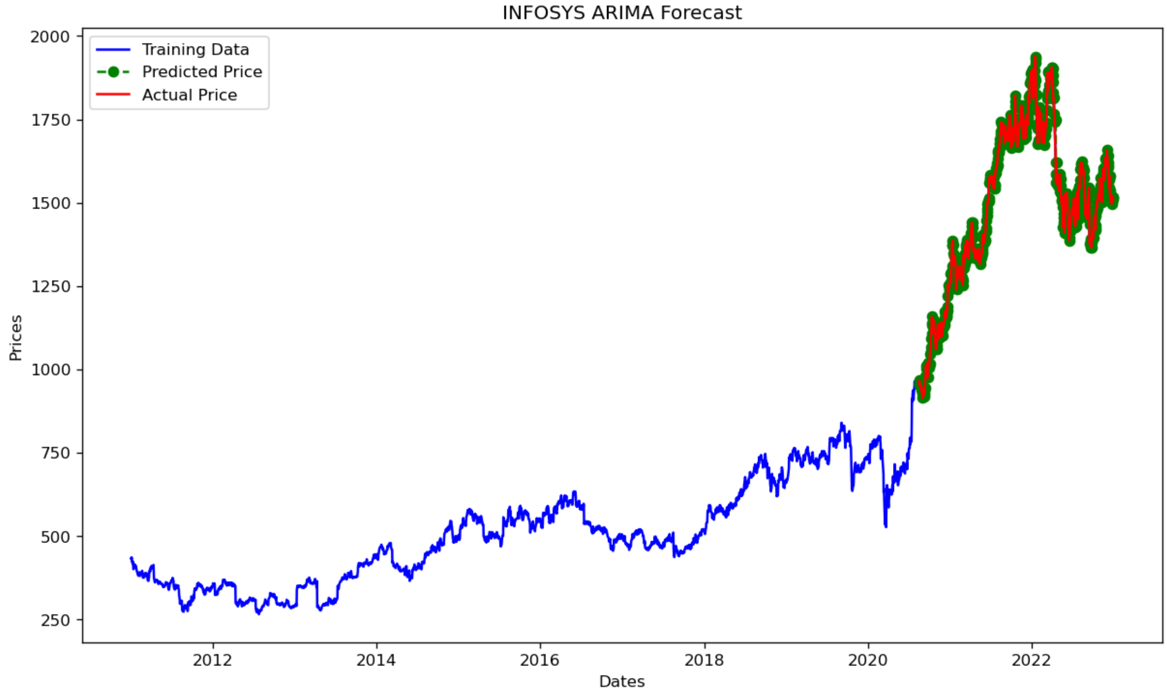

Fig 5.14: Infosys ARIMA Forecast

Fig 5.15 shows the forecasted prices for Hindustani Unilever company. Here, blue line is for training the model, red line is actual prices of stock and dotted green line is forecasted stock prices.

