MACHINE LEARNING APPLICATIONS IN FINANCIAL MARKETS (MBA737)

Team Members

1) ***Aryaman Singhal***
2) ***Prabhu Dayal***
3) ***Avni Agarwal***
4) ***Prabuddha Singh***
5) ***Manish Sharma***

Instructor: Abhinava Tripathi

Introduction

Our Security

Asian Paints Ltd, headquartered in Mumbai, Maharashtra, India, is the country's largest paints company by market share, engaging in the manufacturing, selling, and distribution of paints, coatings, home décor products, bath fittings, and related services. The company, a global leader in the paint industry, originated in 1942 in Mumbai, founded by four friends who seized an opportunity during World War II and the Quit India Movement when paint imports were prohibited.

In 1995, Asian Paints went public, marking a significant milestone in its growth trajectory. The following year, in 1996, it became a part of Nifty 50, and in 2008, it joined the BSE Sensex. The company's manufacturing operations extend across 15 countries globally, with a substantial presence in the Indian subcontinent and the Middle East. It is also the holding company of Berger International.

Asian Paints is a versatile player in various industries and categories, including decorative paints, coatings, industrial finishing products, and home improvement. It has received numerous awards and recognitions for its commitment to quality, innovation, and sustainability. As of May 2023, the company boasts a market capitalization exceeding ₹3 lakh crores, securing its position among the top 10 paint companies by market cap.

Q2 FY2023-24 Performance at a Glance

Net Sales	Net Profit	Net Sales	Net Profit
(Standalone)	(Standalone)	(Consolidated)	(Consolidated)
₹ 7315.7 Crores	₹ 1160.3 Crores ^ 52.0%	₹ 8451.9 Crores	₹ 1205.4 Crores ^ 54.0%

Disclosures under Regulation 46 of SEBI (LODR) Regulations

DEFINING OUR CAPITALS

FINANCIAL CAPITAL

The pool of funds that we utilise, keeping in mind growth and sustainability, to maximise value creation for shareholders and other stakeholders.

₹ 15,585.6 Crores
Shareholders' funds

₹ 93.8 Crores
Debt

Page no. 62

MANUFACTURED CAPITAL

State-of-the-art manufacturing facilities empowered by cutting-edge technologies and green infrastructure, securing high levels of productivity and product excellence.

₹ 3,487.6 Crores
Property, plant and equipment

37
Manufacturing facilities (in-house + outsourced processing centres)

Page no. 70

INTELLECTUAL CAPITAL

Robust innovation platforms crafting new and innovative products with compelling value propositions in record time, demonstrating agility and execution excellence.

240+
Scientists at Research & Technology Centre

₹ 108.0 Crores
Spend on R&D

Page no. 78

HUMAN CAPITAL

Empowered and motivated employees who are crucial for enhancing customer-centricity and delivering business outcomes of strategic, operational and financial importance.

32,036
Total workforce (permanent + temporary/contractual)

₹ 23.7 Crores
Investment in Learning & Development (L&D)

Page no. 86

SOCIAL AND RELATIONSHIP CAPITAL

The relationships that give us the social license to operate, play an important role in helping us pursue our vision and stay true to our purpose in good and challenging times.

₹ 77.2 Crores
Spend on CSR activities

150,000+
Retail touchpoints

Page no. 100

NATURAL CAPITAL

Being in the paints industry, we are dependent on natural resources. Thus, being cognisant of our environmental footprint and pursuing a low-carbon strategy is critical for business success and sustainability.

62.2%
Renewable electricity consumption out of total consumption

₹ 23.4 Crores
Spend on water conservation

Page no. 112

As of Wed 15 Nov 2023, Asian Paints Share price is ₹3,091.80 which is ₹476.20 below 52-week high of ₹3,568.00 and ₹405.95 above 52-week low of ₹2,685.85

52-WEEK HIGH LOW

- *Asian Paints 52-Week High is ₹3,568.00*
- *Asian Paints 52-Week Low is ₹2,685.85*

ALL TIME HIGH

- *The all-time high or the highest price that Asian Paints shares have ever touched was ₹3,590.00 and this occurred on Mon 10 Jan 2022.*
- *The highest closing price at which Asian Paints shares have ever closed was ₹3,576.30 recorded on Fri 07 Jan 2022.*

BSE and the SENSEX

Introduction to the Bombay Stock Exchange (BSE)

Established in 1875 under the name Native Share and Stock Brokers' Association, the Bombay Stock Exchange (BSE) holds a significant position as India's inaugural and largest securities market. Operating from Mumbai, it is a major global exchange alongside counterparts like NYSE and Nasdaq, with a listing of nearly 6,000 companies. The BSE has played a pivotal role in shaping India's corporate landscape, fostering growth in the retail debt market, and providing a dynamic platform for small-and-medium enterprises (SMEs).

The BSE implemented its electronic trading system in 1995, ensuring seamless, error-free trade execution across a diverse range of financial instruments, including equities, currencies, debt instruments, derivatives, and mutual funds. This comprehensive portfolio extends beyond trading to encompass clearing, settlement, and risk management.

At the core of BSE's performance is the Sensex, a benchmark index tracking 30 of the exchange's most influential stocks across 12 sectors. Debuting in 1986, the Sensex is India's oldest stock index, commonly referred to as "BSE 30," and serves as the pulse of the exchange, providing insights into market performance.

Dalal Street, with roots dating back to the 1850s, functions as a hub for numerous financial institutions, akin to the significance of Wall Street in the United States. The activities on Dalal Street symbolize India's financial prowess, reflecting the vibrancy of the nation's financial industry. It is the epicenter for trading various financial instruments, including stocks, stock futures, stock options, index futures, index options, and weekly options, creating a dynamic *ecosystem mirroring the dynamism of the Indian economy. As of June 2023, the BSE stands among the world's largest stock exchanges by market capitalization, boasting a substantial market cap of US$ 3.8 trillion.*

Introduction to NSE and the NIFTY-50

The National Stock Exchange of India Limited (NSE) is the largest financial market in India, established in 1992 with operations commencing in 1994. It has evolved into a sophisticated and electronic market, ranking fourth globally by equity trading volume. The NSE has been a pioneer in Indian financial markets, being the first to introduce modern, fully automated electronic trading, setting up the first electronic limit order book to trade derivatives and ETFs.

Headquartered in Mumbai, the NSE operates across the wholesale debt, equity, and derivative markets. It is renowned for its flagship index, the NIFTY 50, which tracks the largest assets in the Indian equity market. As of April 11, 2023, the NSE's total market capitalization is approximately USD 3.26 trillion, making it the ninth-largest stock exchange in the world.

