THE IMPACT OF MACROECONOMIC ON STOCK RETURN: EVIDENCE FROM COMPANIES LISTED ON NASDAQ

BY
CANDIDATE NUMBER : 710067845
NAME: CIJITH JOSE
MODULE: BEMM466J

A DISSERTATION
SUBMITTED TO BUSINESS SCHOOL
UNIVERSITY OF EXETER
IN FULFILMENT OF THE REQUIREMENTS
FOR BUSINESS ANALYTICS
JANUARY 2022

EXECUTIVE SUMMARY

The purpose of this study is to investigate the impact of macroeconomic and microeconomic variables on stock returns on the Nasdaq Stock Exchange. The research aims to identify the specific variables that have the greatest effect on stock prices and to understand the nature of the relationship between these variables and stock returns. Additionally, the study aims to examine how these variables interact with macroeconomic variables to influence stock prices. The research used a combination of univariate, bivariate, and multivariate analysis techniques to examine the relationship between macroeconomic and microeconomic variables and stock prices.

The literature review highlights the mixed findings of previous studies on the impact of macroeconomic variables such as GDP, inflation, and exchange rate, on stock returns. The research questions focus on identifying the specific microeconomic variables that have the greatest effect on stock prices, including the impact of macroeconomic factors, trends in stock prices in relation to macroeconomic variables, the impact of unemployment, world oil prices, industrial production, Consumer price index, international trade, monetary and fiscal policy, national security and financial regulations, US government spending on Nasdaq stock exchange. The data has been collected for a time span of 20 years from January 2002 to November 2022 using a monthly data time series to portray a larger view of the relationship.
The study found that the distribution of all variables in the data was not normal and there were outliers present in all attributes, except for GDP, Consumer Price Index, Industrial Production Index, Oil Prices, and Trade Exports. The observed trend and pattern among these variables suggest that most of these outliers can be attributed to the impact of the COVID-19 pandemic, during which America experienced a total shutdown.

An initial evaluation of a conceptual framework model was conducted to assess the effect of independent variables on a dependent variable using multivariate analysis techniques such as scatter plots and correlation analysis. The results indicated a strong positive correlation between the Nasdaq and various macroeconomic variables, including GDP, CPI, trade exports and imports, and a negative correlation with unemployment. The analysis suggests a correlation between macroeconomic variables and stock prices, with CPI, GDP, and trade exports and imports appearing to be the most influential factors on Nasdaq pricing. However, it should be noted that this model should not be relied upon without further statistical measures and significance testing.

To improve the accuracy of our dataset and mitigate the presence of high multicollinearity as identified through correlation analysis and the Bartlett test of sphericity, a dimension reduction technique was implemented. Specifically, Principal Component Analysis was conducted, which effectively grouped our variables into just three dimensions. Additionally, scaling and outlier treatment was applied to further enhance the precision of the results. The analysis revealed that economic growth, as represented by a composite variable comprising GDP, CPI, trade exports and imports, and the Industrial Production Index, emerged as the most significant factor influencing stock prices. The model demonstrated that for every unit increase in economic growth, there is a corresponding decrease of 0.281 units in stock prices, holding all other factors constant.
Model three did well under Linear regression using Ordinary Least Square method after removing the insignificant variables (derived using p-value at significance level of 0.05). This Model shows that there is higher impact of Unemployment rate on the stock price followed by Consumer price index and Industrial production index. The model demonstrated that for every unit increase in unemployment, there is a corresponding decrease of 376 units in stock prices, holding all other factors constant.
Hypothesis testing reflected that the variables such as Oil prices, National security, Trade policy, GDP, Tax are poor predictors of Nasdaq stock prices as the p value for all these variables was higher than alpha 0.05.

Other machine learning models such as decision tree, random forest, artificial neural network was also used, and the models and hyperparameters were tuned to arrive at higher accuracy scores. GridSearchCv helped in identifying the best hyperparameters to be used. Random Forest regressor was chosen as the best model and the model predicted Consumer price index and GDP has the highest impact on Nasdaq stock prices followed by trade imports and unemployment rate.

The study recommends further research to confirm the validity of the model. It also suggests monitoring the impact of the COVID-19 pandemic on the economy. Investigating the effect of other macroeconomic variables such as interest rate and inflation on stock prices, using the Random Forest Regressor model 1 for prediction and analysis of stock prices on the Nasdaq Stock Exchange. Considering Consumer Price Index, GDP, trade exports and imports, and unemployment rate when making investment decisions in the stock market while exploring other machine learning models and techniques to investigate the effect of other global and local factors on stock prices and evaluating the results in the context of different stock markets and economies around the world.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 Purpose of the study

The stock market has long been considered a leading indicator of the economy with stock prices reflecting the overall health and performance of companies listed on the exchange. While macroeconomic variables such as GDP and inflation have been widely studied in relation to stock returns, the impact of microeconomic variables on stock prices has received less attention. Microeconomic variables refer to factors that affect individual firms and industries, such as company-specific financial performance and industry-specific trends. This research aims to investigate the impact of microeconomic variables on stock returns on the Nasdaq Stock Exchange. The results of this research will contribute to a deeper understanding of the factors that influence stock returns and inform investment decisions in the stock market.

The results of this research will provide insight into the factors that drive stock returns and inform investment decisions in the stock market. It will also add to the existing literature on the topic and potentially contribute to the development of more effective investment strategies. Furthermore, this study will be a valuable resource for financial analysts and economists as it will provide a comprehensive analysis of the factors that drive the stock market, which can be used to improve forecasting models and inform policy decisions.

1.2 Significance of the study

The problem that this study aims to address is the lack of understanding of the impact of microeconomic variables on stock returns on the Nasdaq Stock Exchange. While macroeconomic variables such as GDP and inflation have been widely studied in relation to stock returns, the effect of microeconomic variables such as government policies, tax & expenditure, political factors, financial regulation, national security etc on stock prices has received less attention. This lack of understanding of the impact of microeconomic variables on stock returns can lead to a lack of precision in predicting stock prices and difficulty in making informed investment decisions. This study aims to fill this gap in knowledge by attempting to address the question of whether the volatility of stock prices is affected by macroeconomic variables and identifying the specific microeconomic variables that have the greatest effect on stock prices.

Stock market is a platform where companies can sell shares of their company to the public to raise capital. When a company wants to raise money, it can either borrow from a bank or other financial institution, or it can sell ownership stakes in the form of stocks. The stock market allows companies to tap into a wide pool of investors, including individuals, mutual funds, and pension funds, who can provide the capital they need to grow and operate their businesses. The stock market is also important because it allows investors to buy and sell stocks, which represent ownership in a company. This provides a way for people to invest their money and potentially earn a return on their investment. Stock prices are determined by supply and demand, and they can rise or fall based on a variety of factors, including the company's financial performance, the state of the economy, and investor sentiment.

The stock market is a key component of the financial system, and it plays a vital role in the economy. It provides a way for companies to raise the capital they need to invest in their businesses, and it allows investors to share in the success of those companies. The stock market is also an important source of economic information, as the prices of stocks reflect the collective expectations of investors about the prospects of companies.

One of the things that encourages investors to invest is the stock return since it compensates them for the risks, they take on by investing their money. To lower the risk associated with ambiguity, investors want knowledge. The dynamic and unpredictable nature of the world's economies, which include a variety of capital markets, has an impact on the psychological market on the America’s Stock Exchange.

The next step in this research is to conduct a thorough literature review to understand the existing knowledge on the topic and identify gaps in the literature that this study aims to fill. The literature review will help to establish the research questions, hypotheses, and methods for this study, and provide a foundation for the analysis and interpretation of the results.

2. LITERATURE REVIEW

The performance of the stock market is often used to gauge a nation's overall economic health. The stock market typically does well when the economy is strong, and vice versa. Given that stock market performance is now a sign of a healthy economy, it is crucial to examine stock market performance from a macroeconomic perspective. According to one of the research, just 38% of the variance in the stock market can be attributed to company-specific factors. While macroeconomic factors accounted for 52% of the driving force, industry-related factors account for the remaining 10%.

Despite the fact that a large number of scientific studies have examined the connection between macroeconomic factors and the stock market, the findings still present some conflicting information regarding the variable or set of macroeconomic factors most likely to influence the stock market and the strength of that influence. A number of papers, and their findings showed a variety of conclusions, including the following: "Some researchers have denied of any relationship between macroeconomic variables and the stock market, while others have found somewhat weak links or different results. Others discovered that some variables play dual roles and affect stock performance in both negative and positive ways. Some of the earlier studies are discussed below.

2.1 Background Theories

The study conducted by Krisna & Wirawati (2013) found that inflation has a positive and significant effect on stock returns, which differs from the findings of Tangjitprom (2013) and Azar (2013) that the effect of inflation on negative stock returns is not significant. The exchange rate, often used as a reference for investors and economists, was also studied and research by Al-abdallah & Aljarayesh (2017) found it does not significantly influence the stock index, while other studies such as Amansyah (2013), Liauw (2013), Silim (2012), and Sisbintari (2009) found that it has a significant effect on stock price index. Additionally, research by Gunawan & Wibowo (2012) found that the Rupiah/US $ exchange rate has no significant effect on positive stock returns, differing from the findings of Prasetiono (2010) which found that interest rates have a negative and non-significant effect on stock return. The impact of GDP on the movement of the Stock Price Index also varied in research, with Adam et al (2013) and Sisbintari (2009) finding it has a significant effect and Kewal (2012) finding the opposite. Furthermore, Pícha (2017) found that money supply affects stock market and Suwandy (2014) found that world oil prices have a significant effect on stock return.

While the majority of articles have focused on established countries, emerging countries, notably Asia, continue to be of interest [3]. Additionally, it is apparent that not many studies have been carried out in the United States of America. India & China [5], Thailand [8], Saudi Arabia [4], Vietnam [7] and Malaysia [6]were mentioned in certain research. However, the results show that it is impossible to draw any firm conclusions about how different countries' stock exchanges would react because they appear to vary depending on the country. The fact that the country's parameters were different could have caused different results. This study's arguments have been that the inconsistent results and lack of agreement are signs of a research gap. Determining the relationship between macroeconomic variables and the American stock market (NASDAQ) as well as the effects of each variable on various sector indices is the overall goal of this study.

- *Kalyanaraman and Al-Tuwajri (2014), Arouri and Rault (2010), and Onour (2008)* took into account other economic factors as the independent variables and the Saudi stock return as the dependent variable. This study investigates the three crucial elements that affect the Saudi Stock Exchange's (TASI) returns based on the macroeconomic factors affecting the Saudi economy. The TADAWUL All Stock Index (TASI), a Saudi index, is used as the dependent variable in this study. The three independent variables are WTI oil prices, Saudi exports, and the PE ratio. Saudi exports and the PE Ratio were shown to be substantially connected with TASI at the 1% level of significance, whereas Oil WTI and TASI are considerably correlated at the 5% level, according to correlation analysis. Following Oil WTI and Saudi Exports as the top three determinants of TASI, step-wise regression analysis of the data showed that the numerous regression models are significant at the 1% level. A further 93% of the variation in the TASI Last Price can be explained by the three independent factors.
- *(Hosseini et al., 2011)* This study examines the connections between four macroeconomic indicators—the price of crude oil (COP), the money supply (M2), industrial production (IP), and inflation rate (IR) in China and India—and stock market indices. This study's time frame is from January 1999 to January 2009. The underlying series are tested as non-stationary at the level but stationary in first difference using

the Augmented Dickey-Fuller unit root test. The use of the multivariate cointegration and vector error correction models developed by Johansen and Juselius in 1990 shows that there are both long- and short-term relationships between macroeconomic variables and stock market indices in each of these two nations.

- *(Chee et al., 2015)* This study looks at how the Malaysian stock market responded to specific macroeconomic factors, including industrial production, inflation, money supply (M1), interest rate, and exchange rate, from 1980's first quarter to 2011's third quarter. This study established the existence of a long-run link between share prices and economic activity using the autoregressive distributed lag (ARDL) bounds test. According to the long-run coefficients, inflation and money supply have a favourable impact on Malaysian share values while negatively affecting interest rates. This analysis shows that the exchange rate is a proper predictor of stock returns with explanatory power.
- *(Forson & Janrattanagul, 2014)* Using monthly time series data that span a 20-year period from January 1990 to December 2009, this study examines and evaluates the long-run equilibrium relationship between the Thai stock exchange index (SETI) and a few macroeconomic variables. This study takes into account the money supply (MS), consumer price index (CPI), interest rate (IR), and industrial production index (IP) as macroeconomic factors (as a proxy for GDP). The findings demonstrate the cointegration of the SET Index at I (1) and the long-term significance of the equilibrium link between the SET Index and the chosen macroeconomic variables. The long-term link between the money supply and the SET Index is strongly positive, but the long-term associations between the SET Index and the industrial production index and consumer price index are negative.

2.2 Hypotheses and Research Model

The variables tested in this study are Unemployment rate(U_R), Industrial Production Index(I_P_I), Real Gross domestic product (Real_GDP), Consumer Price Index(C_P_I), Trade Exports(T_E), Trade Imports(T_I), Oil Price(O_P), Monetary Policy index(M_P)Fiscal Policy Index(F_P), TAX index(TAX), Government Spendinds(GOV_S), National Security(N_S), Financial Regualtions(F_R), Trade_Policy(T_P) as independent varibales and stock price index (NASDAQ) as the dependent variable.

Considering the results of the previous research and literature reviews we have formulated the following research hypothesis:

H1: How does Unemployment rate(U_R),Industrial Production Index(I_P_I), Real Gross domestic product (Real_GDP), Consumer Price Index(C_P_I), Trade Exports(T_E), Trade Imports(T_I), Oil Price(O_P), Monetary Policy index(M_P)Fiscal Policy Index(F_P), TAX index(TAX), Government Spendings(GOV_S), National Security(N_S), Financial Regulations(F_R), Trade Policy(T_P) simultaneously affect the stock price index (NASDAQ).
H2: How Unemployment Rate(U_R) effect the stock price index(NASDAQ).
H3: How Industrial Production Index(I_P_I) effect the stock price index(NASDAQ).
H4: How Real Gross domestic product (Real_GDP) effect the stock price index(NASDAQ).
H5: How Consumer Price Index(C_P_I) effect the stock price index(NASDAQ).
H6: How Trade Exports(T_E) effect the stock price index(NASDAQ).
H7: How Trade Imports(T_I), effect the stock price index(NASDAQ).
H8: How Oil Price(O_P), effect the stock price index(NASDAQ).
H9: How Monetary Policy(M_P) effect the stock price index(NASDAQ).
H10: How Fiscal Policy(F_P) effect the stock price index(NASDAQ).
H11: How Tax (TAX) effect the stock price index(NASDAQ).
H12: How Government Spendings (GOV_S) effect the stock price index(NASDAQ).
H13: How National Security (N_S) effect the stock price index(NASDAQ).
H14: How Financial Regualtions (F_R) effect the stock price index(NASDAQ).
H15: How Trade Policy (T_P) effect the stock price index(NASDAQ)

Based on the above literature the below theoretical model is proposed :

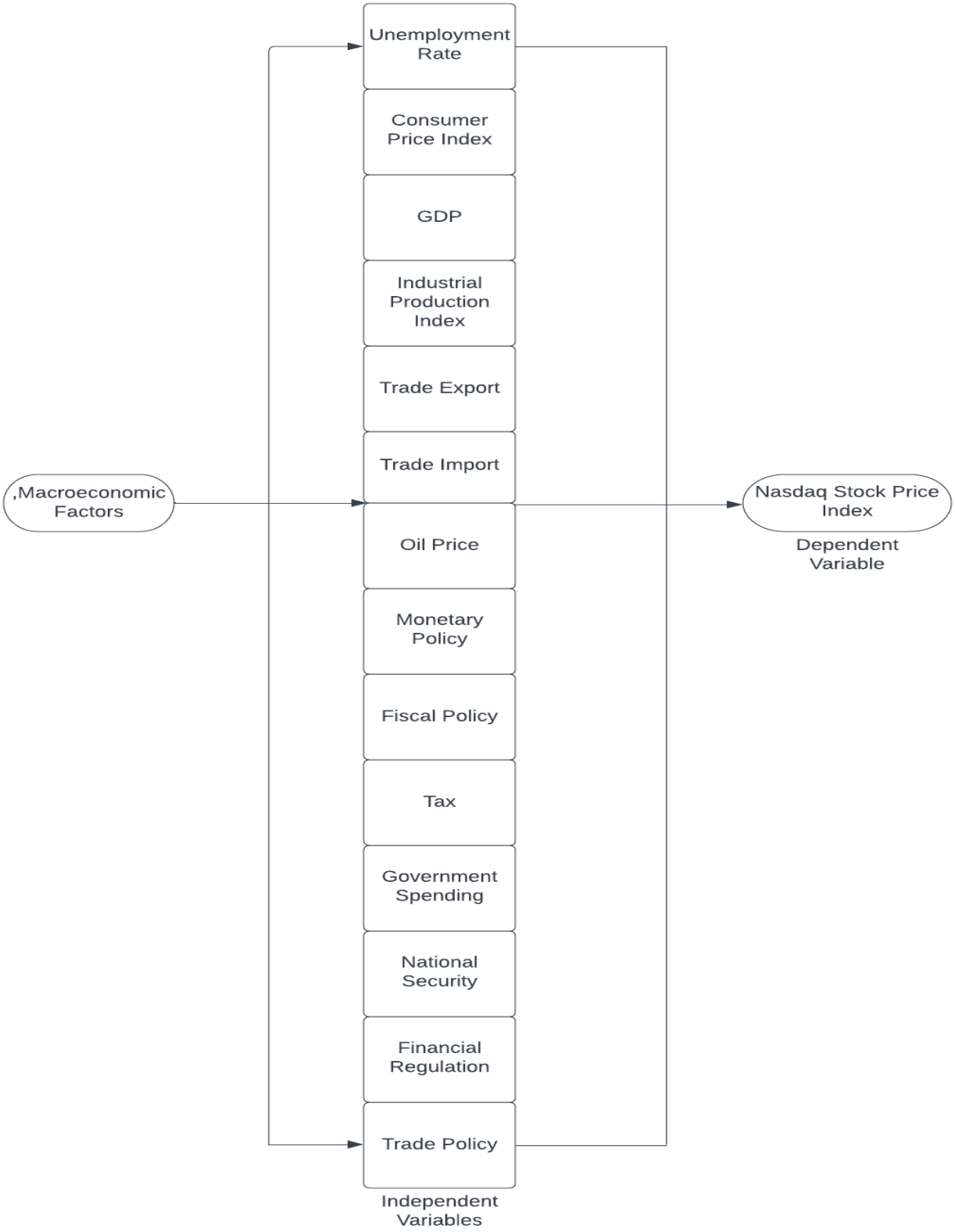


Figure 1:Conceptual Model

3. METHODOLOGY

The research's primary motive is to study the impact or influence of the macroeconomic variables on the stock market price. The specific aim of the study is determine if there exists a relationship and if it does then which factors are influencing the stock prices significantly. The research is important as it might provide knowledge with the help of detailed analysis and statistical insights to investors, which in turn would lower investors risk associated with ambiguity and uncertainty of macroeconomic circumstances. To answer the research questions, we will be collecting the data on stock market prices and macroeconomic indicators over a specific period and apply mathematical and statistical analysis to examine the relationship between both the variables.