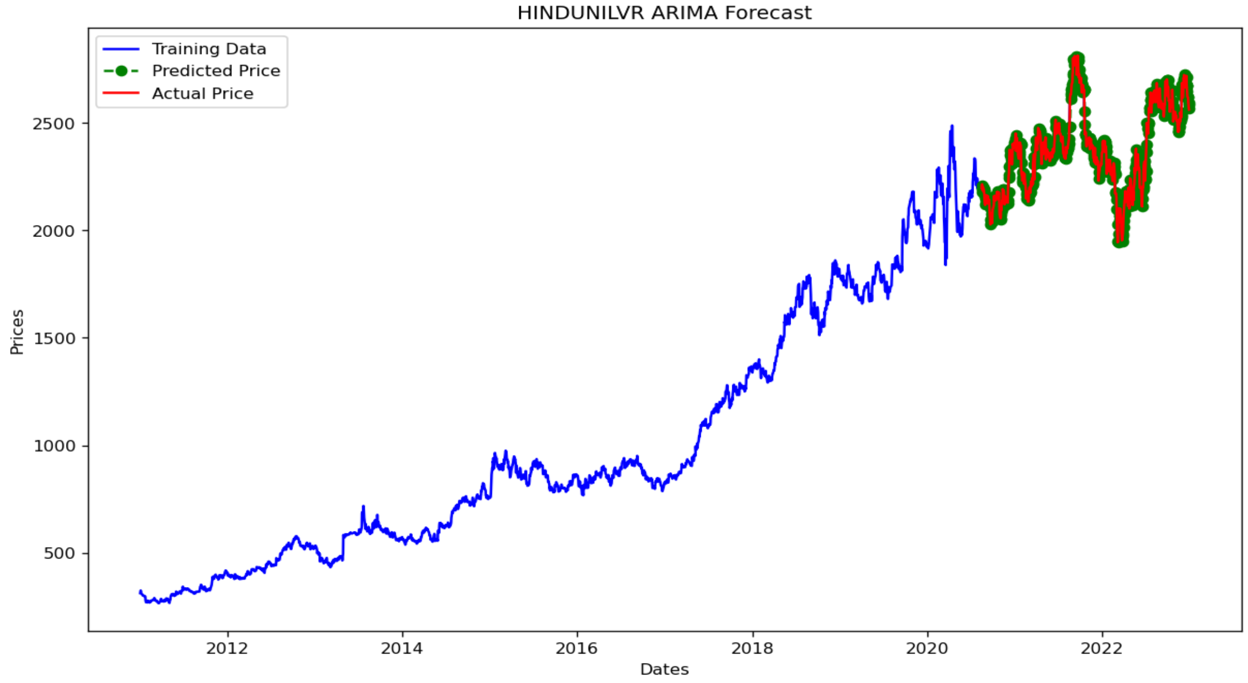

Fig 5.15: Hindustani Unilever ARIMA Forecast

Fig 5.16 shows the forecasted prices for ICICI Bank. Here, blue line is for training the model, red line is actual prices of stock and dotted green line is forecasted stock prices.

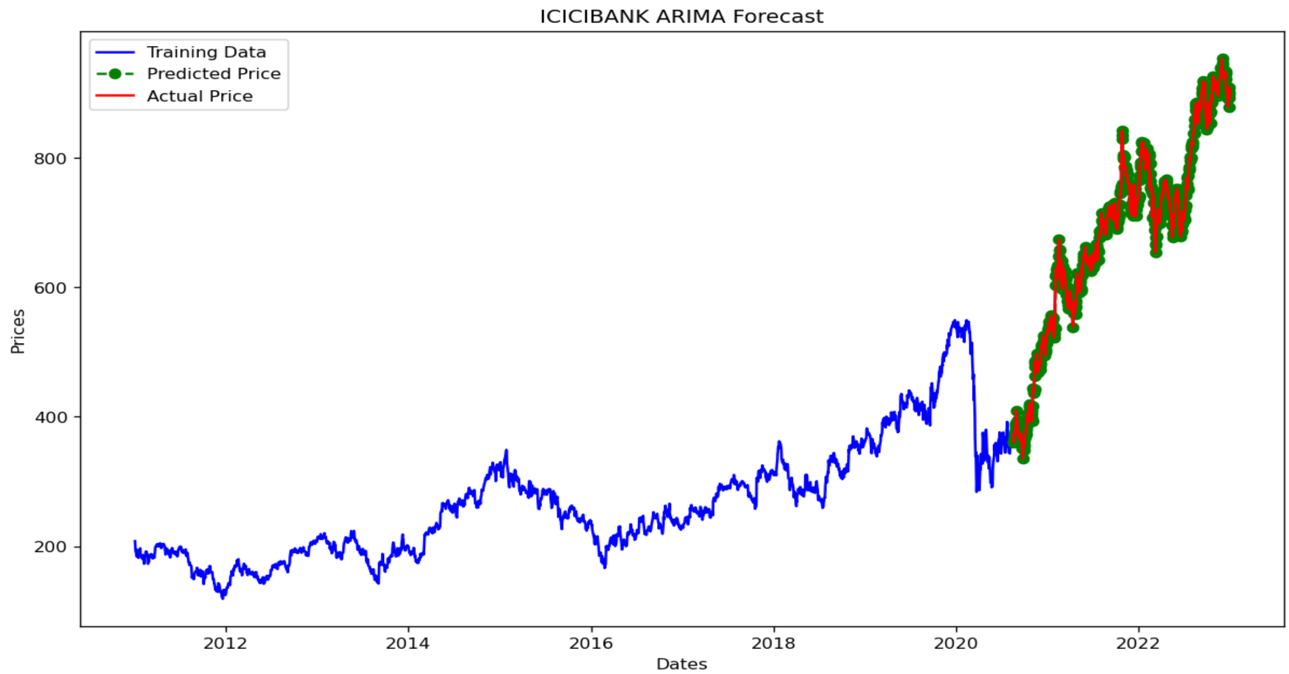

Fig 5.16: ICICI Bank ARIMA Forecast

Fig 5.17 shows the forecasted prices for HDFC company. Here, blue line is for training the model, red line is actual prices of stock and dotted green line is forecasted stock prices.