The NSE's functions include creating a nationwide trading platform for equities, debt, and hybrid instruments, providing equitable access to investors across the country through a robust communication network, and ensuring a fair, efficient, and transparent securities market via electronic trading systems.

The exchange operates an order-driven market through the National Exchange for Automated Trading (NEAT), an entirely automated screen-based trading system. The NSE's market segments include the WholeSale Debt Market Division, offering trading in diverse fixed-income instruments, and the Capital Market Division, facilitating trading in securities like debentures, equity shares, exchange-traded funds, preference shares, and retail government securities.

The NIFTY 50 index, launched in 1996, is a benchmark Indian stock market index representing the weighted average of 50 of the largest Indian companies listed on the NSE. It is the world's most actively traded contract and has become the largest single financial product in India. The index covers 13 sectors of the Indian economy and is widely used by investors as a barometer of the Indian capital market. As of January 2023, the NIFTY 50 index is a free float market *capitalization-weighted index, with a base period starting on November 3, 1995, and a base value set at 1000 with a base capital of ₹2.06 trillion.*

Dividend and Stock Prices

Dividends play a significant role in shaping how stock prices move. The past dividend track record of a stock affects its overall appeal, and when a company announces and distributes dividends, it usually has a noticeable impact on stock prices. Typically, following the ex-dividend date, the stock price decreases by the same amount as the dividend.

Investors value dividends as a reliable source of income. For companies, distributing dividends is a way of sharing profits with shareholders, expressing appreciation for their support, and motivating continued investment. Additionally, dividends act as an indicator of a company's success, as only consistently profitable companies can consistently issue dividends from their earned profits.

Dividends can be provided as either cash or additional shares of stock, and the amount each investor receives is linked to their current ownership stake. Investors generally favor companies that regularly distribute dividends, even though common stock does not guarantee dividend payouts.

Before a company issues dividends, it discloses the amount, payment date, and the ex-dividend date, typically set one day before the date when the company reviews its list of shareholders. This announcement attracts investors willing to pay a premium before the ex-dividend date in anticipation of receiving dividends, often causing the stock price to rise.

On the ex-date, the stock price may decrease by the dividend amount since new investors won't qualify for dividends and are unwilling to pay a premium. However, if there is significant optimism about the stock leading up to the ex-dividend date, the price increase might surpass the automatic reduction, resulting in an overall increase.

Certain investors strategically buy shares just before the ex-dividend date and sell them shortly after the date of record to capitalize on dividend payments. Stock dividends, while not directly enhancing investor value at the time, can impact stock prices similarly to cash dividends. Nevertheless, stock dividends dilute the book value per common share, leading to a decrease in stock price.

The dividend discount model (DDM), also known as the Gordon growth model (GGM), is a valuation method based on the present value of future dividend payments. According to the DDM, a stock's value is determined by dividing the next annual dividend by the discount rate less the dividend growth rate. This model focuses on evaluating a stock's worth solely based on expected future income from dividends, assuming a regular dividend growth rate and a discount rate higher than the dividend growth rate for validity.

While the DDM is considered a prudent approach to valuing stocks, it relies on assumptions about a company's dividend payments, growth patterns, and future interest rates. To apply the DDM, three essential pieces of information are needed: the current or most recent dividend amount, the rate of growth of dividend payments, and the required rate of return. This information is typically available in a company's financial statements and historical stock information.

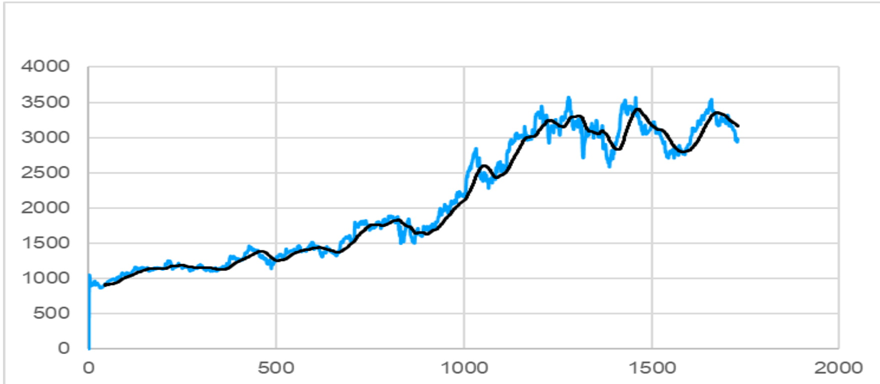
Data Analysis

We have used the following formulae to calculate Stock/Index price and return:

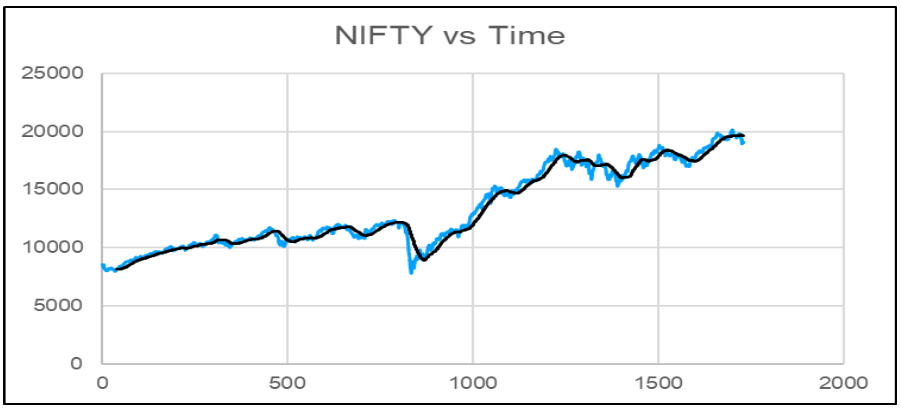
$$Stock\ Price\ =\ \frac{(Opening\ Price + Closing\ Price)}{2}$$

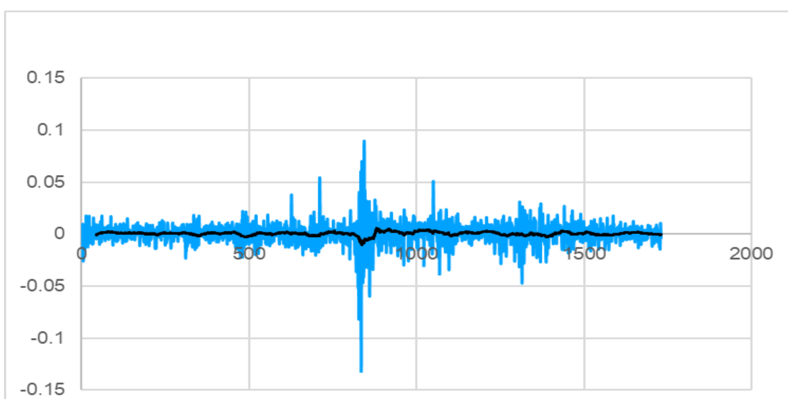
$$Stock\ Return\ =\ \frac{(Price\ Today - Price\ Yesterday)}{Price\ Yesterday}$$