In this paper the aim to see influence of macroeconomic factors on Nasdaq Composite price index which is the dependent variable. To find the significant macroeconomic factors among all the macroeconomic variables, various regression model including principal component analysis will be utilized . The independent variables that are being focused on for this study are Unemployment Rate, Industrial production index, Real Gross Domestic Product, consumer price index, oil prices, monetary policy index, fiscal policy index, government spending index, national security index, trade policy index, financial regulation index, trade exports , trade imports and Tax. Correlation analysis as part of EDA will also be conducted, which would aim to determine the strength and direction of the relationships between variables.

3.1 Purpose of study

This paper examines the influence of macroeconomic factors in United States of America on US stock Exchange, which in this study is NASDAQ composite index (consisting of 3700 listed companies). The study will try to understand if there exists any correlations between the macroeconomic indicators that influences the increase or decrease in Nasdaq stock prices. This study will explore potential factors associated with the upward or downward trend in stock prices over the years. Ultimately the results and analysis will provide detailed insights into the role of macroeconomic variables which are specific to US market and help investors and companies of US and around the world to take informed decisions before investing their money on Nasdaq stocks.

3.2 Research Questions

1. Does macroeconomic factor influence the drops and rise of Nasdaq stock prices?
2. What is the trend of Nasdaq's stock prices in comparison with the macroeconomic variables?
3. Which macroeconomic variables are the most significant indicator of stock returns?
4. Does the increase in the unemployment of US residents have a direct impact on the stock returns?
5. Does world oil prices effect the stock returns?
6. Is there a dependency of stock return based on industrial productions in the United States of America?
7. Does the Consumer price index effect the stock returns?
8. Is there an impact on stock return prices based on the increase or decrease of united states international trade exports & imports?
9. Does the monetary or fiscal policies have any positive or negative impact on stock prices?
10. Does the National security or Financial regulations of US have an impact on stock exchange market?
11. Does the US Government spendings have an influence on Nasdaq stock exchange?

3.3 Research framework

1. An extensive research: on the existing data of macroeconomic variables and stock market to be considered. This involves review of past literature work on the relationship between the independent and dependent variables. It may include examining articles from academic journals, government reports and research papers to gather maximum information on the influence of macroeconomic variables on stock market price.
2. Data collection: of desired stock prices and macroeconomic variables such as data on unemployment rate , industrial production rate, consumer price index, oil prices of the desired country of study. The data can be collected through various sources such as government databases, yahoo finance, Fred , world bank etc.

3. Analyse the data: once the data is extracted it will then be analysed to understand the relationship between the independent variables and dependent variable, through different statistical methods and techniques such as Linear regression, ANN, Random Forest, PCA correlation matrices etc.
4. Interpreting the results: After the data analysis has been done , the insights would then need to be interpreted to get the answers to our research questions. This would involve statistical significance of the results.
5. Research Report: Lastly , the results of the analysis will need to be written in a research report with the detailed findings and their implications for understanding the relationship between the independent variables and dependent variable.

3.4 Research Methodology

The research was conducted using quantitative research methods aimed at answering the our research questions of the presence of an impact of macroeconomic factors on stock price index and which macroeconomic variables have a significance influence on the stock price index by using mathematical models and statistical models.
The models used are as below :

- Correlation Analysis
- Principal Component Analysis
- Linear Regression
- Decision Tree Regression
- Random Forest Regression
- Artificial Neural Networks

ANOVA was performed to evaluate the statistical significance of the variables and incorporated Cross Validation techniques in models to tune and identify the best hyperparameters for optimizing our results and find the best model.

3.5 MACHINE LEARNING MODELS

ANOVA

The formal definition of analysis of variance (ANOVA): ANOVA is a statistical technique that assumes that the observed response is coming from more than one population and tests the hypothesis that at least one population mean is different from the rest. The basic concept of ANOVA is to separate the total variability in a dataset into two types; the variability that can be attributed to specified causes and the variation that can be attributed to chance or error
This technique is part of the domain called “Experimental Designs”. This helps in establishing in a precise fashion the Cause -Effect relation amongst variables. From the Statistical Inference Point of View, ANOVA is an extension of independent t test for testing the equality of two population means (Kenton, 2021).

Principal Component Analysis

The concept of principal components is quite intuitive. Instead of dealing with a large number of possibly correlated variables, principal components are constructed as suitable linear combination of the observed variables such that the components have two important properties:

- The principal components (PCs) carry the total variance present in the data•
- The PCs are orthogonal, i.e. uncorrelated, to one another.

Information content in the data is determined by the variance of the attributes. A random variable whose variance is 0, is completely non-informative because for each unit this variable has the same value; in other words, this is a constant. Reduction of dimension involves sacrificing certain amount of variance. A balance must be struck so that significant reduction in the number of dimensions is achieved by sacrificing the least possible amount of variance (Sartorius, 2020).

Linear Regression

Linear Regression falls under the supervised learning algorithm. The input and the output is connected based on a linear relationship. The independent variable is usually called y and the dependent variable x and then a connection is established between them. Linear Regression fits the given data into a straight line, predicting the most suitable value with minimum variation from the actual value. If a single independent variable is to be predicted from the model, then it is termed as Simple Linear Regression. If many independent variables are to be predicted from the given data, then it is called Multiple Linear Regression.

Consider data $D = \{(y_i,x_i) : i = 1,..., n\}$, where yi is the ith response, measured on a continuous scale; $x_i = (x_{i1},...,x_{ip})^t \in R^p$ is the associated predictor vector; and n (>>p) is the sample size. The linear model is specified as

$yi= \beta0+ \beta1_x i1+ \cdots +\beta_{pxip}+ \varepsilon_i$ with $\varepsilon_i \sim N(0,\sigma^2)$, for i = 1,...,n. (Lehmann & Scheffé, 1950)

Decision Tree Regression

It is possible to represent decisions and all of their potential outcomes using a decision-making tool called a decision tree. A decision or classification problem is represented by a tree structure. This type of regression is famous for its efficiency and simple nature in handling a large set of variables. However, it can often lead to incorrect decisions for small sample spaces. Regression trees are a type of additive model of the form:

$m(x) = \sum^{l}_{i=1} k_i \times I(x \in D_i)$
where,
k_i are constants;
I(.) is an indicator function which returns 1 if its argument is true and 0 false;
and Di are disjoint partitions of the training data D such that $Ul_{i=1} D_i = D$ and $\bigcap^{l}_{i=1.} D_i = \Phi$.
(Hastie & Tibshirani, 1990)

Random Forest Regression

A supervised learning technique called Random Forest Regression leverages the ensemble learning approach for regression. The ensemble learning method combines predictions from various machine learning algorithms to provide predictions that are more accurate than those from a single model. (Bakshi, 2020)
A random forest is a collection of tree predictors $h(x;\theta_k)$, k = 1,...,K where x represents the observed input vector of length p with associated random vector X and the θk are independent and identically distributed (iid) random vectors. The observed (training) data is assumed to be independently drawn from the joint distribution of (X,Y) and comprises n(p+1) - tuples $(x_1 ,y_1),....,(x_n ,y_n)$.
The random forest prediction for regression is the unweighted average of the series:

$h(x) = (1/K)\sum^{K}_{k=1} h(x;\theta_k)$.
As k→∞, the Law of Large Numbers ensures,
$E_{X,Y}(Y-ħ(X))\ 2\rightarrow E_{X,Y}\,(Y – E_\theta h(X;\theta))^2$ (Segal, 2004)

ANN MODEL

A Machine Learning algorithm that is roughly modelled around what is currently known about how the human brain functions

Representation of a neuron

$$y(x) = f\left(\sum_{i=1}^{n} w_i x_i\right)$$

Representation of a neuron:

•w -weights
•n – number of inputs
•xi – input
•f(x) – activation function
•y(x) – output axon

3.6 MODEL EVALUATION METHOD

The various models used have to be checked for their efficiency. A variety of errors and tests are used for this process. This is done in order to find the right model for the given data.

Residuals

Residuals are the difference between training data and fitted values. Forecast error is defined as the difference between test data and predicted values. Residuals are used to evaluate the performance of the model, while errors are used to evaluate the accuracy of the future forecast.

R2 Score

An R2 score evaluates how well a model fits the data. The R2 coefficient of determination in regression is a statistical indicator of how closely the regression predictions match the actual data points. When the R2 value is 1, the regression's predictions accurately reflect the data (Casella & Berger, 2002).

MAPE

In statistics, the mean absolute percentage error (MAPE) is a measure of forecasting method accuracy. MAPE = 100% $n\ t$=1 $n \sum A\ t - F\ t\ A\ t$ |||||| where At is the value of the actual series and Ft is the value of the forecasted series. Ft is subtracted from At and this value is then divided by the actual value At . For each predicted time point, this ratio's absolute value is added, then divided by the n fitted points (Aiolli et al., 2016).

Root Mean Square Error

The Root Mean Square Error (RMSE) is found out by calculating the standard deviation of the residuals. The RMSE represents the degree of dispersion of these residuals. In other words, it provides information on how tightly the data is clustered around the line of best fit. RMSE = 1 $n\ i$=1 $n \sum (S\ i - O\ i\ )$ 2 Where Oi are the original values and Si are the forecasted values and n is the total number of observations (Christie & Neill, 2022).

Jarque Bera Normality Test

The Jarque-Bera test in statistics determines if sample data has skewness and kurtosis that are consistent with a normal distribution. The test statistic is never negative. It indicates that the data does not have a normal distribution if it is far from zero (Jarque & Bera, 1980).

Durbin Watson Statistic :

This test is used test for autocorrelation. The null hypothesis for this test is positive autocorrelation does not exist which means residuals are not correlated. The alternate hypothesis says positive autocorrelation is present. The possible range is $0 \geq D \leq 4$. D should be close to 2 if the null hypothesis is true. D less than 2 may signal positive autocorrelation and more than 2 would mean a negative autocorrelation (Kenton, 2021).

GridSearchCV

Optimizing the performance of a model is crucial, and one way to achieve this is through the process of hyperparameter tuning. The GridSearchCV method is utilized to determine the optimal values for the given model's hyperparameters, as the performance of a model is closely tied to the values of these parameters.. However, this process can be both time-consuming and resource-intensive when done manually. To alleviate this challenge, GridSearchCV, a function from the model selection package of Scikit-learn (SK-learn) is employed to automate the tuning process, thus making it more efficient. Hence after using this function we get accuracy/loss for every combination of hyperparameters and we can choose the one with the best performance. (Mujtaba, 2020)

4. FINDINGS

The findings of this study provide a comprehensive examination of the relationship between microeconomic variables and stock prices on the Nasdaq Stock Exchange.

4.1 Data Set

The data consists of 14 independent variables and 1 dependent variable having 251 observations collected for a specific time period from 2002 January to 2022 November. The below Table describes the symbols, variables used, their units of measurements and variables used as proxy for empirical purposes in other words data dictionary. There are 14 float types variables which are the macroeconomic factors as well as the dependent variable NASDAQ stock index and 1 datetime data type which is the time period from 2002 to 2022.

Symbol	Variable	Units of Measurement
U_R	Unemployment Rate	Percentages
I_P_I	Industrial production Index	Index
Real_GDP	Real Gross Domestic Product	USD Billion
C_P_I	Consumer Price Index	Index
T_E	Trade Exports	US Dollars
T_I	Trade Imports	US Dollars
O_P	WTI Oil Prices	Cushing, OK WTI Spot Price FOB (Dollars per Barrel)
M_P	Monetary Policy	Index
F_P	Fiscal Policy	Index
TAX	Taxes	Index
GOV_S	Government Spending	Index
N_S	National Security	Index
F_R	Financial Regulations	Index
T_P	Trade Policy	Index
NASDAQ	Nasdaq stock index	Stock Index

Table 1:Units of Measurement

4.2 Data Definitions

- The Consumer Price Index (CPI) serves as the primary source of information for calculating the US inflation rate. This variable has been taken into account as both an index and an inflation rate.
- The Industrial Production Index (IPI) according to the National Statistics Institute, is a cyclical indicator that assesses the industrial sector's productive activity (excluding construction). The growth rate (seasonally adjusted) is employed as a percentage and the factor has been analysed as an index (in levels).
- Oil WTI (West Texas Intermediate), commonly referred to as Texas light sweet, is a grade of crude oil that serves as a standard for oil pricing. Because of its comparatively low density and low sulphur level, this grade is referred to as light and sweet. The underlying material for oil futures contracts on the Chicago Mercantile Exchange is oil WTI.
- Trade exports and imports, the US economy (GDP) is significantly impacted by trade and exports, just like every other economy in the globe. Similar to how the exports of one country have an impact on the stock exchanges.
- The unemployment rate represents the total number of individuals who are not working but are actively seeking employment.
- Gross Domestic Product (GDP) is a measure of a country's economic activity and health. It is the total value of all goods and services produced within a country over a given period of time, typically a year.
- Monetary Policy, Fiscal Policy, Tax, Government Spending, National Security, Financial Regulations, Trade policy are other macroeconomic factors calculated as uncertainty index derived using keywords/results from the Access World News database of over 2,000 US newspapers. (*Economic Policy Uncertainty Index*, 2012)

4.3 Descriptive Statistics

Initial data inspection was performed and it is observed that there were two missing/null values present in Trade Export and Trade Import variables which were treated using measures of central tendency. There are no duplicated values or anomalies present in the dataset.

Variables	count	mean	std	min	25%	50%	75%	max
U_R	251	5.51035857	1.91000845	3.2	4.1	5	6.7	14.2
I_P_I	251	97.4325339	4.85911504	84.7277	93.6555	98.7302	101.374	104.8129
Real_GDP	251	16590.023	1836.15427	13344.5108	15295.4967	16254.9852	18070.1182	20248.0625
C_P_I	251	227.353223	28.5798604	177.7	205.596	228.807	245.809	298.349
T_E	251	1.1184E+11	3.0461E+10	5.2667E+10	8.8128E+10	1.2025E+11	1.3307E+11	1.821E+11
T_I	251	1.7549E+11	4.2651E+10	8.3473E+10	1.4666E+11	1.8258E+11	1.9859E+11	2.9652E+11
O_P	251	65.715498	24.9055411	16.55	46.93	62.26	86.06	133.88
M_P	251	85.5584834	52.9528255	17.6162287	44.9418548	71.1810665	112.355779	304.069276
F_P	251	116.649852	73.9074204	23.0520648	66.5889505	99.1162375	140.39632	433.293459
TAX	251	121.718761	76.7516867	24.4412169	70.5318649	102.478109	150.425074	471.899606
GOV_S	251	93.5119419	95.5183283	5.7772429	37.2149007	64.755707	108.062741	635.271777
N_S	251	93.1182269	73.618623	23.737117	49.2393837	71.1313724	110.379411	593.459985
F_R	251	123.273166	120.750902	0	47.0101321	84.2470506	153.96918	877.545946
T_P	251	122.846569	218.65365	7.67261843	28.2103417	50.425833	92.6114495	1946.683
NASDAQ	251	4768.91283	3644.18942	1172.06006	2162.09009	2977.22998	6462.31006	15644.9697

Table 2:Descriptive statistics of all variables

These statistics give an overview of the distribution of values for each variable and can be used to identify patterns and trends in the data

From the above descriptive statistics we can identify that :

- There are 251 data points for each variable.
- The data skewness can be observed and distribution by looking at the difference between mean of all variables and mode (50%).
- Example on how to infer the data: Unemployment rate has a distribution which is almost a bell curve since its mean and mode is very close to each other. 75% of the time the unemployment rate has been under 6.7% and 50% of the times below 5% for the last 20 years.
- The mean of Real_GDP is 16590.02296 and the standard deviation is 1836.154271, indicating that most of the observations fall within a range of 14754 to 18426.
- Similarly, the mean and standard deviation of NASDAQ is 4684.0163 and 3423.693766 respectively. This suggests that most of the observations fall within a range of 1360.322534 and 8007.700006.
- Nasdaq stock index has been 4768.91 on an average with maximum of 15644.96 and minimum of 1172.06 for last 20 years.
- The min and max values of NASDAQ shows that there is quite a variation in the minimum and maximum values.