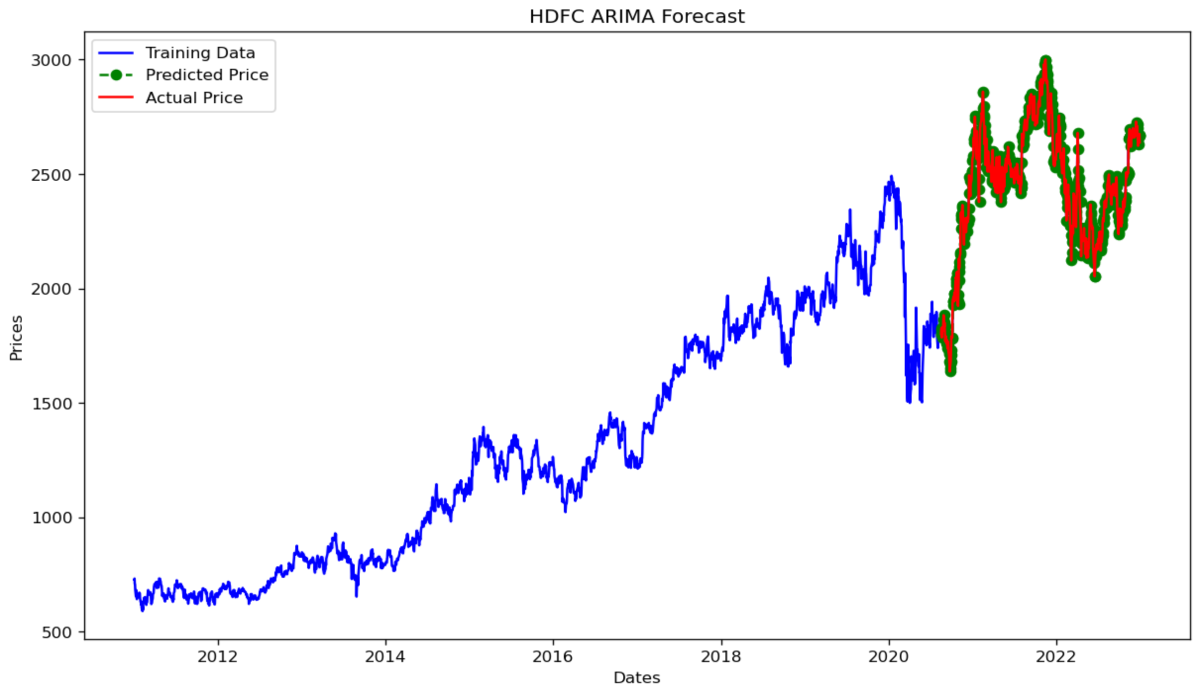

Fig 5.17: HDFC ARIMA Forecast

Fig 5.18 shows the forecasted prices for ITC company. Here, blue line is for training the model, red line is actual prices of stock and dotted green line is forecasted stock prices.