Smoothed and Raw Data for Asian Paint Price vs Time

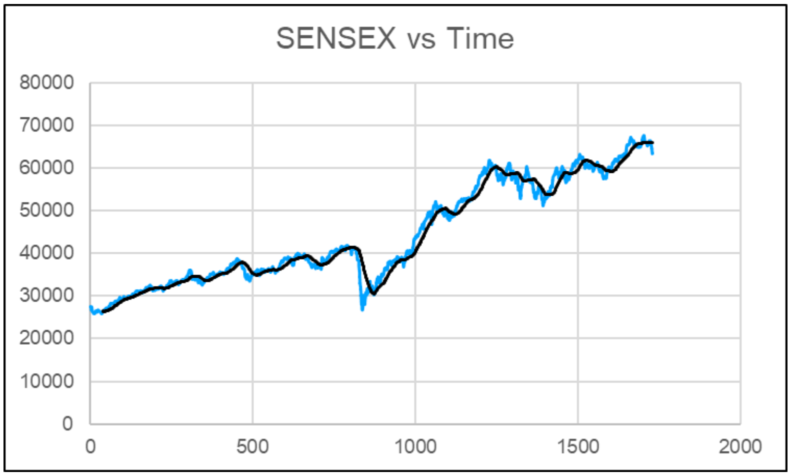


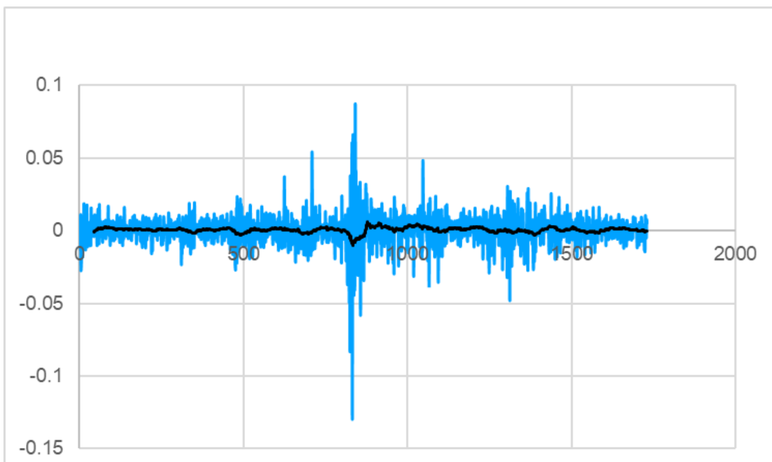
Smoothed and Raw Data for percentage change in NIFTY vs Time





Smoothed and Raw Data for percentage change in Sensex vs Time





Summary Statistics

```
> summary(Data$Asian_Paints)
      Min.    1st Qu.     Median       Mean    3rd Qu.       Max. 
-0.1402790 -0.0074890  0.0002320  0.0007391  0.0083480  0.0885270 
> summary(Data$Nifty)
      Min.    1st Qu.     Median       Mean    3rd Qu.       Max. 
-0.1298046 -0.0044504  0.0007925  0.0005323  0.0062753  0.0876321 
> summary(Data$Sensex)
      Min.    1st Qu.     Median       Mean    3rd Qu.       Max. 
-0.1315258 -0.0043817  0.0008070  0.0005493  0.0061676  0.0897490 
```

Some Financial Statistics

Stationarity:- *Stationarity is a critical concept in the realm of financial markets, influencing both their use and interpretation. In financial contexts, stationarity refers to the statistical property of a time series where key characteristics, such as mean and variance, remain constant over time. The assumption of stationarity is fundamental for many financial models and analyses, providing a foundation for making informed decisions. Time series data that exhibits stationarity is easier to analyze and model, as it allows for the application of various statistical tools and techniques with confidence in their validity. However, the financial markets are inherently dynamic and subject to various external factors, making it challenging to maintain strict stationarity. Understanding and addressing non-stationarity is crucial for accurate financial modeling and forecasting. Researchers and analysts often employ techniques such as differencing or transformation to induce stationarity in time series data, ensuring that their models are robust and reflective of the underlying patterns. Stationarity, or the lack thereof, significantly shapes how financial data is interpreted, affecting investment strategies, risk assessments, and overall decision-making processes in the dynamic landscape of financial markets.*

KPSS Test:- *The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test is a pivotal tool in the realm of financial markets, employed to assess the stationarity of* *time series data. Unlike tests that assume stationarity as a null hypothesis, the KPSS test explicitly tests for non-stationarity, making it particularly relevant in the context of financial markets where data often exhibits trends and structural breaks. This test is valuable for analysts and researchers seeking to understand the long-term behavior of financial variables. In financial markets, where the assumption of stationarity is critical for reliable modeling, the KPSS test helps identify whether a time series is stationary around a deterministic trend or if it has a unit root, indicating non-stationarity. The interpretation of the KPSS test involves comparing the test statistic to critical values. If the test statistic exceeds these critical values, the null hypothesis of stationarity is rejected, signaling the presence of a unit root and non-stationarity. This information is crucial for refining financial models, as non-stationarity can introduce spurious relationships and undermine the validity of statistical inferences, ultimately influencing investment strategies and risk management decisions in the dynamic landscape of financial markets.*

Kurtosis:- *Kurtosis, in the context of financial markets, is a statistical measure that characterizes the shape of the probability distribution of a financial asset's returns. It provides insights into the tails of the distribution, indicating the likelihood of extreme outcomes or outliers. High kurtosis suggests fatter tails, meaning that the probability of extreme events is higher than what would be expected in a normal distribution. In financial markets, understanding kurtosis is crucial for risk assessment and portfolio management. A distribution with high kurtosis indicates that extreme events, such as market crashes or rapid price movements, are more probable, emphasizing the need for robust risk management strategies. On the other hand, low kurtosis implies a distribution with thinner tails, suggesting that extreme events are less likely. Analysts and investors use kurtosis alongside other statistical measures to evaluate the risk-return profile of assets and portfolios. It provides a nuanced perspective on the distribution of returns, allowing for a more comprehensive assessment of potential risks and rewards in the dynamic and uncertain environment of financial markets. The Kurtosis of a Normal distribution is considered to be 3, anything above that is considered to be excess kurtosis, and the distribution is said to be leptokurtic.*