4.4 Exploratory Data Analysis

Univariate & Bivariate Analysis

- Unemployment Rate

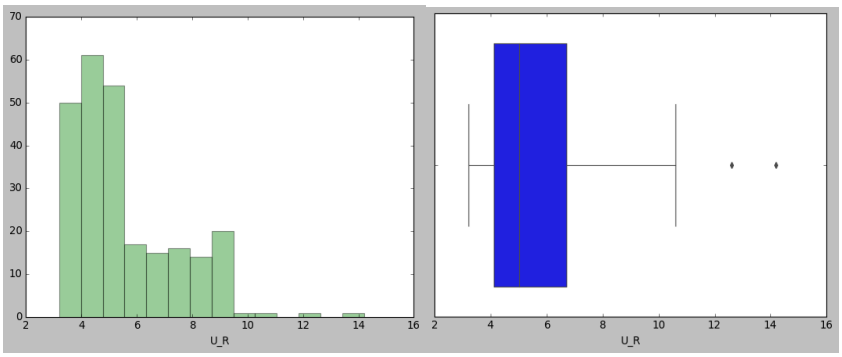


Figure 2:Histogram and Box Plot of U_R

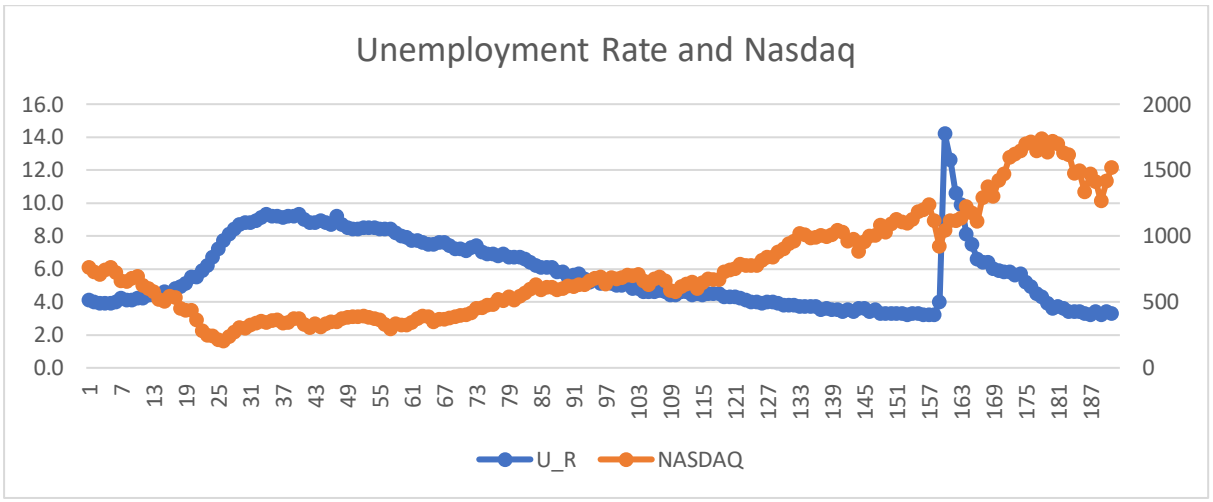


Figure 3:Time series chart of U_R & Nasdaq

- Distribution plot / histogram shows skewness in the data which can be described as a right skewed data.
- The box plot confirms this inference as the outliers can be clearly identified on the right hand side of the box. It also shows that 50% of the data lie between 4% to 7%
- Time series chart shows monthly data points for 20 years. A trend in the plot can be observed which is negative correlation as and when unemployment increases, the stock prices decrease and when unemployment decreases, the prices increase. This shows low employment will lead to high productivity and profits for the companies. Hence the initial evaluation through graphical representation shows that there is an impact of this macroeconomic factor on stock price.

- Industrial Production Index

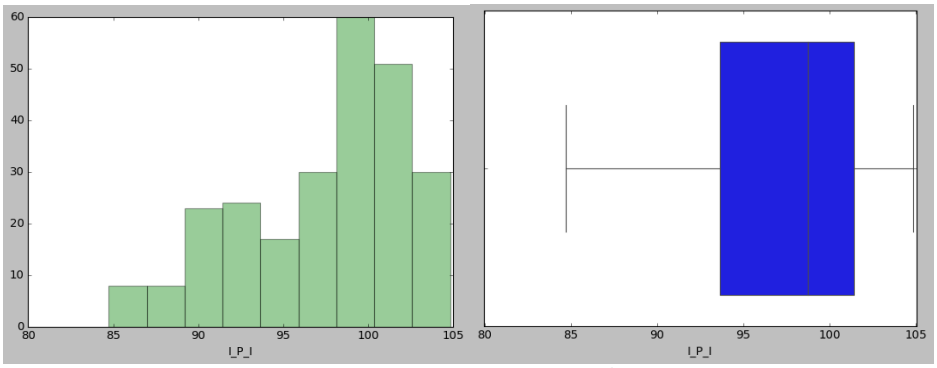


Figure 4:Histogram and box plot of I_P_I

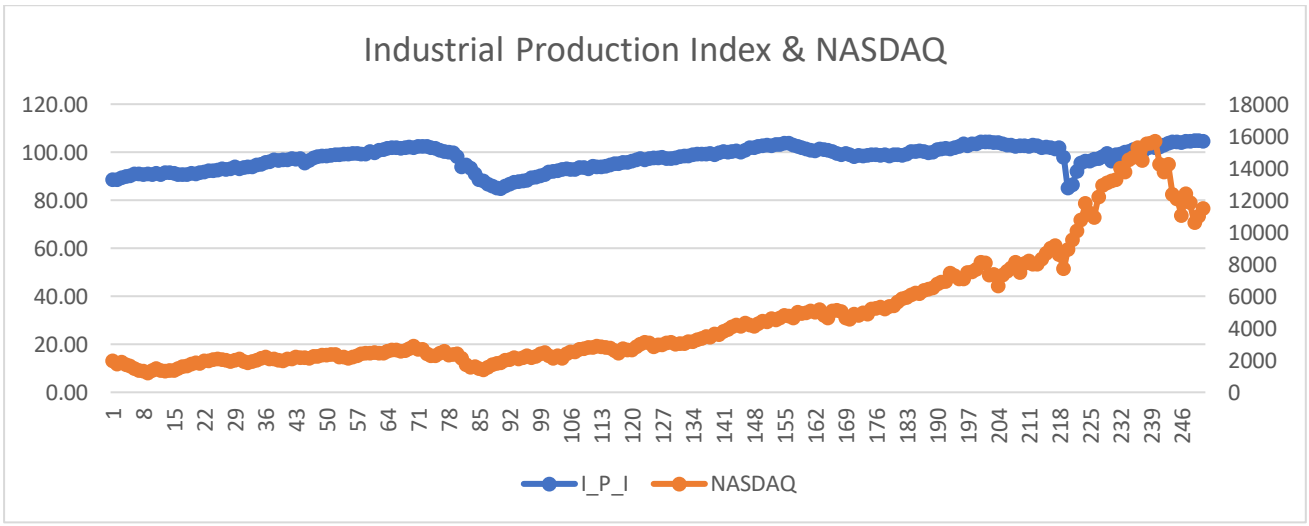


Figure 5:Time series chart of I_P_I vs Nasdaq

- Distribution plot / histogram shows skewness in the data which can be described as a left skewed data.
- The box plot confirms this inference it can be clearly seen that 50% of the data is between 94 – 102 approx, also without any outliers.
- Time series chart shows monthly data points for 20 years. A trend in the plot can be observed which is positive correlation as when IPI increases, the stock prices increase .This shows when industrial production index is high, economy is producing more goods and services, which is likely to lead to higher stock prices. Hence, the initial evaluation through graphical representation shows that there is an impact of the IPI on stock price. However, there is a need to investigate further as certain factors like economic policies and inflation rate etc effect the relationship between the two variables.

- Real Gross domestic Product

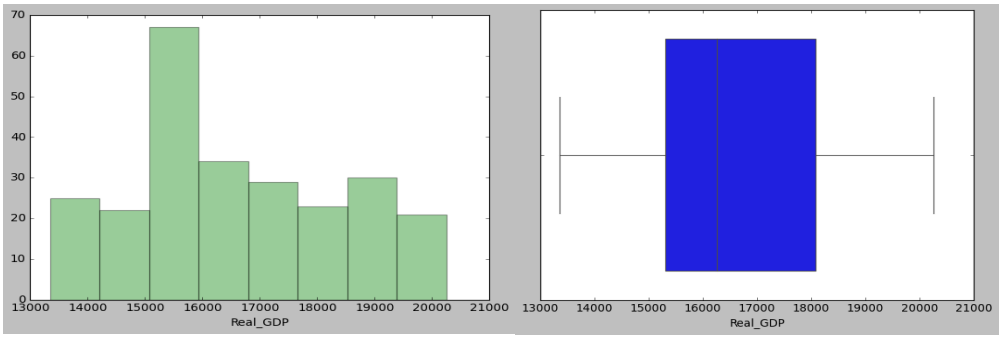


Figure 6:Histogram & Box plot of Real_GDP

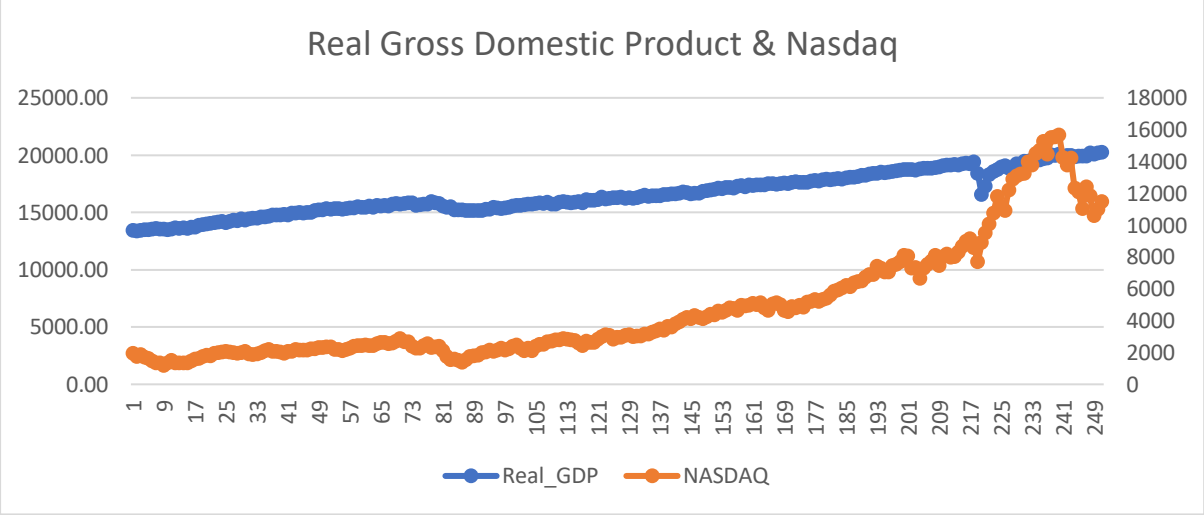


Figure 7:Time series chart of Real GDP vs Nasdaq

- Distribution plot / histogram shows a bell shaped curve, which indicates that the data is normally distributed
- The box plot confirms that there are no outliers present in the data.
- Time series chart shows monthly data points for 20 years. A trend in the plot can be observed which is strong positive correlation, as when GDP increases, the stock prices increase. This shows when GDP is growing, companies experience increased profits, which attracts more investors and pushes the stock prices up. Hence, the initial evaluation through graphical representation shows that there is significant impact of the GDP on stock price. However, there is a need to investigate further as the relationship is not always linear due to factors like natural disaster or pandemic like situation and the investors may sell stocks anticipating the low GDP growth in the future, even if the GDP is currently growing.

- Consumer Price Index

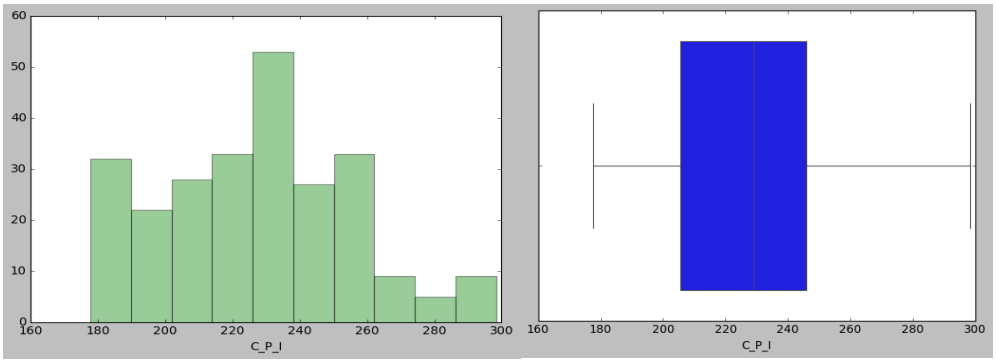


Figure 8:Histogram & Box Plot of C_P_I

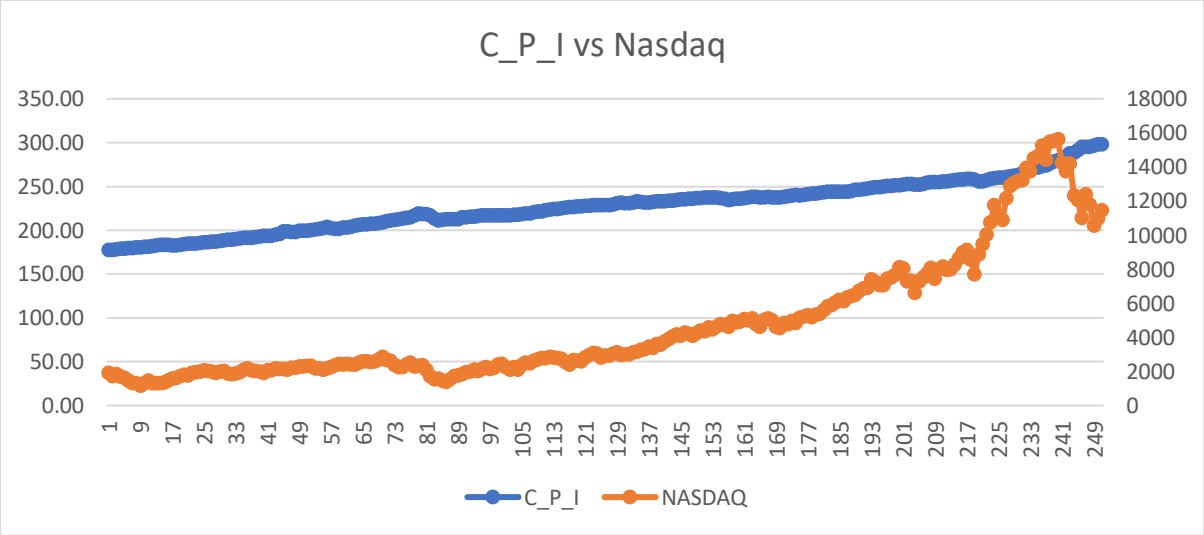


Figure 9:Time series chart of C_P_I vs Nasdaq

- Distribution plot / histogram shows skewness in the data which can be described as a right skewed data.
- The box plot confirms that is no outlier in the data.
- Time series chart shows monthly data points for 20 years. A trend in the plot can be observed which is positive correlation, indicating when inflation goes up, the stock prices also go up. Hence, the initial evaluation through graphical representation shows that there is significant impact of the CPI on stock price. However, this needs to be investigated further as the relationship is not always linear due to factors like government policies , interest rates wherein investors may sell stocks anticipating high inflation in the future, even if there is low inflation currently.