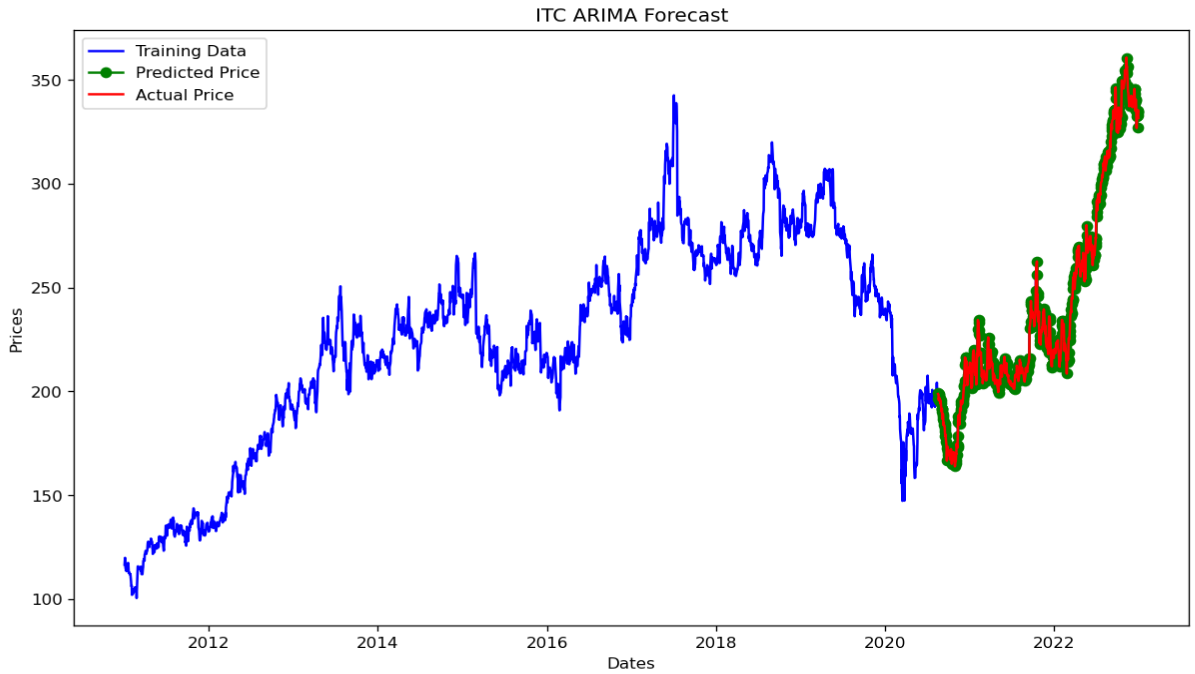

Fig 5.18: ITC ARIMA Forecast

Fig 5.19 shows the forecasted prices for SBIN company. Here, blue line is for training the model, red line is actual prices of stock and dotted green line is forecasted stock prices.