Skewness:- *Skewness, a statistical measure assessing the asymmetry of a probability distribution, is a vital tool in the analysis of financial markets. In the financial context, skewness helps quantify the extent and direction of* *asymmetry in the distribution of asset returns. Positive skewness indicates a distribution with a longer right tail, suggesting a higher probability of extreme positive returns. Conversely, negative skewness implies a longer left tail, indicating a higher likelihood of extreme negative returns. Understanding skewness is crucial for investors and analysts as it provides insights into the potential risks and rewards associated with an investment. In financial markets, where the occurrence of extreme events can significantly impact portfolio performance, skewness complements other risk measures, such as standard deviation and kurtosis, to offer a more nuanced view of the return distribution. Investors can use skewness to adjust their risk tolerance and optimize portfolios based on their preferences for avoiding or embracing asymmetry in return distributions. As a result, skewness serves as a valuable metric for crafting well-informed investment strategies in the dynamic and unpredictable landscape of financial markets.*

Mean, median and Standard deviation:- *In the realm of financial markets, the trio of mean, median, and standard deviation constitutes fundamental statistical measures crucial for understanding and managing risk. The mean, or average, provides a central tendency, summarizing the expected return on an investment. However, in situations where extreme values can distort the average, the median offers a more robust measure of central tendency, as it represents the middle value in a dataset. Standard deviation, on the other hand, gauges the dispersion of returns around the mean, offering insights into the volatility of an asset. Investors and analysts use these measures collectively to assess risk and make informed decisions. A higher standard deviation indicates greater price volatility, signifying increased risk, while a lower standard deviation suggests more stable returns. The interplay of mean, median, and standard deviation allows for a comprehensive understanding of the distribution of financial returns. This statistical toolkit aids in portfolio construction, risk management, and the formulation of investment strategies, contributing to a more nuanced and informed approach in navigating the uncertainties of financial markets.*

Multicollinearity: *Multicollinearity, a phenomenon in statistical analysis, is particularly relevant and consequential in the context of financial markets. It occurs when independent variables in a regression model are highly correlated, making it challenging to disentangle their individual effects on the dependent variable. In financial modeling, multicollinearity can lead to inflated standard errors and imprecise coefficient estimates, reducing the reliability of the model. This complicates the interpretation of relationships between variables and hinders the identification of key drivers influencing financial outcomes.*

Investors and analysts rely on regression models to understand the complex interplay of various factors impacting asset prices, returns, and other financial metrics. High multicollinearity can distort these models, impeding the ability to make accurate predictions and informed investment decisions. Diagnostic tools such as variance inflation factors (VIF) are commonly used to detect and mitigate multicollinearity, allowing analysts to refine models and enhance their effectiveness in capturing the nuances of financial market dynamics. Addressing multicollinearity is crucial for maintaining the integrity of financial models, ultimately supporting more accurate risk assessments, portfolio management, and strategic decision-making within the intricate landscape of financial markets.

Correlation: *Correlation, a key statistical measure, plays a crucial role in the analysis and interpretation of financial markets. It quantifies the degree of association between two variables, offering insights into their linear relationship. In financial markets, correlation is widely used to understand the co-movements of different assets, helping investors and analysts diversify portfolios effectively. A correlation coefficient ranges from -1 to 1: a value of 1 indicates a perfect positive correlation, -1 signifies a perfect negative correlation, and 0 suggests no linear relationship. Understanding correlations is vital for constructing well-balanced portfolios that can potentially mitigate risk through the inclusion of assets with diverse performance patterns. However, it's important to note that correlation does not imply causation, and the financial landscape can be influenced by various factors beyond linear relationships. Additionally, correlations can change over time due to shifting market conditions. Continuous monitoring and adaptation to evolving correlations are essential for investors to make informed decisions, manage risk, and navigate the complex dynamics of financial markets successfully.*

Normality: *Normality, a fundamental assumption in statistical analysis, holds particular significance in the interpretation of financial markets data. In the context of financial modeling, normality refers to the distribution of residuals, or the differences between observed and predicted values in a regression model. The assumption of normality is crucial because many statistical tests and methods rely on the assumption that residuals follow a normal distribution. Deviations from normality can signal potential issues with the model, such as misspecification or the presence of outliers, impacting the reliability of statistical inferences. In financial markets, where accurate predictions and risk assessments are paramount, adhering to the normality assumption ensures the validity of various modeling techniques. Analysts often employ diagnostic tests, such as the Shapiro-Wilk test or visual inspection of* *residual plots, to assess the normality of residuals. Ensuring that residuals are approximately normally distributed enhances the robustness of financial models, providing a solid foundation for making informed decisions regarding investment strategies, risk management, and other critical aspects within the intricate and dynamic landscape of financial markets.*

Autocorrelation: *Autocorrelation, a statistical concept, is highly relevant in the analysis of financial markets, particularly in understanding the persistence of trends or patterns in time series data. Autocorrelation measures the degree to which a variable is correlated with itself over different time intervals. In financial markets, where the concept of momentum and trend-following strategies is prevalent, autocorrelation helps identify whether past price movements can predict future ones. Positive autocorrelation implies that past values positively influence future values, while negative autocorrelation suggests an inverse relationship. Traders and analysts use autocorrelation to refine their understanding of market dynamics and optimize trading strategies. However, it's important to note that autocorrelation doesn't necessarily imply causation, and other factors may influence market movements. Identifying and interpreting autocorrelation patterns aids in predicting potential turning points, market reversals, or continuation of trends, contributing to more informed decision-making in the dynamic and complex environment of financial markets. Monitoring autocorrelation is a valuable tool for traders and investors seeking to uncover patterns and exploit potential opportunities in the ever-changing landscape of financial markets.*

Null-Hypothesis: *In financial markets, the null hypothesis is a foundational concept used in statistical testing to evaluate the validity of assumptions or assertions about market phenomena. It often involves statements about the absence of an effect or the equality of parameters. For instance, when testing the efficacy of a trading strategy or the impact of a financial policy, the null hypothesis might state that there is no significant difference or effect. Analysts then use statistical tests to either accept or reject this null hypothesis based on the observed data. The interpretation of results is crucial in guiding decision-making. A failure to reject the null hypothesis does not prove its truth but suggests a lack of evidence to the contrary. On the other hand, rejecting the null hypothesis implies that the observed data provide sufficient evidence to challenge the assumption of no effect. The proper formulation and testing of null hypotheses in financial research contribute to the robustness and reliability of findings, guiding investors, policymakers, and analysts in making well-informed decisions within the complex and uncertain dynamics of financial markets.*

Note: In this report, the critical p-value is taken to be 0.05.