- Trade Exports

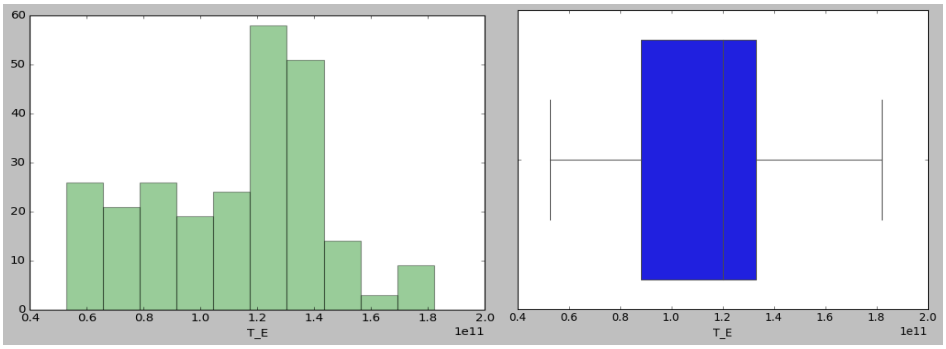


Figure 10:Histogram & Box Plot of T_E

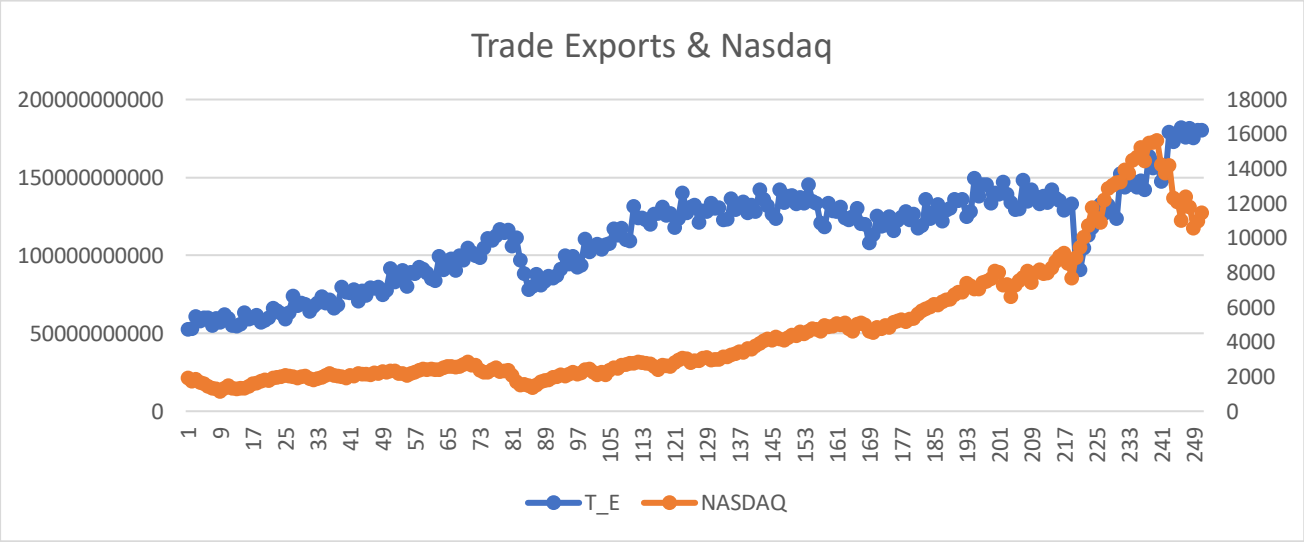


Figure 11:Time series chart of T_E vs Nasdaq

- Distribution plot / histogram shows almost a bell shaped distribution.
- The box plot confirms that is no outlier in the data.
- Time series chart shows monthly data points for 20 years. A trend in the plot can be observed which is positive correlation, indicating when exports are high, the stock prices rise. Hence, the initial evaluation through graphical representation shows that there is significant impact of the exports on stock price. This is because companies are selling more goods and services hence a strong export market can attract more investors, which pushes stock prices up. However, this needs to be investigated as the relationship is not always linear due to factors like exchange rates, trade agreements wherein investors may sell stocks anticipating of lower future exports and profits

- Trade Imports

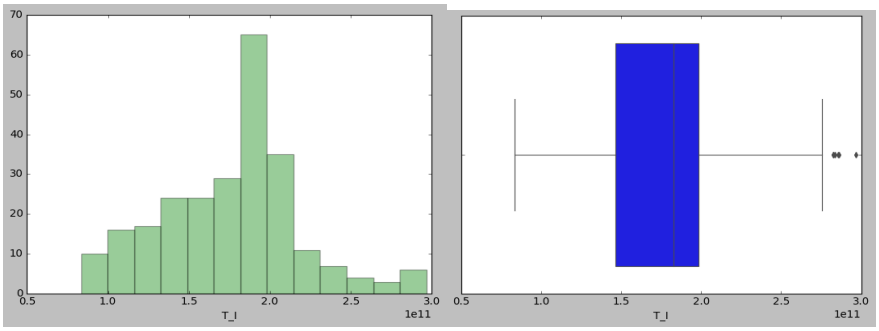


Figure 12:Histogram & Box Plot of T_I

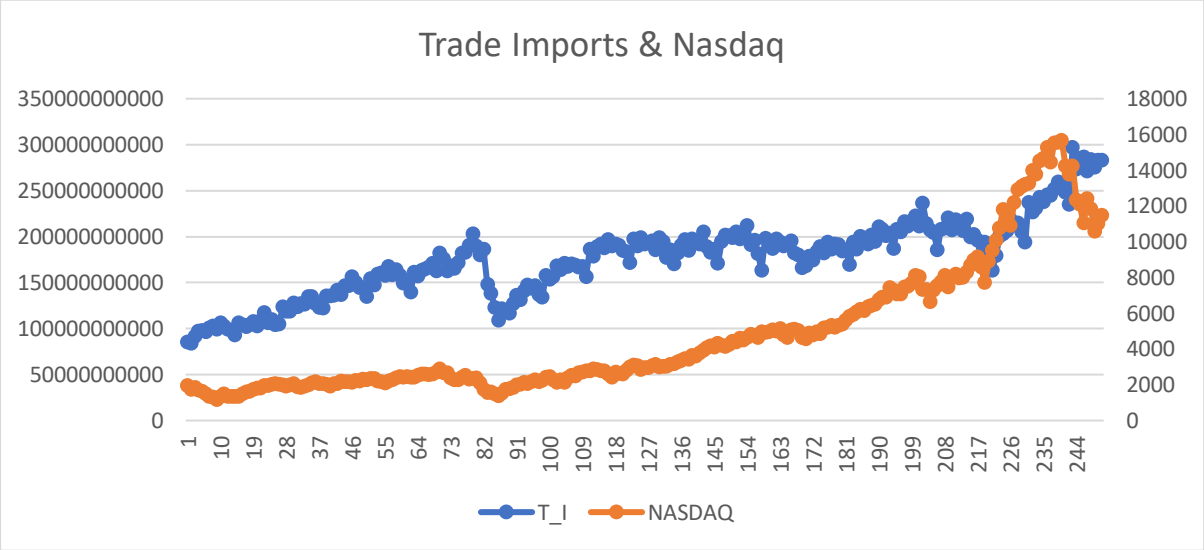


Figure 13:Time series chart of T_I vs Nasdaq

- Distribution plot / histogram show right skewed data.
- The box plot confirms that are outliers in the data.
- Time series chart shows monthly data points for 20 years. A trend in the plot can be observed which is positive correlation, indicating when imports are high, the stock prices rise. Hence, the initial evaluation through graphical representation shows that there is significant impact of the imports on stock price. This is because companies will have access to large number of products, raw materials etc, hence more sales which pushes stock prices up. However, this needs to be investigated as the relationship is not always linear due to factors like exchange rates, trade agreements.

- Oil Prices

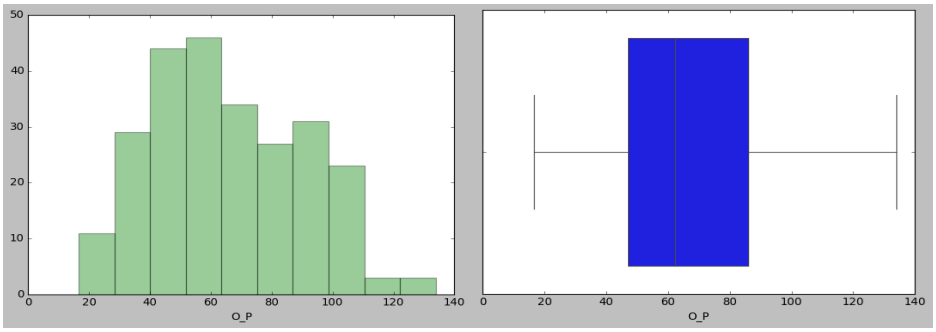


Figure 14:Histogram & Box Plot of Oil Price

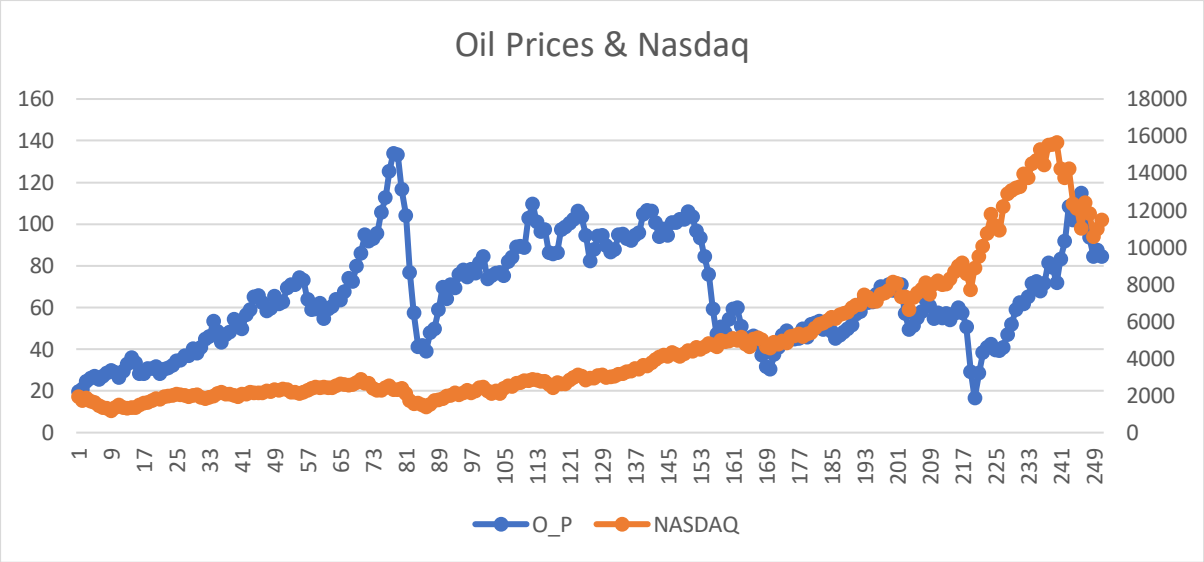


Figure 15:Time Series Chart of O_P vs Nasdaq

- Distribution plot / histogram shows almost a bell shaped distribution.
- The box plot confirms that is no outlier in the data.
- Time series chart shows monthly data points for 20 years. There is a certain trend indicating when oil prices are high, the stock prices remain constant and when it decreased the prices went high. This is because when the oil price are low the cost of production is also low, which improves the purchasing power. Hence, the initial evaluation through graphical representation shows that there exists a certain relationship of oil price on stock price though insignificant and needs to be investigated through statistical models.

- Monetary Policy Index

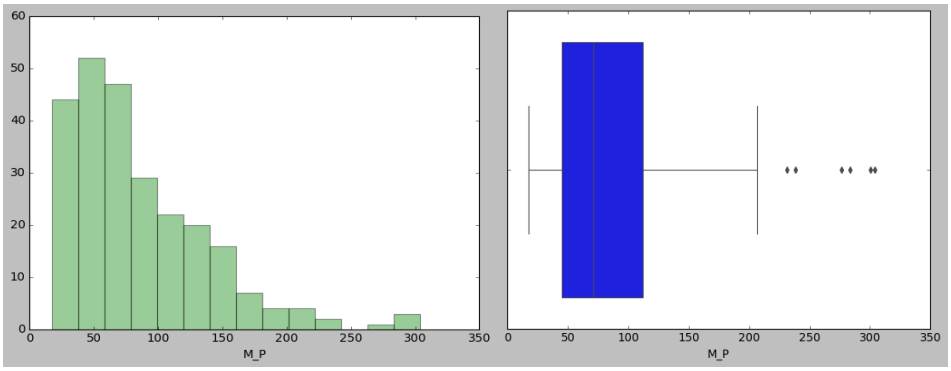


Figure 16:Histogram & Box Plot of Monetary Policy

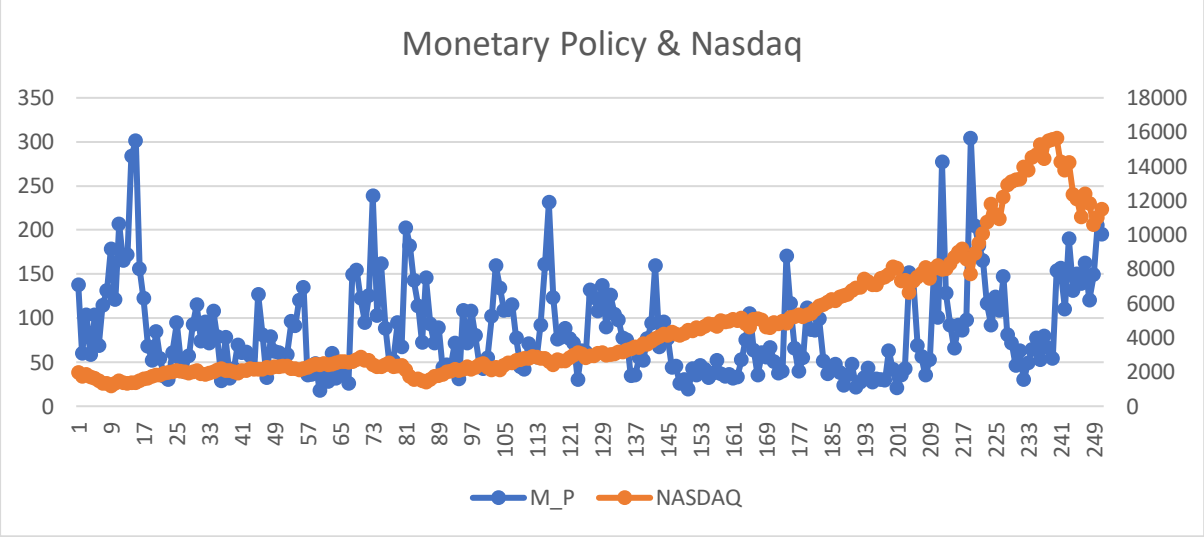


Figure 17:Time Series chart of M_P vs Nasdaq

- Distribution plot / histogram shows skewness in the data which can be described as a right skewed data.
- The box plot confirms this inference as the outliers can be clearly identified on the right hand side of the box.
- Time series chart shows monthly data points for 20 years. There is a certain trend indicating when Monetary policy uncertainty index are high, the stock prices remain constant and as the uncertainty decreases the prices went high. Monetary policy can effect stock returns based on if the policy is expansionary (prices go up) or contractionary(prices go down). Hence, the initial evaluation through graphical representation shows that there exists a certain relationship of Monetary policy on stock price though insignificant and needs to be investigated further through statistical models.