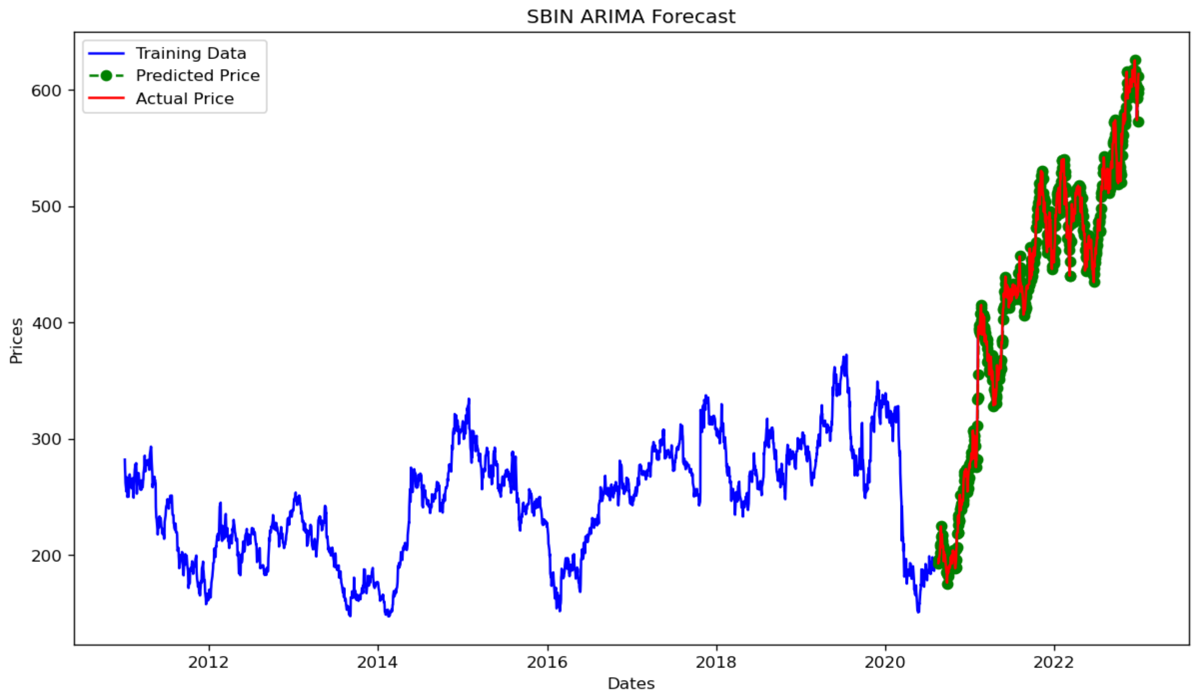

Fig 5.19: SBIN ARIMA Forecast

Fig 5.20 shows the forecasted prices for Bharti Airtel company. Here, blue line is for training the model, red line is actual prices of stock and dotted green line is forecasted stock prices.

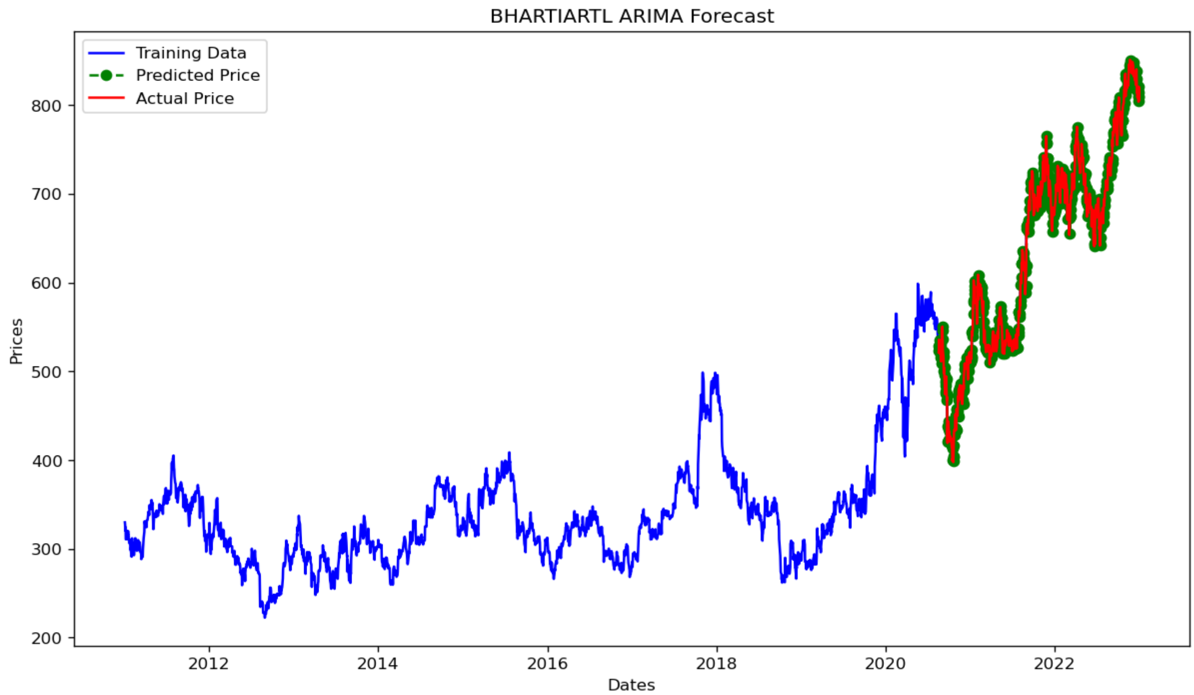

Fig 5.20: BHARTIARTL ARIMA Forecast

5.6 ARCH MODELS

The ARCH (Autoregressive Conditionally Heteroscedastic) model is used in time series when the model is based on the variance. This method was constructed specifically to deal with statistical and financial problems that have experienced considerable volatility.

The aim of this project is to figure determine which model gives the best outcomes. when financial time series are considered in Nifty50 index in National Stock Exchange, and while in specifically, ARCH models they might not always be the best while dealing with time series in finance.

For every situation, the ARCH/GARCH models will be used. For the purpose of analysing the results, the company will be able to do so. The Result analysis will then follow a thorough review that also considers the results of the ARIMA modelling process.