Market Statistics:Normality

```
> jarque.test(Data$Asian_Paints)

        Jarque-Bera Normality Test

data:  Data$Asian_Paints
JB = 2946.2, p-value < 2.2e-16
alternative hypothesis: greater
```

The above is Jarque-Bera Normality Test for Asian Paints and below is for Nifty-50 and SENSEX

```
        Jarque-Bera Normality Test

data:  Data$Sensex
JB = 32380, p-value < 2.2e-16
alternative hypothesis: greater
```

```
        Jarque-Bera Normality Test

data:  Data$Nifty
JB = 30353, p-value < 2.2e-16
alternative hypothesis: greater
```

Inferences: *Jarque-Bera test has the null hypothesis that the data are sampled from a Gaussian distribution. In this case we have a small p-value, and we can reject the Null, and conclude data is not drawn from a Normal distribution.*

Market Statistics: Autocorrelation

```
> ####Autocorrelation###############
> #Durbin watson
> dwtest(Slr)

	Durbin-Watson test

data:  Slr
DW = 2.1035, p-value = 0.9656
alternative hypothesis: true autocorrelation is greater than 0

> ###Breusch-Godfrey Test
> bgtest(Slr)

	Breusch-Godfrey test for serial correlation of order up to 1

data:  Slr
LM test = 3.3586, df = 1, p-value = 0.06686
```

Inferences: *The Durbin-Watson test tests the null hypothesis that linear regression residuals of time series data are uncorrelated, against the alternative hypothesis that autocorrelation exists.*

With such a high p-value (0.97), we can not reject the Null Hypothesis, and thus the data series comes out to be uncorrelated.

Market Statistics: Beta

The Calculation for Beta Is As Follows:

$$\text{Beta coefficient}(\beta) = \frac{\text{Covariance}(R_e, R_m)}{\text{Variance}(R_m)}$$

where:

R_e = the return on an individual stock

R_m = the return on the overall market

Covariance = how changes in a stock's returns are related to changes in the market's returns

Variance = how far the market's data points spread out from their average value

For our data we have the following statistics:

1) *Co-Variance of Asian Paints Returns with SENSEX: 1.64063E-05*
2) *Co-Variance of Asian Paints Returns with NIFTY: 1.92429E-05*
3) *Variance of BSE SENSEX Returns: 0.000120715*
4) *Variance of NIFTY-50 Returns: 0.0001185*
5) *Variance of Asian Paints Returns: 0.000257619*

Market beta with SENSEX: 0.135909122

Market beta with NIFTY: 0.162387761

Generally the market beta being less than 1 reflects the stability of stock, however in this case, it speaks more about the low correlation of Asian Paints stock with the major indices.

Market Statistics:Stationarity

```
> ###KPSS test H0: test of  stationarity
> summary(ur.kpss(Data$Asian_Paints)) #Fail to reject null: Data Staionary

#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 8 lags.

Value of test-statistic is: 0.0831

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
> summary(ur.kpss(Data$Nifty)) #Fail to reject null: Data Staionary

#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 8 lags.

Value of test-statistic is: 0.0432

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

```
> summary(ur.kpss(Data$Sensex))

#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 8 lags.

Value of test-statistic is: 0.0397

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

Inferences: *The following statistics for stationarity have critical values, even at 10% significance levels as 0.347 for our data set. We fail to reject the Null Hypothesis (H0) for all three market basic market statistics (BSE SENSEX, NIFTY-50, and ASIAN PAINTS stock/index returns). Thus we can conclude that the time series is stationary for all the 3 returns.*

CORRELATION MATRIX

	Asian_Paints	Sensex	Nifty	Sentiment
Asian_Paints	1.0000000	0.1260345	0.5227928	0.3319063
Sensex	0.1260345	1.0000000	0.2041098	0.2291688
Nifty	0.5227928	0.2041098	1.0000000	0.6373280
Sentiment	0.3319063	0.2291688	0.6373280	1.0000000

```
####correlation across these variables seems to be pretty high
##A more direct measure of multicollinearty is variance inflation factor VIF
vif(Mlr)
      Sensex       Sentiment          Nifty DividendAnnounced
    1.062421        1.713727       1.694218          1.000641
```

Inferences: *The VIF-test here, shows low levels of correlation between the various variables, thus in the coming sections we can use all the variables without worrying about multicollinearity.*

REGRESSION ANALYSIS :NIFTY PRICES

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.955666588
R Square	0.913298627
Adjusted R Square	0.913248424
Standard Error	251.4646459
Observations	1729

ANOVA

	df	SS	MS	F
Regression	1	1150358057	1150358057	18191.94644
Residual	1727	109205926.4	63234.46812	
Total	1728	1259563984		

	Coefficients	Standard Error	t Stat	P-value
Intercept	-1032.065468	23.88412348	-43.21136041	3.6504E-277
X Variable 1	0.234127234	0.001735851	134.8775239	0

REGRESSION ANALYSIS: NIFTY RETURNS

SUMMARY OUTPUT

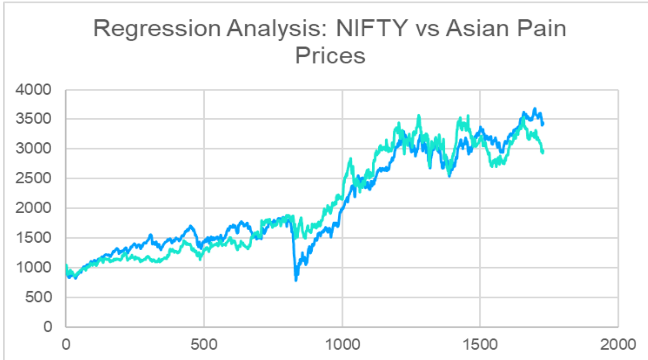
Regression Statistics	
Multiple R	0.160890818
R Square	0.025885855
Adjusted R Square	0.025321805
Standard Error	0.010847047
Observations	1729

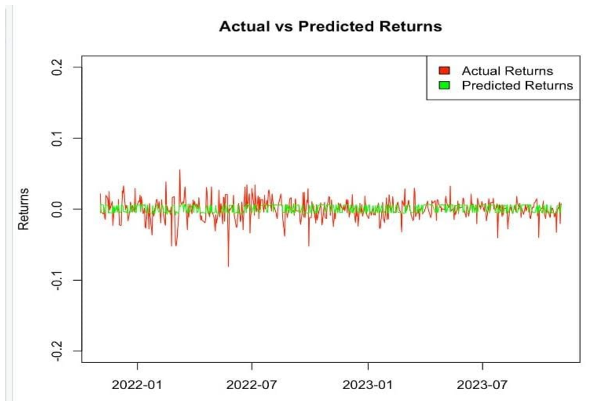
ANOVA

	df	SS	MS	F	Significance F
Regression	1	0.00539968	0.00539968	45.89284754	1.70733E-11
Residual	1727	0.20319609	0.000117658		
Total	1728	0.20859577			

	Coefficients	Standard Error	t Stat	P-value
Intercept	0.000462852	0.000261176	1.772186136	0.076539941
X Variable 1	0.162387761	0.023970704	6.774425993	1.70733E-11

NIFTY VISUAL PLOTS :Linear Regression





***Inferences**: In the above regression analysis, we have compared two charts, first one is the Asian Paints price vs NIFTY Price regression analysis, and the second one is the Asian Paints returns vs NIFTY Returns analysis, which was originally specified in the problem statement.*

The Asian Paints price vs NIFTY price analysis was done outside of the scope of the statement, despite an excellent R-squared score of 95%, the model is not very useful due to its failure in predicting daily prices, and high error range.