- Fiscal Policy Index

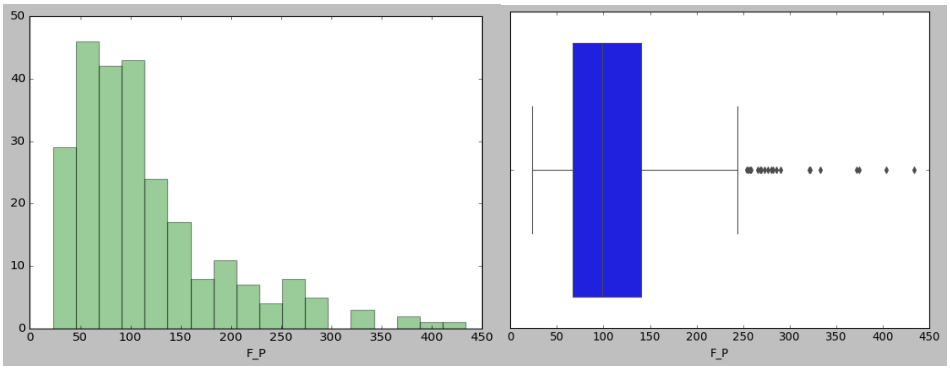


Figure 18:Histogram & Box Plot of F_P

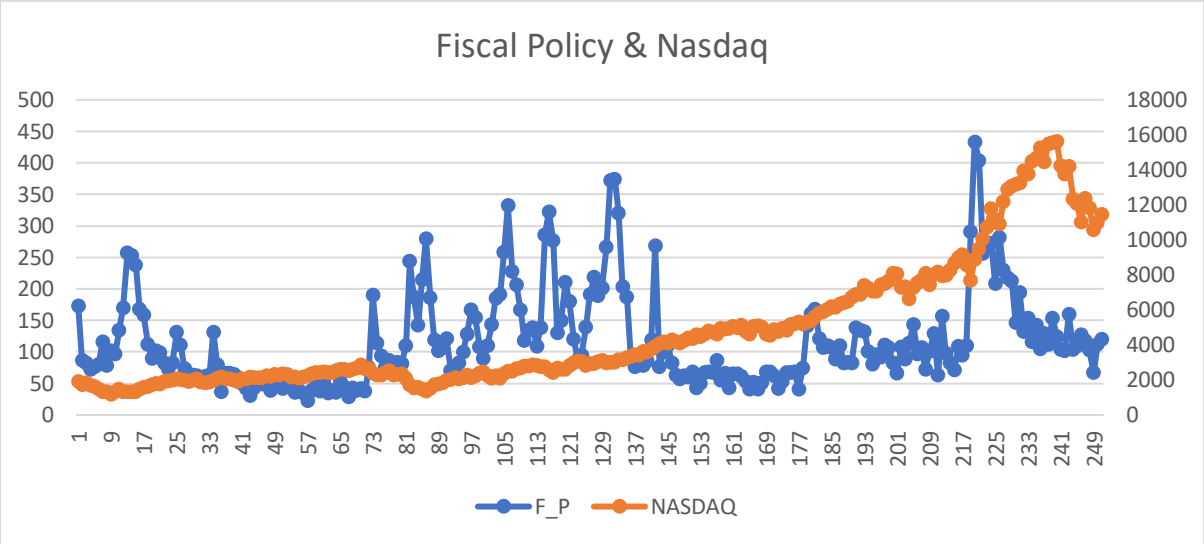


Figure 19:Time series chart of F_P vs Nasdaq

- Distribution plot / histogram shows skewness in the data which can be described as a right skewed data.
- The box plot confirms this inference as the outliers can be clearly identified on the right hand side of the box.
- Time series chart shows monthly data points for 20 years. There is a certain trend observed, indicating when Fiscal policy uncertainty index are high, the stock prices remain constant and when it decreases the prices went high. Fiscal policy index can effect stock returns based on if the policy is expansionary (prices go up) i.e. high government spendings and low taxes which stimulates demand and increases customer spending or contractionary(prices go down). Hence, the initial evaluation through graphical representation shows that there exists a certain relationship of Fiscal policy on stock price and needs to be investigated further through statistical models.

- Tax Index

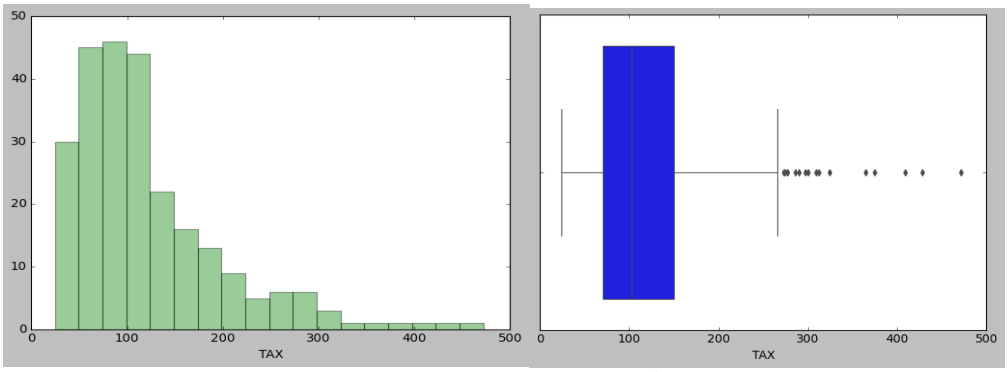


Figure 20:Histogram & Box Plot of Tax

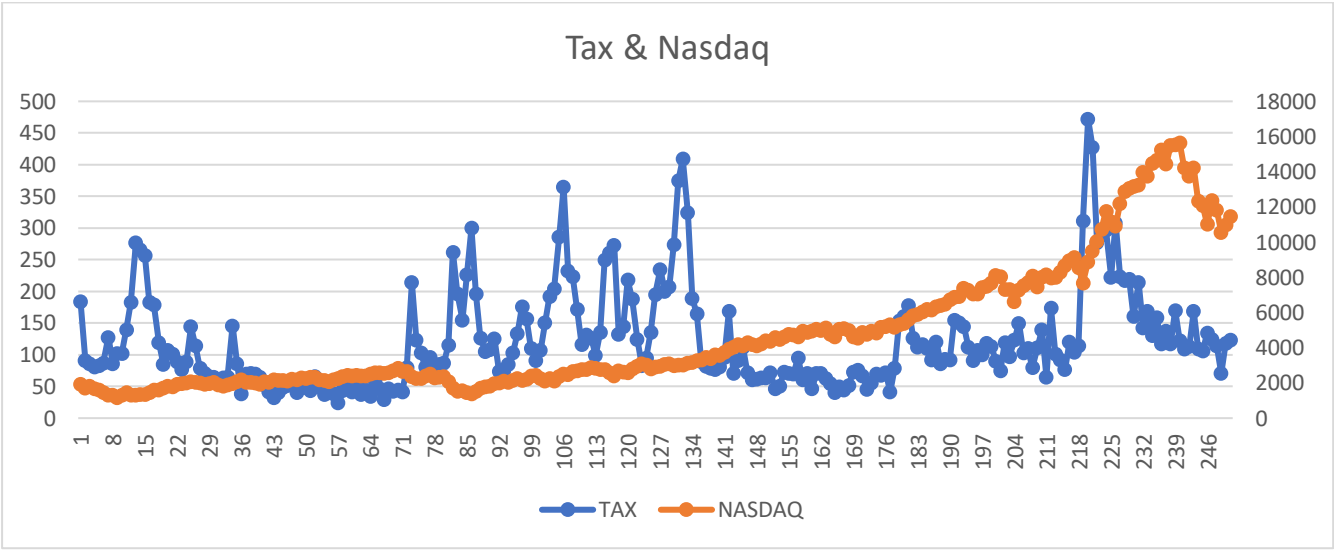


Figure 21:Time series chart of Tax vs Nasdaq

- Distribution plot / histogram shows skewness in the data which can be described as a right skewed data.
- The box plot confirms this inference as the outliers can be clearly identified on the right hand side of the box.
- Time series chart shows monthly data points for 20 years. A certain trend is observed, indicating when Tax uncertainty index are high, the stock prices are low and when it decreases the prices are high ,this is because when tax are lower the disposable income increases which boosts economic growth and attracts investors to buy more stocks.. Hence, the initial evaluation through graphical representation shows that there exists a certain relationship of Tax uncertainty index on stock price and needs to be investigated further through statistical models.

- Government Spendings Index

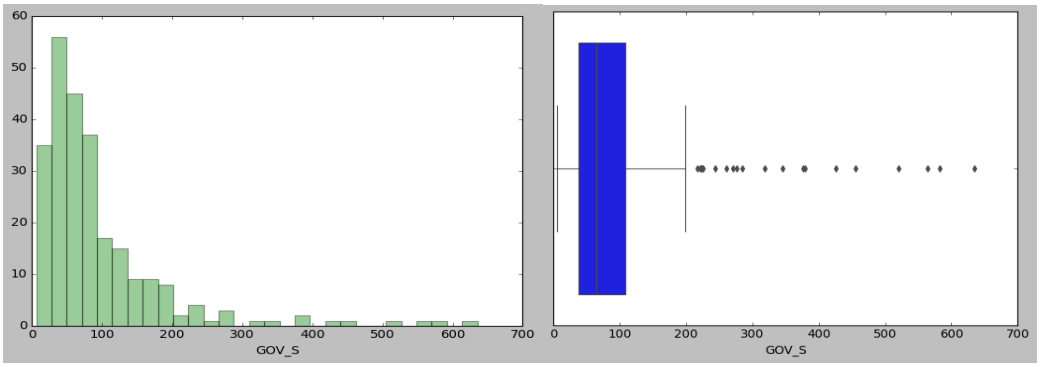


Figure 22:Histogram & Box Plot of Gov_S

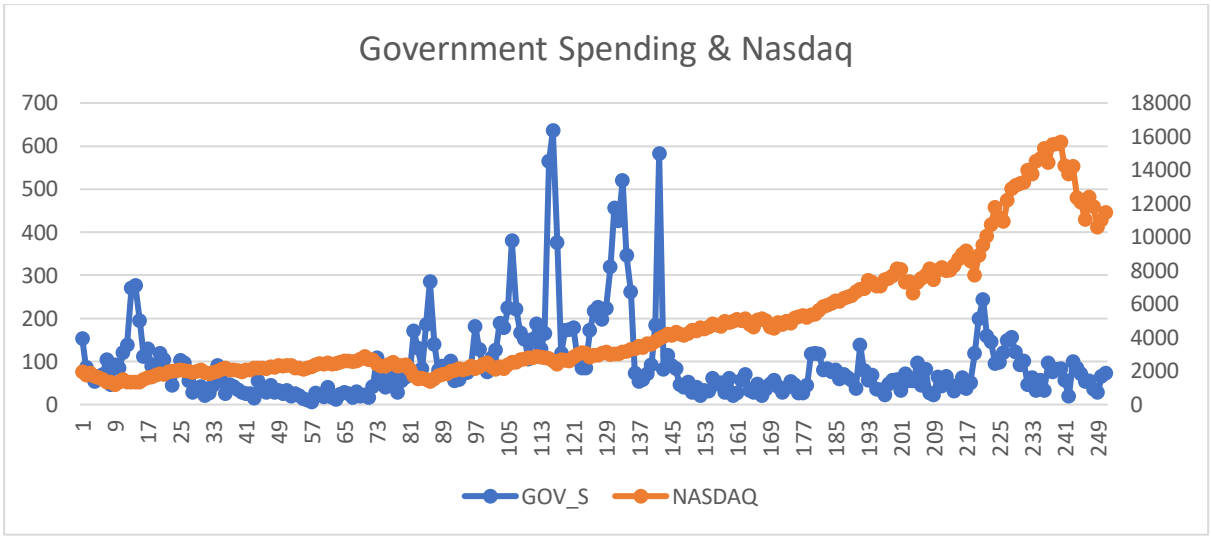


Figure 23:Time series chart of Gov_S & Nasdaq

- Distribution plot / histogram shows skewness in the data which can be described as a right skewed data.
- The box plot confirms this inference as the outliers can be clearly identified on the right hand side of the box.
- Time series chart shows monthly data points for 20 years. A certain negative trend is observed, indicating when Government spending uncertainty index are high, the stock prices remain low and constant and when it decreases the prices shoot up. Hence, the initial evaluation through graphical representation shows that there exists a certain relationship of government spending uncertainty index on stock price and needs to be investigated further through statistical models.

- National Security Index

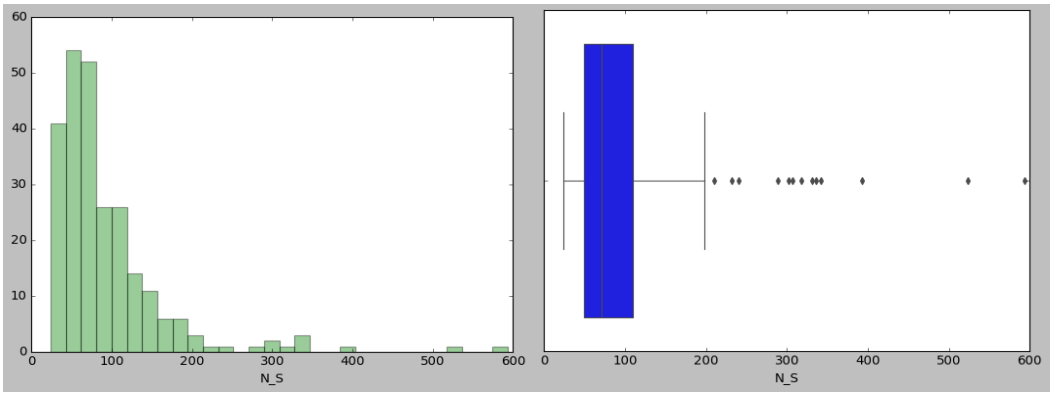


Figure 24:Histogram & Box Plot of N_S

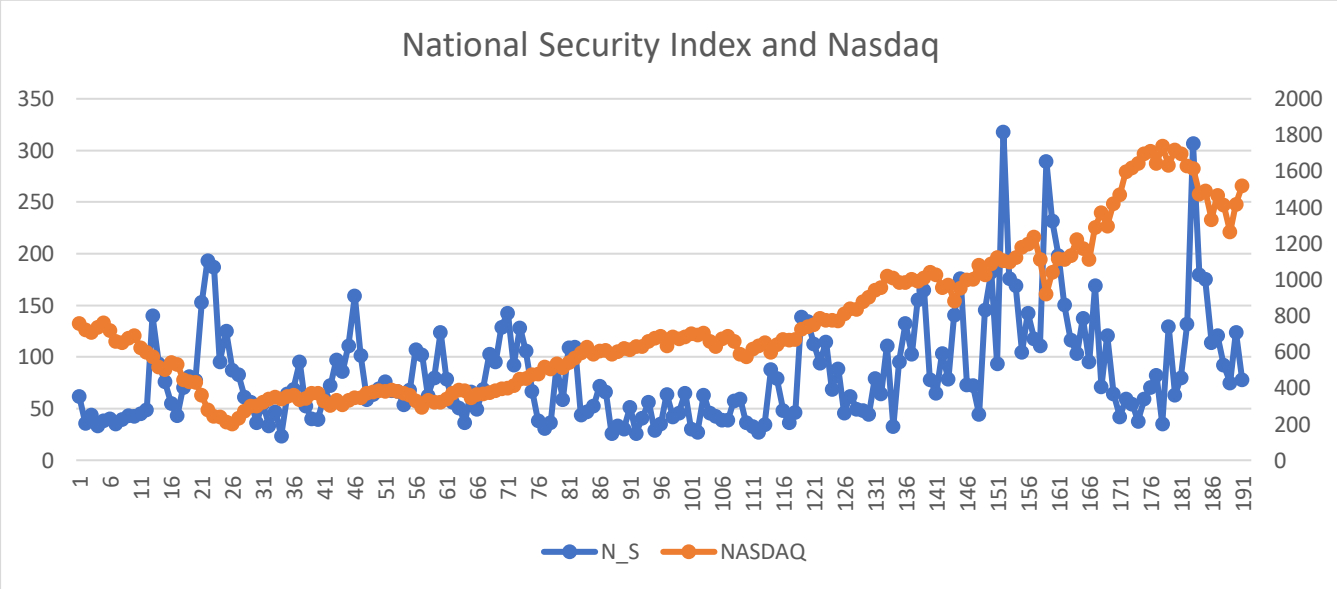


Figure 25:Time series chart of N_S vs Nasdaq

- Distribution plot / histogram shows skewness in the data which can be described as a right skewed data.
- The box plot confirms this inference as the outliers can be clearly identified on the right hand side of the box.
- Time series chart shows monthly data points for 20 years. There is a certain trend observed indicating when National security uncertainty index are high, the stock prices remain low and constant and when it decreases the prices shoot up. Hence, the initial evaluation through graphical representation shows that there exists a certain relationship of national security uncertainty index on stock price and needs to be investigated further through statistical models.