It is important to execute the model. as the initial step. By using this method, the residuals will be available for analysis. Then decide if the ARCH family of models is appropriate for the purpose of study.

Using the output of analysis for RELIANCE:

```
                      Constant Mean - ARCH Model Results
==============================================================================
Dep. Variable:                 RELIANCE   R-squared:                       0.000
Mean Model:               Constant Mean   Adj. R-squared:                  0.000
Vol Model:                         ARCH   Log-Likelihood:               -19482.1
Distribution:                    Normal   AIC:                           38978.3
Method:              Maximum Likelihood   BIC:                           39020.2
                                          No. Observations:                 2958
Date:                  Thu, Jun 15 2023   Df Residuals:                     2957
Time:                          01:07:45   Df Model:                            1
                                 Mean Model
============================================================================
                 coef    std err          t      P>|t|      95.0% Conf. Int.
----------------------------------------------------------------------------
mu           433.1978      3.172    136.556      0.000 [4.270e+02,4.394e+02]
                               Volatility Model
=============================================================================
                 coef    std err          t      P>|t|       95.0% Conf. Int.
-----------------------------------------------------------------------------
omega         45.4771      6.294      7.226  4.989e-13      [ 33.141, 57.813]
alpha[1]       0.9617  5.298e-02     18.151  1.257e-73      [  0.858,  1.066]
alpha[2]       0.0000  6.322e-02      0.000      1.000      [ -0.124,  0.124]
alpha[3]       0.0000  7.291e-02      0.000      1.000      [ -0.143,  0.143]
alpha[4]       0.0326  4.791e-02      0.680      0.497   [-6.134e-02,  0.126]
alpha[5]       0.0000  3.267e-02      0.000      1.000 [-6.404e-02,6.404e-02]
=============================================================================
```

Fig 5.10: Reliance ARCH Model

The following table shows the ARCH model results of all 10 companies:

Name Of Company	Parameters	Standard Error	AIC	BIC
Reliance ARCH(5)	α1= 0.9617	0.0530	38978.3	39020.2
	α2= 0.0000	0.0632		
	α3= 0.0000	0.0729		
	α4= 0.0326	0.0479		
	α5= 0.0000	0.0327		
TCS ARCH (1)	α1= 1.0000	0.0659	42708.4	42726.4
HDFCBANK ARCH(4)	α1= 0.8222	0.1340	40943.6	40979.6
	α2= 0.0507	0.2300		
	α3= 0.0430	0.1110		
	α4= 0.0809	0.1020		
INFOSYS ARCH(5)	α1= 1.0000	0.4130	37548.6	37590.5
	α2= 2.5214e-11	0.0797		
	α3= 0.0000	0.8150		
	α4= 0.0000	0.3930		
	α5= 0.0000	0.0888		
HINDUNILVR ARCH(4)	α1= 0.9290	0.0567	43099.4	43135.4
	α2= 0.0293	0.0645		
	α3= 0.0219	0.0651		
	α4= 0.0187	0.0409		
ICICUBANK ARCH(1)	α1= 1.0000	0.0071	34008.3	34026.3
HDFC ARCH(5)	α1= 0.9245	0.1510	43960.1	44002.1
	α2= 3.5885e-15	0.1750		
	α3= 3.7884e-16	0.0480		
	α4= 0.0755	0.0547		
	α5= 3.2445e-14	0.0290		
ITC ARCH(2)	α1= 0.9798	0.0466	27892.1	27916.1
	α2= 0.0000	0.0434		
SBIN ARCH(2)	α1= 0.8064	0.0870	30502.2	30526.2
	α2= 0.1925	0.0931		
BHARTIARTL ARCH(2)	α1= 0.9519	0.0600	31275.8	31299.8
	α2= 0.0460	0.0604		

Table 5.4: ARCH Model Results

6. Result and Discussion

We observed that the share prices have increasing trend. In the period of pandemic, they show decreasing trend, however; there is a huge growth in share prices after the pandemic.

ARIMA models perform better as compare to the ARCH model in this context of study. Even with the long-range forecasting the ARIMA model gives the better predicted values. The most effective way to apply this method would be use ARIMA to produce forecasts for the future and then evaluate how accurate the predicted values.

The ARCH method was also considered into evaluating the data. The Analysis of the data indicated that one of the crucial elements to assess whether or not the ARCH model family would be suitable. In this study ARCH results are not satisfactory.