The Asian Paints daily return vs NIFTY Returns is the real analysis, despite its poor performance in R-squared score (2.5%), the NIFTY variable has a p-value of order of 1e-11, indicating it cannot be rejected. The intercept however had a p-value greater than 0.05, indicating its not statistically significant from 0.

REGRESSION ANALYSIS: SENSEX PRICES

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.958284394
R Square	0.91830898
Adj R Square	0.918261677
Standard Error	244.0906303
Observations	1729

ANOVA

	df	SS	MS	F
Regression	1	1156668917	1156668917	19413.63442
Residual	1727	102895067.2	59580.23579	
Total 1728	1259563984			

	Coefficients	Standard Error	t Stat	P-value
Intercept	-951.3702601	22.56476828	-42.16175625	1.0853E-267
X Variable 1	0.068305212	0.000490231	139.3328189	0

REGRESSION ANALYSIS: SENSEX RETURNS

SUMMARY OUTPUT

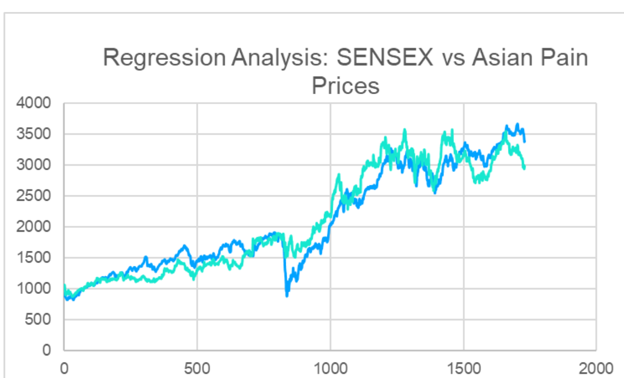
Regression Statistics	
Multiple R	0.093033689
R Square	0.008655267
Adj R Square	0.00808124
Standard Error	0.015985539
Observations	1729

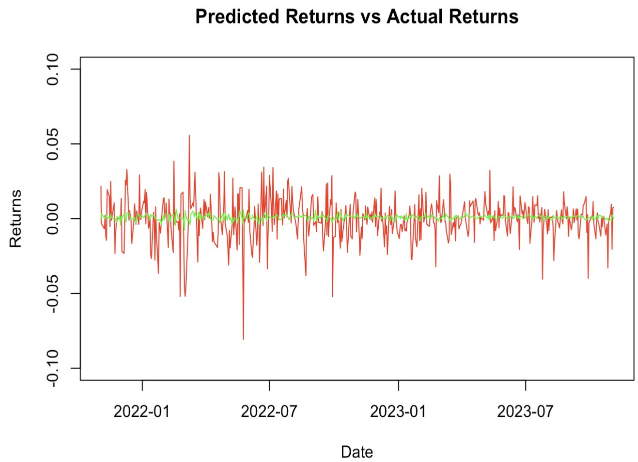
ANOVA

	df	SS	MS	F	Significance F
Regression	1	0.003853033	0.003853033	15.07815202	0.000107036
Residual	1727	0.441313215	0.000255537		
Total 1728	0.445166248				

	Coefficients	Standard Error	t Stat	P-value
Intercept	0.000664452	0.000384922	1.726200203	0.084490251
X Variable 1	0.135909122	0.035000524	3.883059621	0.000107036

SENSEX VISUAL PLOTS: Linear Regression





***Inferences**: In the above regression analysis, we have compared two charts, first one is the Asian Paints price vs SENSEX Price regression analysis, and the second one is the Asian Paints returns vs SENSEX Returns analysis, which was originally specified in the problem statement.*

The Asian Paints price vs SENSEX price analysis was done outside of the scope of the statement, despite an excellent R-squared score of 92%, the model is not very useful due to its failure in predicting daily prices, and high error range.

The Asian Paints daily return vs SENSEX Returns is the real analysis, despite its poor performance in R-squared score, the SENSEX variable has a p-value of order of approx 1e-3, indicating it cannot be rejected at any significance level. The intercept however had a p-value greater than 0.05, indicating its not statistically significant from 0.

MLR SUMMARY (without NIFTY)

```
Residuals:
      Min        1Q    Median        3Q       Max 
-0.131072 -0.008585 -0.000988  0.007161  0.089930 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)       -0.0049847  0.0006790  -7.341 3.85e-13 ***
Sensex             0.0755317  0.0396347   1.906   0.0569 .  
Sentiment          0.0107242  0.0009259  11.583  < 2e-16 ***
DividendAnnounced  0.0007123  0.0047624   0.150   0.8811    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01572 on 1228 degrees of freedom
Multiple R-squared:  0.1128,	Adjusted R-squared:  0.1106 
F-statistic: 52.05 on 3 and 1228 DF,  p-value: < 2.2e-16
```

Inferences: *Here, a multiple variable linear regression was run using the variables, Intercept, SENSEX, Sentiment, Dividend. Though we took 4 variables, the p-value of Intercept, and Sentiment only were statistically significant.*

Despite theoretical reports that Dividend leads to a spike in share prices, we did not observe this effect. This probably is due to the effect of Dividend being translated as Sentiment in the surrounding days; since the Dividend variable gains significance only on the day of release.

Another possible reason can be our calculation of daily stock price, which was taken as average of opening and closing price. This would dilute the price change should the dividend be released midway through the day.

SENSEX returns is another variable that shows high p-value, despite being statistically significant in the single variable model. Although only marginally greater than 0.05, SENSEX returns might have been translated as Market sentiment.

We have a higher R-squared value of 11%, but it still is not practically good enough.