- Financial Regulations

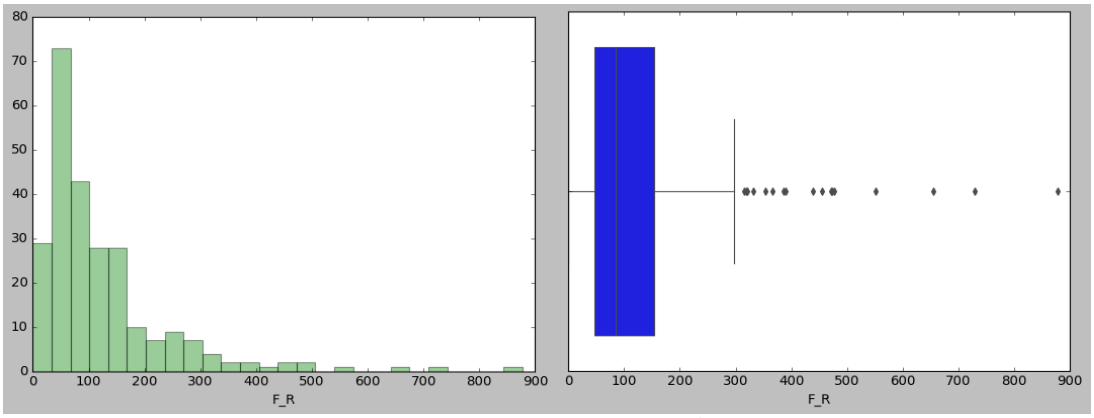


Figure 26:Histogram & Box Plot o f F_R

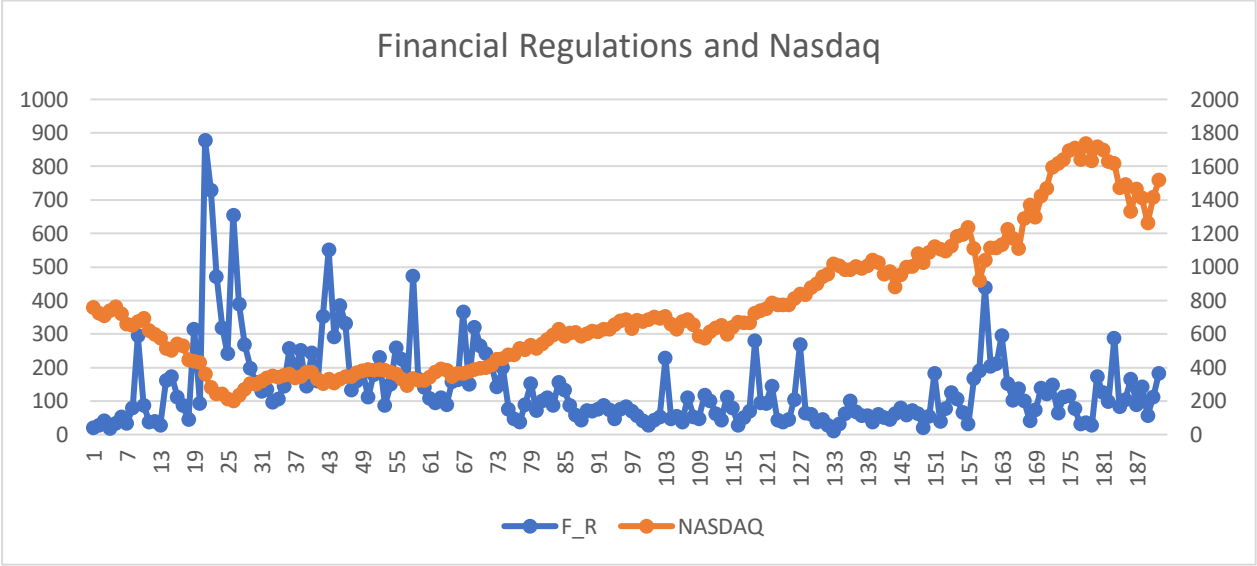


Figure 27:Time series chart of F_R & Nasdaq

- Distribution plot / histogram shows skewness in the data which can be described as a right skewed data.
- The box plot confirms this inference as the outliers can be clearly identified on the right hand side of the box.
- Time series chart shows monthly data points for 20 years. A negative trend is observed, indicating when financial regulations uncertainty index are high, the stock prices remain low and constant and when it decreases the stock prices increase. Hence, the initial evaluation through graphical representation shows that there exists a certain relationship of financial regulations uncertainty index on stock price and needs to be investigated further through statistical models.

- Trade Policy Index

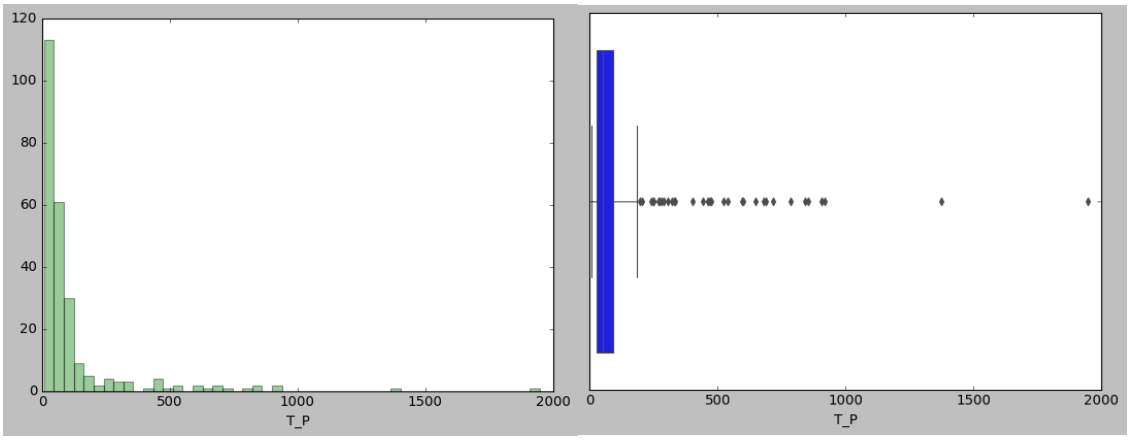


Figure 28:Histogram & Box Plot of T_P

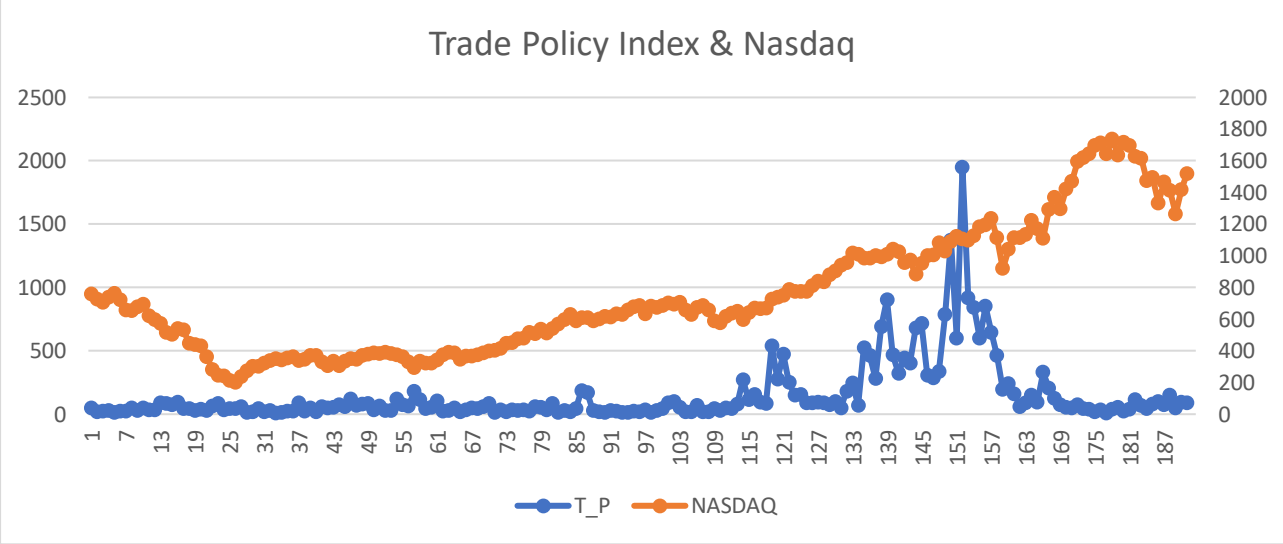


Figure 29:Timer series chart of T_P and Nasdaq

- Distribution plot / histogram shows skewness in the data which can be described as a right skewed data.
- The box plot confirms this inference as the outliers can be clearly identified on the right hand side of the box.
- Time series chart shows monthly data points for 20 years. There is a trend observed indicating a constant behaviour . Hence, the initial evaluation through graphical representation shows that there is no significant relationship of Trade Policy uncertainty index on stock price and needs to be investigated further through statistical models.

- NASDAQ Index

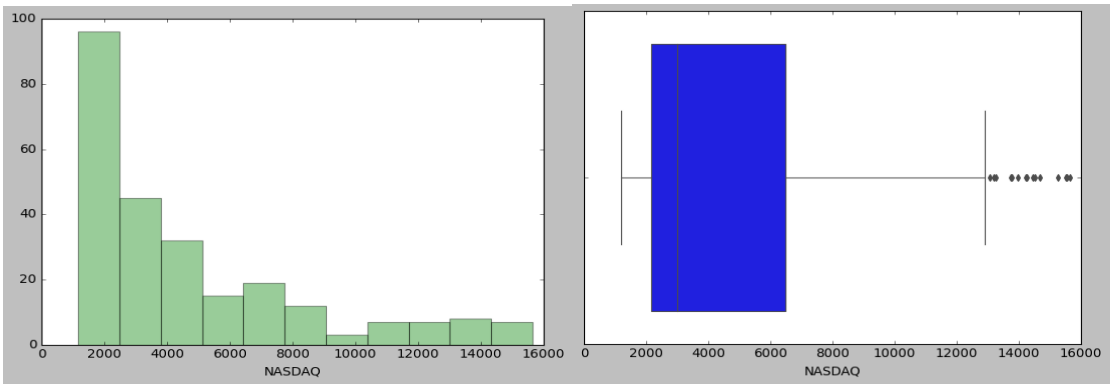


Figure 30:Histogram & Box Plot of Nasdaq

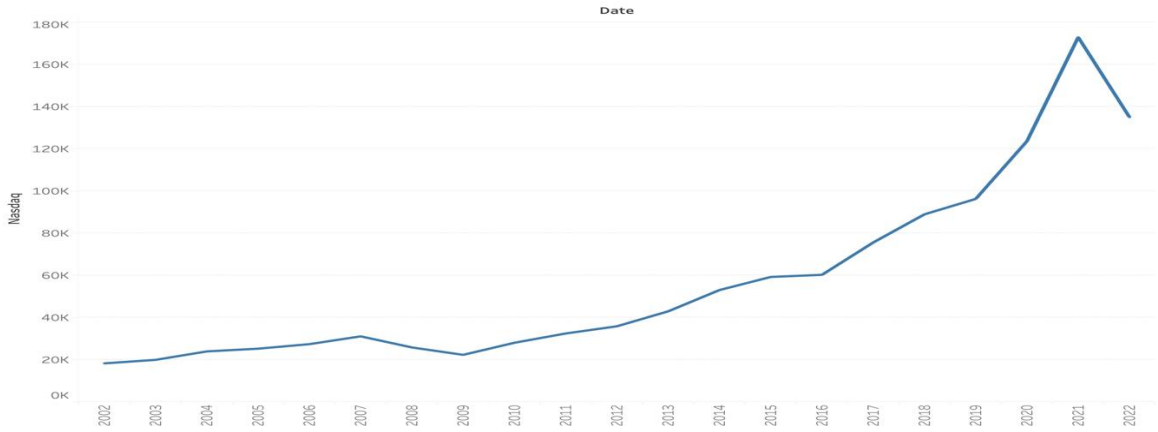


Figure 31:Time series chart of Nasdaq stock index

- Distribution plot / histogram shows skewness in the data which can be described as a right skewed data.
- The box plot confirms this inference as the outliers can be clearly identified on the right hand side of the box.
- Time series chart shows monthly data points of Nasdaq stock returns for 20 years and shows an increasing trend over the years. The highest being in 2021 and then a decreasing trend thereon.

Multivariate Analysis

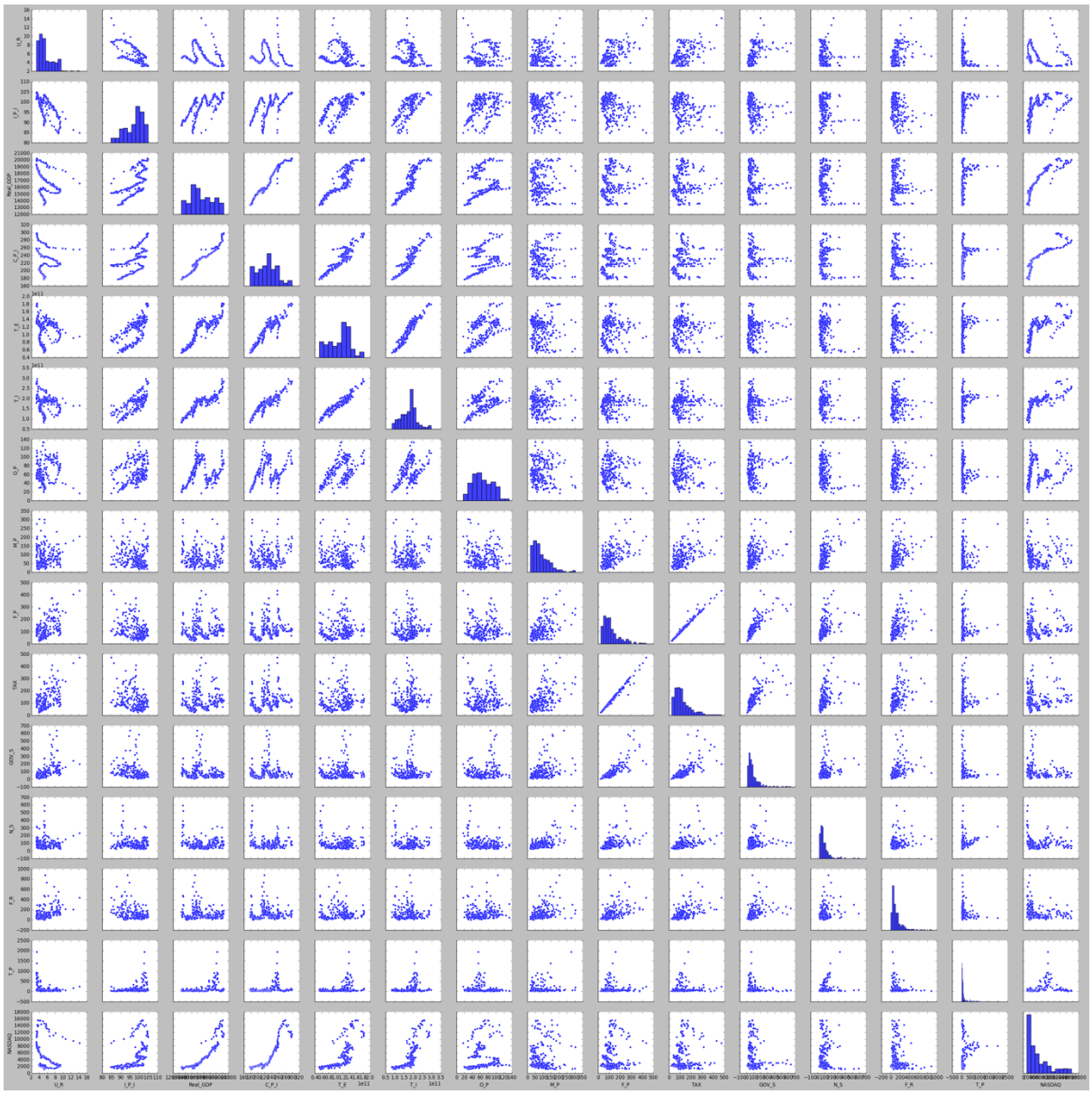


Figure 32:Scatter Plot

- The scatter plot here shows the relationship between each variable, and it can be inferred from the graph that there are several variables which are correlated to each other the scatter plot shows a linear trend for some of the variables.
- The scatter plot shows a high positive relation between GDP , CPI , Trade Exports and Imports & Nasdaq
- Observation also shows a negative relationship between Unemployment rate and Nasdaq in the graph.

Correlation Analysis

A statistical measure called correlation expresses how closely two variables are related to one another. Positive and negative correlations are the two main categories. When two variables move in the same direction, there is a positive correlation; when one increases, the other increases as well. When two variables change in the opposite directions—when one rises, the other falls—there is a negative correlation.

The association between two variables can be valued using a correlation coefficient. The range of correlation coefficients is from -1 to 1. No relationship between the variables is shown by a "0," while a perfect negative or positive correlation is indicated by a "-1 & +1"(Glen, 2021).

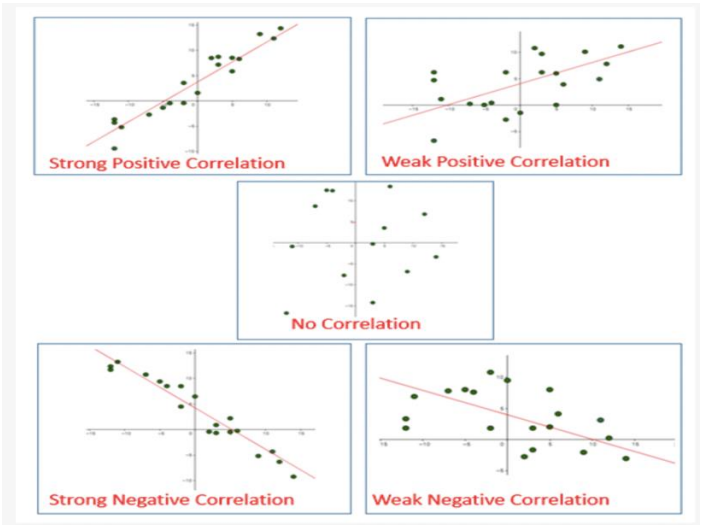


Figure 33:Correlation using scatter plot

The most common correlation coefficient is the Pearson Correlation Coefficient. The Pearson correlation coefficient is used in many different fields, including psychology, sociology, and biology. Pearson correlation coefficient between two variables x (independent variable) and y (dependent variable) can be calculated using the below formula x bar is the mean value of x and y bar is the mean value of y.

$$r = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\sqrt{[\Sigma(x - \bar{x})^2 (y - \bar{y})^2]}}$$

The below table shows the values of the coefficients for our dataset using Pearson correlation formula.