The top-performing NIFTY 50 companies behaved quite well during the study's focused period, with no noticeable oscillations. This finding explains why models that were more basic, such as ARIMA, performed better when trying to forecast market returns. When it comes to other markets, ARCH models typically perform better during the market's high oscillations caused by crises and other significant influences, such as a variety of other events. If there were more volatility during the time period, the outcomes would be entirely different in the context of this study.

7. Limitation and Future Scope

In stock market study and time series analysis, the ARIMA (Autoregressive Integrated Moving Average) and ARCH/GARCH (Autoregressive Conditional Heteroskedasticity/Generalized Autoregressive Conditional Heteroskedasticity) models are frequently used. However, it's critical to acknowledge their drawbacks and take into account the potential for future development. The following are some limitations and future application considerations for stock market studies using ARIMA and ARCH/GARCH models:

- Limitation

Limited Incorporation of External Factors: ARIMA models do not explicitly include external factors or economic indicators that could affect stock market behaviour, instead it focusses primarily on the historical values of the target variable. Incorporating relevant external factors could enhance the predictive power of the models.

Model Specification: For ARCH/GARCH models, the correct lag order for the autoregressive and moving average terms must be specified. It can be difficult to find the optimal lag order, and you may require to use iterative model selection techniques or more complex modelling techniques.

- Future scope

State Space Models: State space models offer a flexible method for modelling time series data, such as the Kalman filter or structural time series models. These models may incorporate both observed and unseen components, making it feasible to model stock market movements in a more advanced and accurate way.

Extensions to Higher Moments: Modelling volatility, which relates to the subsequent moment of the distribution, is the primary focus of ARCH/GARCH models. Asymmetric and fat-tailed distributions, which are frequently observed in financial data, can be captured by extensions to higher moments, such as GARCH-M or EGARCH models.

8.REFERENCE

- Afeef Mustafa, Ihsan, Anjum, Zada, H. (2018). “Forecasting Stock Prices through Univariate ARIMA Modeling”. https://www.researchgate.net/publication/368576271
- Rob J. Hyndman and George Athanasopoulos (2021), “Forecasting: Principles and Practice”. https://otext.com/fpp3/
- Granger, C. W. J. (1992). “Forecasting stock market prices: Lessons for forecasters. International Journal of Forecasting”. https://doi.org/10.1016/0169-2070(92)90003-R
- Ayodele Ariyo Adebiyi, Aderemi Adewumi, Charles Ayo (2014),”Stock Price Prediction using the ARIMA model”. https://www.researchhate.net/publication/2119224
- Ravi Summinga-Sonagadu (2019). “Risk Model Validation An Intraday VaR and ES Approach Using the Multiplicative Component GARCH”. https://www.researchgate.net/publication/
- Ratnadip Adhikari, R. K. Agrawal (2013), “An Introductory Study on Time series Modeling and Forecasting”. https://www.researchgate.net/publication/235219651
- https://www.kaggle.com/code/prashant111/arima-model-for-time-series-forecasting
- https://finance.yahoo.com/
- https://www.nseindia.com/
- Jason Brownlee (January 13, 2017), ‘A Gentle Introduction to the Box-Jenkins Method for Time Series Forecasting’ https://machinelearningmastery.com/
- Hardikkumar Dhaduk (2021), “Stock market forecasting using Time Series Analysis With ARIMA model”. https://www.analyticsvidhya.com/blog/2021/07/
- Shumway, R. H., & Stoffer, D. S. (2011). Time Series Analysis and Its Applications With R Examples. Book. https://doi.org/10.1007/978-1-4419-7865-3

Study of stock market volatility using time series

ORIGINALITY REPORT

12%	11%	2%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	hdl.handle.net	Internet Source	5%
2	mafiadoc.com	Internet Source	1%
3	dokumen.tips	Internet Source	1%
4	Submitted to Lebanese University	Student Paper	1%
5	baixardoc.com	Internet Source	<1%
6	Submitted to National College of Ireland	Student Paper	<1%
7	www.mdpi.com	Internet Source	<1%
8	Submitted to University of Stellenbosch, South Africa	Student Paper	<1%
9	research-repository.griffith.edu.au	Internet Source	<1%