MULTICOLLINEARITY BETWEEN NIFTY AND SENSEX PRICES

```
SUMMARY OUTPUT

NIFTY VS SENSEX

Regression Statistics
Multiple R      0.997306595
R Square        0.994620445
Adjusted R Square       0.99461733
Standard Error  878.7750204
Observations    1729

ANOVA
                df      SS      MS      F       Significance F
Regression      1       2.46581E+11     2.46581E+11     319303.2617     0
Residual        1727    1333668042      772245.5365
Total           1728    2.47914E+11

                Coefficients    Standard Error  t Stat          P-value
Intercept       -1183.146569    83.46609133     -14.17517641    3.13799E-43
X Variable 1    3.427794643     0.00606615      565.0692539     0
```

MULTICOLLINEARITY BETWEEN NIFTY AND SENSEX RETURNS

```
SUMMARY OUTPUT

Regression Statistics
Multiple R      0.160890818
R Square        0.025885855
Adjusted R Square       0.025321805
Standard Error  0.010847047
Observations    1729

ANOVA
                df      SS      MS      F       Significance F
Regression      1       0.00539968      0.00539968      45.89284754     1.70733E-11
Residual        1727    0.20319609      0.000117658
Total   1728    0.20859577

                Coefficients    Standard Error  t Stat          P-value
Intercept       0.000462852     0.000261176     1.772186136     0.076539941
X Variable 1    0.162387761     0.023970704     6.774425993     1.70733E-11
```

Inferences: *Here, we are doing a small statistical analysis on NIFTY50, and SENSEX indices. Two distinct analysis are being conducted:*

a) SENSEX and NIFTY Index value
b) SENSEX and NIFTY daily returns

When we compare the Index values, we get expected results with extremely high (99.7%) R-squared scores, however, when we compare the daily returns for these indices, we get very low (16%) R-squared scores. Although the P-value is very small, the low correlation is unexpected.

We hence proceed with NIFTY in the second Multiple Linear Regression Analysis.

MLR Summary (with NIFTY)

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       0.0006012  0.0007018   0.857    0.392
Sensex            0.0296958  0.0359829   0.825    0.409
Sentiment        -0.0001878  0.0010679  -0.176    0.860
Nifty             0.7558992  0.0458457  16.488   <2e-16 ***
DividendAnnounced 0.0003779  0.0043107   0.088    0.930
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01423 on 1227 degrees of freedom
Multiple R-squared:  0.2737,    Adjusted R-squared:  0.2714
F-statistic: 115.6 on 4 and 1227 DF,  p-value: < 2.2e-16
```

Correlation between the predicted object and actual data is 0.2937028

Inferences: *Here, a multiple variable linear regression was run using the variables, Intercept, SENSEX, Sentiment, NIFTY and Dividend. Though we took 5 variables, the p-value of NIFTY only was statistically significant here.*

Possible speculations for this behavior include:

1) *Redundancy of Information among the Sentiment, SENSEX, and NIFTY variables.*
2) *Market prices preconditioned for Dividend.*
3) *High amount of noise trading in this relatively stable stock, leading to little to no intercept*

We have a higher R-squared value among any other model of the report at 27% here.

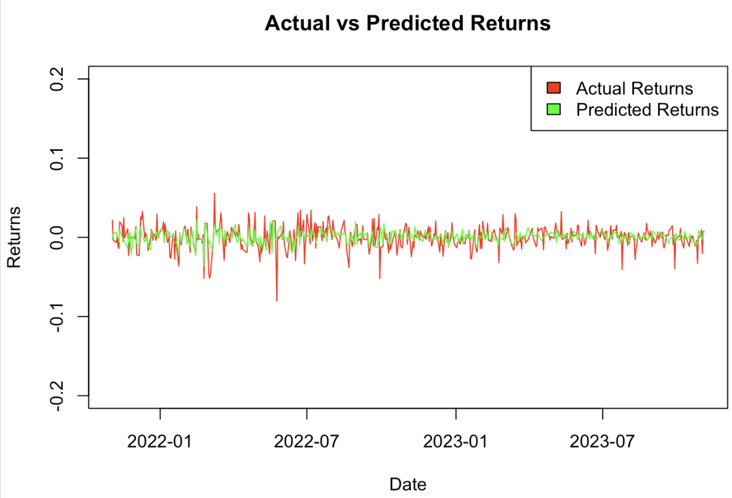
SLR2: Simple Non Linear Regression

```
lm(formula = Asian_Paints ~ Nifty + I(Nifty * Nifty), data = train)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.057649 -0.008107 -0.001070  0.007113  0.085721 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)       0.0006291  0.0004158   1.513    0.131    
Nifty             0.7459534  0.0363682  20.511   <2e-16 ***
I(Nifty * Nifty) -0.7612917  0.6427340  -1.184    0.236    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01421 on 1229 degrees of freedom
Multiple R-squared:  0.2741,	Adjusted R-squared:  0.273 
F-statistic: 232.1 on 2 and 1229 DF,  p-value: < 2.2e-16
```



Inferences: *In this regression we tried to take a NIFTY-squared term for our model, and non linearize it to check whether superior performance is shown. This is tried out assuming NIFTY-squared to be a separate variable.*

The NIFTY-squared term fails to improve our model significantly, and the variable itself fails to have a p-value less than 5%. There is a marginal improvement in the R-squared score.

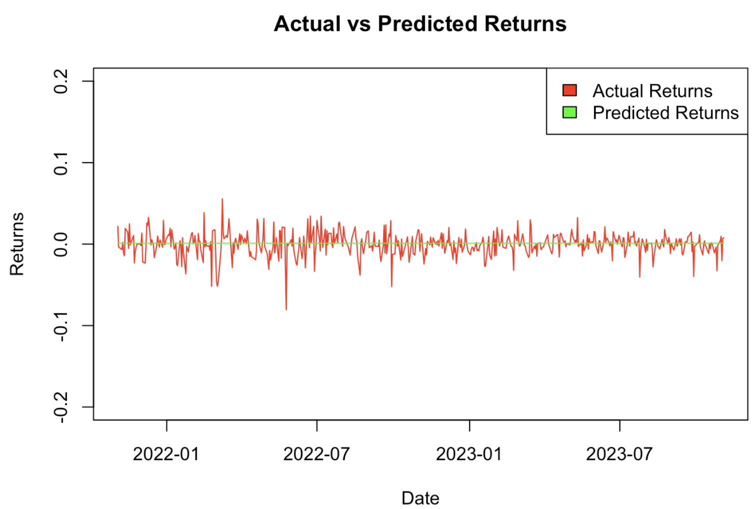
Naive Model