	U_R	I_P_I	Real_GDP	C_P_I	T_E	T_I	O_P	M_P	F_P	TAX	GOV_S	N_S	F_R	T_P	NASDAQ
U_R	1														
I_P_I	-0.6712619	1													
Real_GDP	-0.2995381	0.68964383	1												
C_P_I	-0.139967	0.60357957	0.97264229	1											
T_E	-0.1467479	0.70175911	0.89227063	0.9287147	1										
T_I	-0.2306516	0.74269525	0.90827	0.92963572	0.96097856	1									
O_P	0.16385051	0.36964623	0.21312543	0.31787657	0.53836015	0.51444855	1								
M_P	0.09795169	-0.1568292	-0.0230886	0.05037787	-0.0353668	0.0163148	-0.0426888	1							
F_P	0.56876534	-0.3169824	0.07690077	0.18596535	0.10869375	0.075835	0.01409634	0.52452388	1						
TAX	0.54531641	-0.3187264	0.08840874	0.19003282	0.0935236	0.0684537	-0.0285366	0.51670557	0.98944472	1					
GOV_S	0.52644279	-0.2893424	-0.1196481	-0.0109205	0.04368765	-0.0276045	0.17025819	0.39501458	0.82033696	0.74780395	1				
N_S	-0.0193373	-0.2133903	-0.0999995	-0.0765239	-0.1742902	-0.1488078	-0.2699019	0.69301532	0.47425023	0.492268	0.29529262	1			
F_R	0.4584814	-0.3870114	-0.1307445	-0.0214973	-0.0399842	-0.0730606	0.09652391	0.43755747	0.50967748	0.51440752	0.41700456	0.25969595	1		
T_P	-0.2975865	0.29483796	0.40497122	0.31759939	0.26331027	0.25213859	-0.121456	0.16558991	0.05212531	0.0717237	-0.0907563	0.28273094	-0.0923392	1	
NASDAQ	-0.2538118	0.50375381	0.8975576	0.8763567	0.7126405	0.78549959	0.050956	0.06170439	0.15161836	0.17383612	-0.1274301	0.03448958	-0.1278043	0.28450709	1

Table 3:Pearson correlation coefficients

We can use correlation heatmaps to identify possible links between variables and to gauge how strong these relationships are. Additionally, outliers and linear and nonlinear correlations can be found using correlation graphs. The cells' color-coding makes it simple to quickly spot any links between the variables. Finding both linear and nonlinear associations between variables is possible with the aid of correlation heatmaps.

Generally speaking, multi-collinearity is present when the Pearson correlation coefficient is more than 0.7.

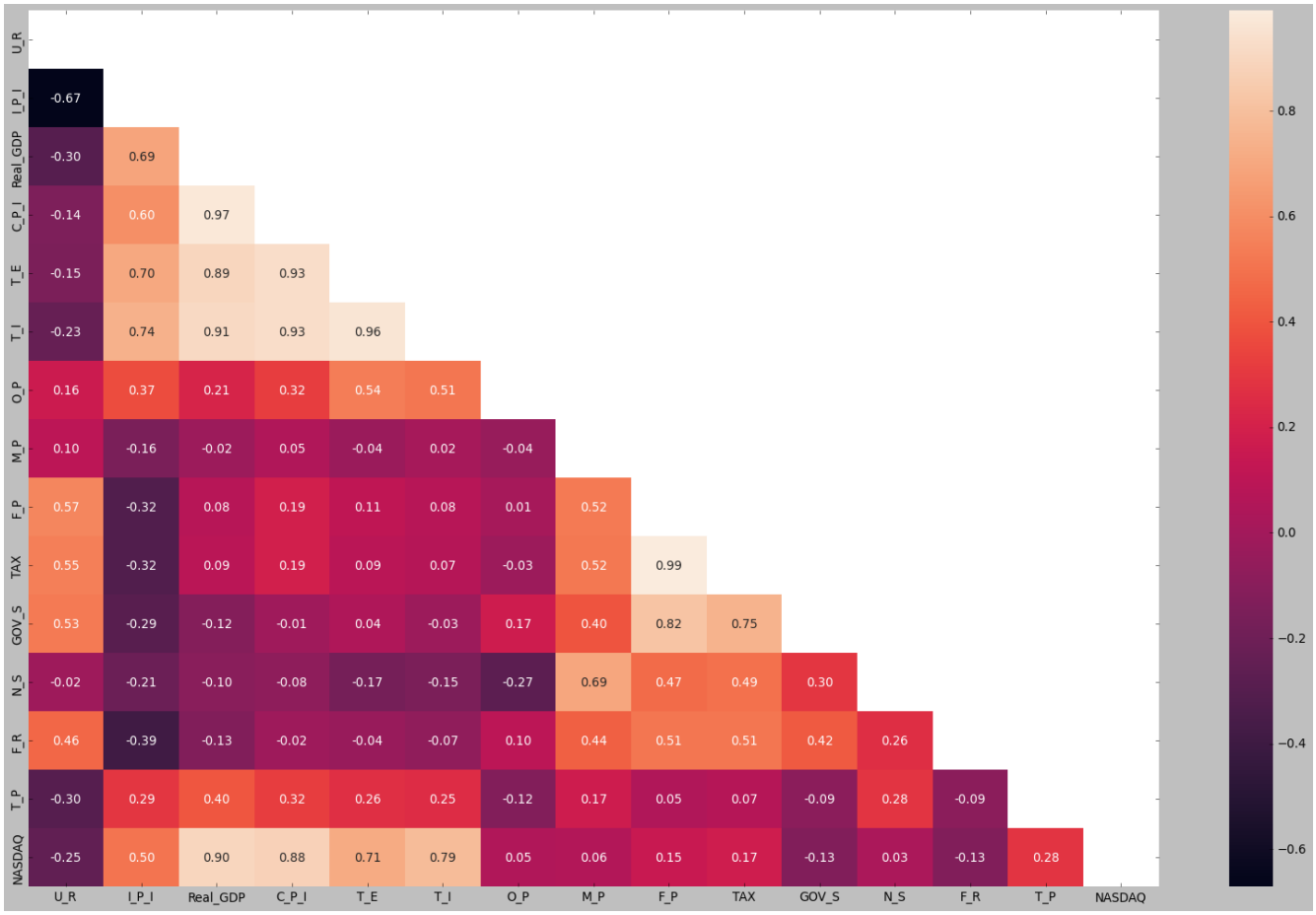


Figure 34:Correlation Heat Map

Inferences from the correlation heat map are as follows :

- Strong positive correlation between Nasdaq & GDP , Consumer price index , Trade exports and imports can be seen.
- A negative correlation can be seen between Nasdaq & Unemployment rate however, it is not very strong, the correlation indicated that when the unemployment increases the stock prices also decrease as a result of less sales and lower profit.
- Multi-collinearity can be observed in the data, various variables like Trade imports and exports, GDP & Consumer price index, Tax & Fiscal Policy index are highly correlated to each other with the coefficient very close to 1.
- There is a strong negative correlation between Unemployment rate and Industrial production index indicating when people are unemployed the production also decreases.

Principal Component Analysis

In order to be able to run PCA successfully, the statistical significance needs to be identified. To verify this assumption below hypothesis tests are conducted:

BARTLETT TEST OF SPHERICITY

This tests the hypothesis that the variables are uncorrelated in the data.
H0: All variables are uncorrelated.
Ha: At least one variable in the data is correlated.

If the null hypothesis cannot be rejected then PCA is not recommended. (alpha = 0.05)

After importing factor analyser and running the code in python it is seen that p value is 0.0 which is lower than alpha value, and hence the null hypothesis is rejected and conclude at least one variable in the dataset is correlated.

KMO Test

The Kaiser-Meyer-Olkin (KMO) - measure of sampling adequacy (MSA) is an index used to examine how appropriate PCA is.

Generally, if MSA is less than 0.5, PCA is not recommended, since no reduction is expected. On the other hand, MSA > 0.7 is expected to provide a considerable reduction is the dimension and extraction of meaningful components.

After running the code in python the score 0.76 was calculated, which is greater than 0.7 hence, it can be inferred that there is enough number of observation and can expect the model to give us meaningful components.

Performing PCA

The PCA components were extracted (eigen vectors) and the PCA explained variance ratio/ eigen values as below, which indicates the variances captured by each new variables and the first three seems to explain the 80% variances hence we choose to work with these 3 variables instead of 14 variables for linear regression and see how the model performs.

Eigen Values:

```
array([0.3557828 , 0.67253856, 0.80446681, 0.86799644, 0.90663925,
       0.94226523, 0.96401543, 0.9813594 , 0.989451  , 0.99460947,
       0.99731432, 0.99915838, 0.99964454, 1.        ])
```

Eigen vectors: Eigen vectors indicates the direction of the principal component, which helps in understanding the linear transformation. The table below shows the eigen vectors with original features

	U_R	I_P_I	Real_GDP	C_P_I	T_E	T_I	O_P	M_P	F_P	TAX	GOV_S	N_S	F_R	T_P
PCA 0	0.2680855	-0.4033233	-0.3638471	-0.331531	-0.3522323	-0.3676475	-0.1542272	0.14426221	0.1923544	0.1906529	0.25186158	0.14624016	0.21228366	-0.1268838
PCA 1	0.16179397	0.02607225	0.23667774	0.28779588	0.26268053	0.24797795	0.12737502	0.26651584	0.40735712	0.40343952	0.3377208	0.25280517	0.27836541	0.17998484
PCA 2	-0.4330183	0.05757789	0.06823149	-0.0208693	-0.1548865	-0.1051022	-0.5035153	0.23773075	0.00711295	0.03638376	-0.1355775	0.46371339	-0.1563551	0.44696344

Table 4:Eigen Vectors with original features

The explicit form of PC0 in terms of eigen vectors are :

(0.27 x U_R) + (-0.40 x I_P_I) + (-0.36 x Real GDP) + (-0.33 x C_P_I) + (-0.35 x T_E) + (-0.37 x T_I) + (-0.15 x 0_P) + (0.14 x S_M_P) + (0.19 x F_P) + (0.19 x TAX) + (0.25 x GOV_S) + (0.15 x N_S) + (0.21 x F_R) + (-0.13 x T_P)

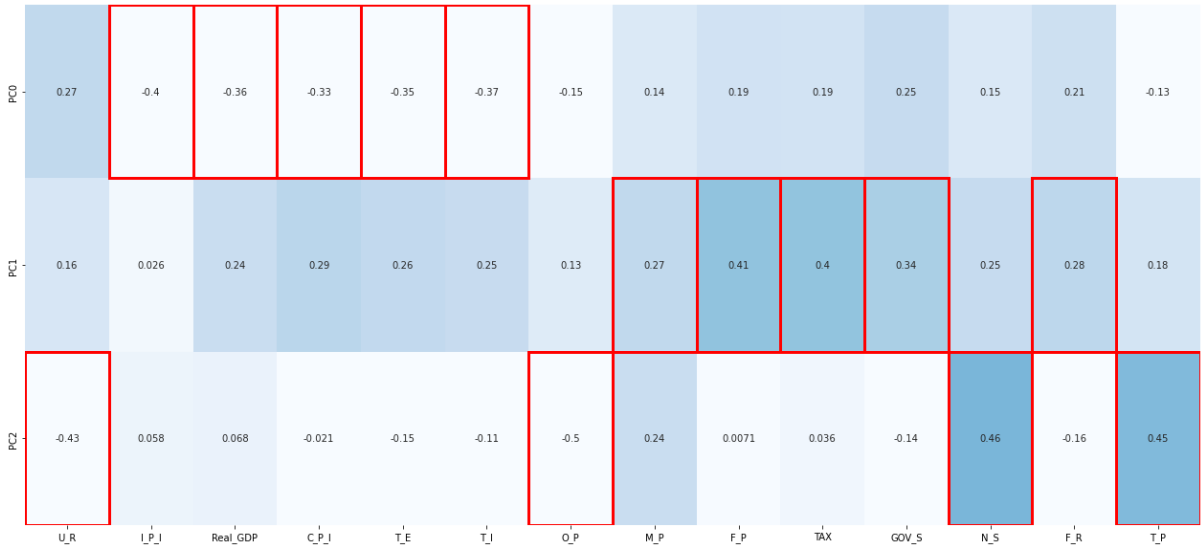


Figure 35:Heat Map of new PC and & original features

Inference from the above Heat map:

- The above heat map shows for a particular variable which PC group has the highest coefficient, i.e. marked as red.
- PC0 shows the most influential factors as I_P_I, GDP, C_P_I & Trade exports and imports hence, named this new variable as Economic Indicators.
- PC1 shows the most significant factors as M_P, F_P, TAX, GOV_S & F_R hence, named this group as Government regulations and Policies
- PC2 has the most influential factors as U_P, O_P, N_S & T_P hence, named this new variable as Unemployment rate & Oil Prices.

Thus, the dimensions were successfully reduced from 14 to just 3 variables. In the next step a model with PCA variables and the original variables will be built to compare the model performance and accuracy.

4.5 MODEL BUILDING

In this stage, data sets for production, testing, and training are created. These data sets give the ability to create an analytical approach, train it, and put aside some data for model testing.
Additionally, depending on the work done in the model planning phase, creation and implementation of models is done in this phase. Necessary consideration should be taken into account; if its current tools will be adequate for running the models or if it would require a more robust environment for doing so (Example – fast hardware and parallel processing).

Linear Regression

To perform linear regression we split data(70:30) into training and test data with 70% of the data in train set and 30% of the data in test set. The training data is used to create the model and the test data shows how well the model performance is , so that the results can be supervised and reworked upon. This data is fitted into a model that takes in an input and predicts the output. The RMSE values and the regression scores ($r^{2)}$ on train and test set are then calculated.

LR MODEL 1 – using PCA Variables

For this particular model we will be running regression analysis on the 3 variables derived after principal component analysis. After applying the regression model on the scaled data as PCA requires the units of the variables to be on the same scale and using scikit learn we go ahead to find the intercept & coefficients of each independent variable as below :

The intercept is -0.042
The coefficient for Economic_Indicator is -0.281
The coefficient for Gov_Policies is 0.22
The coefficient for Unemployment_oil_prices is 0.15

Training Score(r^2)	73
Test Score(r^2)	68
RMSE on Train	0.49
RMSE on Test	0.62

Table 5:LR Model 1 accuracy scores

The intercept of the model is -0.042 which simply means when X (independent variables) are 0 then the price of the stock will decrease by 0.042. The Linear equation is :

*Nasdaq = (-0.042) * Intercept + (-0.281) * Economic indicator + (0.22) * Gov_policies + (0.15) *unemployment_oil_prices*

When there is a unit change increase in unemployment rate and oil prices , Nasdaq stock prices increases by 0.15 units, keeping all other predictors constant .Similarly for Gov_policies with one unit change there is an impact on stock prices by 0.15 units.
similarly, when there is an increases by 1 unit in Economic indicator, stock price decreases by 0.281units, keeping all other predictors constant.

Remarks: The model can be slightly on the overfit zone , the model is not performing at its best with low r^2 scores.

LR MODEL 2 – Full Weighted model

For this particular model regression analysis using unscaled data on the all 14 independent variables was performed, keeping Nasdaq stock index as the dependent variable. After applying the regression model using scikit learn the intercept & coefficients of each independent variable was calculated as below :

The coefficient for U_R is -242.85,	The coefficient for I_P_I is -107.56
The coefficient for Real_GDP is 0.56,	The coefficient for C_P_I is 105.51
The coefficient for T_E is -7.86023619525622e-08,	The coefficient for T_I is 3.7464005231413466e-08
The coefficient for O_P is -8.70 ,	The coefficient for M_P is -6.18
The coefficient for F_P is 27.42,	The coefficient for TAX is -14.60
The coefficient for GOV_S is -8.57 ,	The coefficient for N_S is 3.70
The coefficient for F_R is -2.56 ,	The coefficient for T_P is -2.53

Training Score(r^2)	92.62
Test Score(r^2)	92.60
RMSE on Train	879.95
RMSE on Test	1026.31

Table 6:LR Model 2 accuracy scores

R^2 is not a reliable metric as it always increases with addition of more attributes even if the attributes have no in fluence on the predicted variable. Instead we can use adjusted R^2 which removes the statistical chance hence imp roving R^2 score . Scikit does not provide a facility for adjusted R^2 so the stats model is utilised, a library that give s results similar to what is obtain in R language along with F statistic and the T statistic value to understand the v ariables statistical significance even better . This library expects the X and Y to be given in one single data frame , hence the X&Y data is merged using concat function .
Now, with the 'stats models' using ordinary least square method we try to build the model and pull out the intercepts and coefficients to compare the regression models.