```
lm(formula = Asian_Paints ~ 1, data = train)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.141308 -0.008702 -0.001010  0.007777  0.087498 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)  
(Intercept) 0.0010288  0.0004749   2.166   0.0305 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01667 on 1231 degrees of freedom
```



Inferences: *In this model, we assume the returns to be normally distributed with the distribution having some finite variance, and a fixed mean (μ). Under this assumption, a natural model is to assume a return of mean (μ), since it is the mean, median and mode of the normal distribution.*

Here μ = 0.001. The small value guarantees that absurdly large returns are absent. Further it has a p-value of 0.03, meaning it is statistically significant

Though the assumption has some merit, it fails to produce good results, some variables have the same kind of randomness (E.g.: NIFTY), which this assumption fails to account for.

CONCLUSION

ERROR MODEL NAME RANK MATRIX

	MSE	RMSE	RAE	MAE	SMAPE	MSLE	RMSLE	RSE	RRSE	Complex
Naive	4	4	4	4	4	4	4	4	4	4
SLR_Mod	1	1	2	2	2	1	1	1	1	1
SLR2_Mod	2	2	1	1	3	2	2	2	2	2
MLR_Mod	3	3	3	3	1	3	3	3	3	3

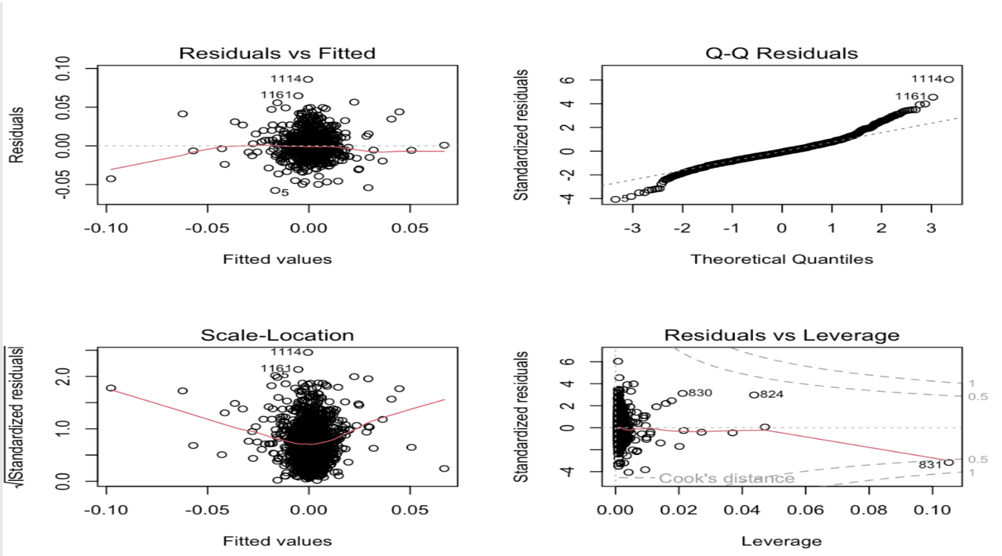
As expected, the naive model performs the worst on the test set. Its assumption is too rigid, and thus very low R-squared scores are obtained.

Multiple Linear Regression also does not perform well in the test set as it modeled the noise in the training dataset, this is not unexpected considering 4 out of 5 variables in the second multiple regression model have p-value above critical levels.

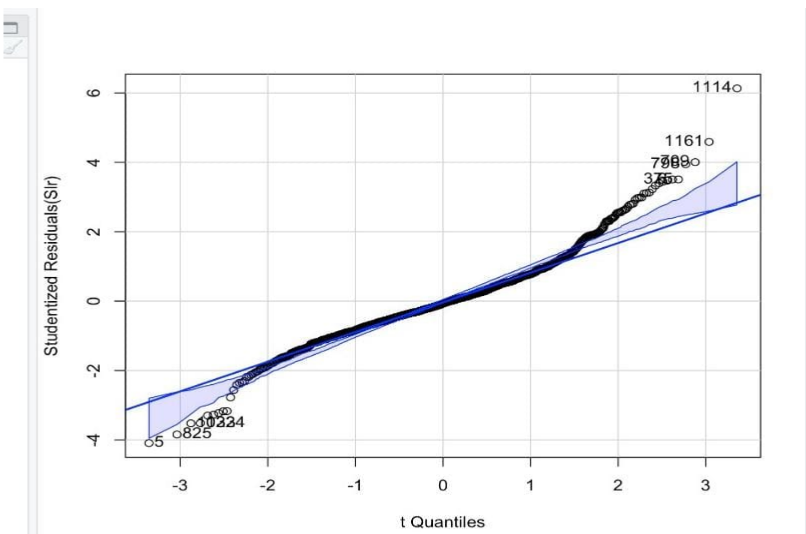
Simple Linear Regression with non-linear NIFTY Variable and Simple Linear Regression has comparable performances but the Simple Linear regression model performs the better on the testing set. Again this can be attributed to the fact that the NIFTY-squared variable was more or less redundant, and the returns did not possess nonlinear characteristics.

Simple Linear Regression model with NIFTY performed the best.

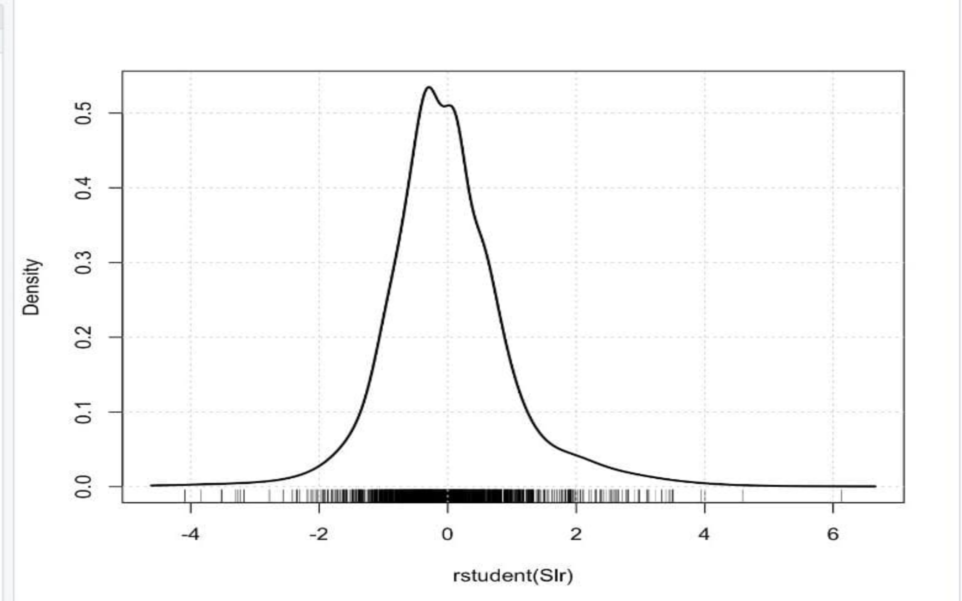
APPENDIX: Remaining Graphs



Residual vs Fitted, Q-Q Residuals, Scale Location and Residuals vs Leverage plots for SLR model to identify outliers and goodness of fit



QQ PLOT FOR NORMALITY OF RESIDUALS



Density plot of raw residuals(SLR)