Ordinary Least Square Method

Dep. Variable:	NASDAQ	**R-squared:**	0.926
Model:	OLS	**Adj. R-squared:**	0.92
Method:	Least Squares	**F-statistic:**	143.5
Date:	Wed, 18 Jan 2023	**Prob (F-statistic):**	7.91E-83
Time:	14:07:50	**Log-Likelihood:**	-1434.8
No. Observations:	175	**AIC:**	2900
Df Residuals:	160	**BIC:**	2947
Df Model:	14		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-1.40E+04	4319.324	-3.243	0.001	-2.25E+04	-5476.07
U_R	-242.8532	94.083	-2.581	0.011	-428.657	-57.049
I_P_I	-107.5583	44.12	-2.438	0.016	-194.691	-20.425
Real_GDP	0.5621	0.316	1.778	0.077	-0.062	1.186
C_P_I	105.5078	20.863	5.057	0	64.305	146.71
T_E	-7.86E-08	1.16E-08	-6.793	0	-1.01E-07	-5.57E-08
T_I	3.75E-08	8.27E-09	4.529	0	2.11E-08	5.38E-08
O_P	-8.7026	5.404	-1.61	0.109	-19.375	1.97
M_P	-6.1784	2.26	-2.734	0.007	-10.642	-1.715
F_P	27.4163	10.756	2.549	0.012	6.175	48.658
TAX	-14.5972	8.753	-1.668	0.097	-31.884	2.689
GOV_S	-8.566	3.741	-2.29	0.023	-15.954	-1.178
N_S	3.6963	2.758	1.34	0.182	-1.751	9.144
F_R	-2.5559	1.216	-2.103	0.037	-4.956	-0.155
T_P	-2.5319	1.661	-1.525	0.129	-5.812	0.748

Omnibus:	9.831	**Durbin-Watson:**	1.95
Prob(Omnibus):	0.007	**Jarque-Bera (JB):**	10.464
Skew:	0.472	**Prob(JB):**	0.00534
Kurtosis:	3.738	**Cond. No.**	1.33E+13

Table 7:Statistical measures of regression model

H0 : Independent variables have no impact Dependent variable.
Ha : Independent variables have an impact on Dependent variables . (alpha= 0.05)

Based on the above hypothesis the p-values were derived and it is seen that attributes such as Real_GDP, Oil Prices, National Security, TAX and trade policy have value higher than alpha (0.05) therefore, statistically their coefficients are not reliable, hence are called poor predictors of Nasdaq.

The F-statistic value of 143.5, and the low Prob (F-statistic) value of 7.91E-83 indicate that the model is statistically significant.

Adjusted R^2 is 92 which is similar to the model seen in sk-earn.
Prob(JB) is 0.00534 less than alpha 0.05 indicating the overall model is good for deployment minus the insignificant variables.
Durbin-Watson score is 1.95 which is close to 2 indicating there is no autocorrelation i.e. residuals are not correlated.

The intercept of the model is -1.40E+04 which simply means when X (independent variables) are 0 then the price of the stock will decrease by decrease by -1..40E+04 . The Linear equation is :

*Nasdaq = (-1.40E+04) * Intercept + (-242.85) * U_R + (-107.56) * I_P_I + (0.56) *Real_GDP + (105.50)*C_P_I + (-7.86E-08)*T_E + (3.75E_08)*T_I + (-8.70)*O_P + (-6.18)*M_P + (27.42)*F_P + (-14.60)*TAX + (-8.57)*GOV_S + (3.70)*N_S + (-2.56)*F_R + (-2.53)*T_P*

Remarks: The significant influencers of Nasdaq stock price in this model is Unemployment Rate , Industrial Production Index and Consumer Price Index. If the unemployment rate increases this will reduce the stock price by 242.85; similarly a unit change in Industrial Production Index will decrease the stock price by 107.55 and so on keeping all other factors constant. However, if there is an increase in consumer price index the stock market prices increase by 105.50 as result of inflation.

However to understand the reliability of the coefficients and the model another model is built removing the variables with high p-value which are useless for interpretation as the hypothesis suggest these variables are not a significant predictor of Nasdaq.

LR MODEL 3 -Removing Poor Predictors

For this particular model regression analysis was performed using unscaled data on 9 independent variables hence removing the attributes which were not of statistical significance and keeping Nasdaq stock index as the dependent variable. Below is the result after performing regression on sk-learn and OLS model.

Training Score($r^{2)}$	92.01
Test Score($r^{2)}$	91.48
RMSE on Train	915.84
RMSE on Test	1101.42

Table 8:LR Model 3 accuracy score

Dep. Variable:	NASDAQ	**R-squared:**	0.92
Model:	OLS	**Adj. R-squared:**	0.916
Method:	Least Squares	**F-statistic:**	211.1
Date:	Wed, 18 Jan 2023	**Prob (F-statistic):**	1.04E-85
Time:	15:48:02	**Log-Likelihood:**	-1441.8
No. Observations:	175	**AIC:**	2904
Df Residuals:	165	**BIC:**	2935
Df Model:	9		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	-1.30E+04	4276.109	-3.04	0.003	-2.14E+04	-4556.77
U_R	-376.6755	75.354	-4.999	0	-525.458	-227.893
I_P_I	-106.0815	39.967	-2.654	0.009	-184.993	-27.17
C_P_I	150.5358	9.086	16.567	0	132.596	168.476
T_E	-8.73E-08	1.10E-08	-7.905	0	-1.09E-07	-6.55E-08
T_I	3.21E-08	7.76E-09	4.136	0	1.68E-08	4.74E-08
M_P	-6.3883	1.933	-3.305	0.001	-10.205	-2.572
F_P	13.26	3.3	4.019	0	6.745	19.775
GOV_S	-7.4443	3.531	-2.108	0.037	-14.416	-0.472
F_R	-3.1101	1.214	-2.562	0.011	-5.507	-0.714

Omnibus:	13.101	**Durbin-Watson:**	1.916
Prob(Omnibus):	0.001	**Jarque-Bera (JB):**	15.92
Skew:	0.523	**Prob(JB):**	0.000349
Kurtosis:	4.044	**Cond. No.**	1.28E+13

Table 9:Statistical measures of regression model

Based on the hypothesis and the p-values, it is seen that all attributes have p-values less than alpha (0.05) hence the null hypothesis is rejected and say that statistically their coefficients are reliable, therefore can be called significant predictors of Nasdaq.

Adjusted R^2 is 92 which is similar to the model seen in sk learn.
Prob(JB) is 0.00034 less than alpha 0.05 indicating the overall model is good for deployment.
Durbin-Watson score is 1.95 which is close to 2 indicating there is no autocorrelation i.e residuals are not correlated.
The F-statistic value of 211.1, and the low Prob (F-statistic) value of 1.04E-83 indicate that the model is statistically significant.
The intercept of the model is -1.30E+04 which simply means when X (independent variables) are 0 then the price of the stock will decrease by decrease by -1..30E+04 which is almost 0. The Linear equation is :

*Nasdaq = (-1.30E+04) * Intercept + (-376.68) * U_R + (-106.01) * I_P_I + (150.54)*C_P_I + (-8.73E-08)*T_E + (3.21E_08)*T_I + (-6.39)*M_P + (13.26)*F_P + (-7.44)*GOV_S + (-3.11)*F_R*

Remarks: Based on the analysis, it appears that the factors impacting the Nasdaq stock market price in this model have undergone some shifts, with the consumer price index playing a more prominent role as indicated by its higher coefficients compared to the I_P_I index in Model 2. Additionally, evaluation of Model 3 indicates that it has a level of accuracy that is comparable to that of Model 2, with statistical analysis this time proving that all attributes included in Model 3 are significant.

From the analysis of all the linear regression models constructed, it appears that Model 3 is the optimal option. This is evidenced by its high R2 value of 92%, indicating a strong correlation between the model's predictions and the actual Nasdaq stock prices. This model effectively utilizes the unemployment rate, consumer price index, and industrial production index as key predictors of Nasdaq stock prices. These factors were identified as the most important ones in the analysis. However, there high RMSE scores indicates there is room for further improvement.

Artificial Neural Network

The data was segmented into two distinct subsets, one for training and one for validation. Utilizing a method akin to linear regression, the training subset was employed to construct the model, and the validation subset was utilized to evaluate the model's precision. The Artificial Neural Network regression technique was applied to the data, enabling prediction of outcomes. Subsequently, various metrics such as R2 score and RMSE were calculated to assess model performance.

Model 1: Under this model various hyperparameters to tune the model were used such as :

- Hidden Layer Size = (200)
- random state = 123
- max iteration=10000

The results of the model were as under :

Train RMSE	365.39
Test RMSE	457.57
R^2 Training Score	98.83
R^2 Test Score	98.29

Table 10:ANN model 1 accuracy score

Model 2: Under this model the GridSearchCv was utilized and the cross validation functions to optimise the model by tuning the hyperparameters as the performance of a model is closely tied to the values of these parameters.
Results of the new optimum hyperparameters were as below:

- Hidden Layer Size = (15,15)
- Activation = 'relu'
- Solver = 'adam'
- random state = 123,
- max iteration=10000
- CV= 3

The results of this model were as under:

Train RMSE	295.06
Test RMSE	424.74
R^2 Training Score	99.23
R^2 Test Score	98.52

Table 11:ANN model 2 accuracy score

Remarks: The R2 score for both models on the training set and test set are both close to 99 it suggests that the model is performing very well on both the training and test data. A high R2 score on both the training and test sets suggests that the model has a good fit to the data and is able to generalize well to unseen data. This is a good indication that the model is not overfitting or underfitting the data. However, even though the R2 score is high, it would be important to check other performance metrics such as root mean squared error or mean absolute error to have a more complete picture of the model performance. The RSME is slightly better in the second model .

The Second model seems to be better under artificial neural network , with good R^2 scores on train and test data , RMSE scores also lesser in model 2 show better accuracy.

Decision Tree Regression

The data was divided into two distinct groups, one for training and one for validation purposes. Similar to previous models, the training group was utilized to construct the model, and the validation group was used to evaluate the model's accuracy. The Decision Tree regression method was applied to the data, which enabled the prediction of outcomes. Afterwards, a variety of metrics such as R2 score and RMSE were calculated to evaluate the performance of the model.

Model 1: Under this model various hyperparameters to tune the model were used such as :

- random state = 123

The results of the model were as under :

Train RMSE	0.00
Test RMSE	369.37
R^2 Training Score	1.00
R^2 Test Score	98.88

Table 12: DT Model 1 accuracy score

Model 2: Under this model the GridSearchCv and the cross validation functions were utilised to optimise the model by tuning the hyperparameters as the performance of a model is closely tied to the values of these parameters.
Results of the new optimum hyperparameters were as below:

- Max_depth = 10
- Min_sample_leaf = 5
- Min_sample_split = 2
- random state = 123,
- estimator =dtr
- CV= 3

The results of this model were as under:

Train RMSE	280.32
Test RMSE	415.87
R^2 Training Score	99.30
R^2 Test Score	98.58

Table 13:DT Model 2 accuracy score

Remarks: The RMSE scores are better in Model 2 because of the less difference between them, and the RMSE and R^2 scores of Model 1 seem to reflect that the model is overfitting. This means that the model is fitting the noise in the training data rather than the underlying pattern, and thus is not generalizing well to unseen test data.

Random Forest Regression

The data was divided into two distinct groups, one for training and one for validation purposes. Similar to previous models, the training group was utilized to construct the model, and the validation group was used to evaluate the model's accuracy. The Random Forest regression method was applied to the data, which enabled the prediction of outcomes. Afterwards, a variety of metrics such as R2 score and RMSE were calculated to evaluate the performance of the model.

Model 1: Under this model various default hyperparameter to tune the model was used such as :

- random state = 123

The results of the model were as under :

Train RMSE	176.81
Test RMSE	308.86
R^2 Training Score	99.73
R^2 Test Score	99.22

Table 14:RF Model 1accuracy score

Model 2: Under this model the GridSearchCv and the cross validation functions were utilised to optimise the model by tuning the hyperparameters as the performance of a model is closely tied to the values of these parameters.
Results of the new optimum hyperparameters were as below:

- Max_depth = 7
- Max_features = 10
- Min_sample_leaf = 1
- Min_sample_split = 3
- random state = 123,
- n_estimator = 50
- CV= 3

The results of this model were as under:

Train RMSE	176.24
Test RMSE	338.69
R^2 Training Score	99.72
R^2 Test Score	99.06

Table 15:RF Model 2 accuracy score

Remarks: The RMSE scores are better in Model 1 because of the less difference between them, and the RMSE and R^2 scores of Model 1 and Model 2 are very similar. Since the RMSE scores are less in model 1 which means less error, we will choose Model 1 between both the models.

Macroeconomic Variables	Importance
C_P_I	0.623015
Real_GDP	0.272483
T_I	0.064212
U_R	0.010852
T_P	0.00632
T_E	0.005204
I_P_I	0.005095
O_P	0.003341
F_P	0.00233
TAX	0.002289
M_P	0.00187
N_S	0.001295
F_R	0.00085
GOV_S	0

Table 16:Feature Importance RF Model 1

Using the function in model 1 called important features under random forest the list of macroeconomic variables were extracted with its details of importance level and it is observed that consumer price index has the most impact on Nasdaq price followed by GDP . The remaining macroeconomic variables have lower importance scores, with Unemployment Rate being the fourth most important variable, Trade_Policy, Trade_Exports, IPI, O_P, F_P , TAX, M_P, N_S and F_R having a less importance. Government Spending has no importance on the stock returns.

4.6 RESULTS

- Based on the univariate and bivariate analysis, it was identified that the distribution of all variables in the data is not normal and there are outliers present in all attributes, with the exception of GDP, Consumer Price Index, Industrial Production Index, Oil Prices and Trade Exports. The observed trend and pattern among these variables suggest that the majority of these outliers can be attributed to the impact of the COVID-19 pandemic, during which America experienced a total shutdown.

- An initial evaluation of a conceptual framework model was conducted to assess the effect of independent variables on a dependent variable through the use of multivariate analysis techniques, such as scatter plots and correlation analysis. This allowed for the calculation of Pearson's coefficient and the determination of the strength of the relationship between the variables. The results indicated a strong positive correlation between the Nasdaq and various macroeconomic variables, including GDP, CPI, trade exports and imports, and a negative correlation with unemployment. The analysis suggests a correlation between macroeconomic variables and stock prices, with CPI, GDP, and trade exports and imports appearing to be the most influential factors on Nasdaq pricing.
  However, it should be noted that this model should not be relied upon without further statistical measures and significance testing.

- To improve the accuracy of the dataset and mitigate the presence of high multicollinearity as identified through correlation analysis and the Bartlett test of sphericity, a dimension reduction technique was implemented. Specifically, we conducted a Principal Component Analysis, which effectively grouped our variables into just three dimensions. Additionally, scaling and outlier treatment was performed to further enhance the precision of the results. The analysis revealed that economic growth, as represented by a composite variable comprising GDP, CPI, trade exports and imports, and the Industrial Production Index, emerged as the most significant factor influencing stock prices. The model demonstrated that for every unit increase in economic growth, there is a corresponding decrease of 0.281 units in stock prices, holding all other factors constant.

- Model three which did fairly well under Linear regression using Ordinary Least Square method after removing the insignificant variables (derived using p-value at significance level of 0.05) shows that there are higher impact of Unemployment rate on the stock price followed by Consumer price index and Industrial production index. The model demonstrated that for every unit increase in unemployment, there is a corresponding decrease of 376 units in stock prices, holding all other factors constant.
  Hypothesis testing reflected that the variables such as Oil prices , National security, Trade policy , GDP and Tax are poor predictors of Nasdaq stock prices as the p value for all these variables was higher than alpha 0.05.

- Other machine learning models were used such as decision tree, random forest, artificial neural network and tuned the models and hyperparameters to arrive at higher accuracy scores. GridsearchCv helped in identifying the best hyperparameters to be used. Random Forest regressor was chosen as the best model and the model predicted Consumer price index and GDP has the highest impact on Nasdaq stock prices followed by trade imports and unemployment rate. The performances of all model with and without hyperparameter tuning is as under:

	Train RMSE	Test RMSE	Training Score	Test Score
Decision Tree Regressor	0.000000	369.369810	1.000000	0.988858
Random Forest Regressor	176.813391	308.863002	0.997251	0.992209
ANN Regressor	365.394911	457.566926	0.988261	0.982902

Table 17:Model performance with default hyperparameter

	Train RMSE	Test RMSE	Training Score	Test Score
Decision Tree Regressor	280.320501	415.871804	0.993091	0.985876
Random Forest Regressor	176.237258	338.685780	0.997269	0.990632
ANN Regressor	295.056906	424.735635	0.992345	0.985267

Table 18:Model performance with hyperparameter tuning

Based on thorough analysis, it is recommended to utilize the Random Forest Regressor model 1 for deployment due to its high performance as evidenced by its strong R-squared and low root mean squared error scores. Additionally, our research and modelling efforts have highlighted there is significant impact and influence of Consumer Price Index on the Nasdaq stock exchange, as demonstrated through correlation analysis, principal component analysis, linear regression, and random forest regression.

4.7 Summary

In summary, this analysis aimed to assess the effect of various independent variables on the Nasdaq stock exchange through univariate, bivariate, and multivariate analysis techniques. It was found that the distribution of all variables in the data was not normal and there were outliers present, except for GDP, Consumer Price Index, Industrial Production Index, Oil Prices, and Trade Exports, which were attributed due to the impact of the COVID-19 pandemic. A conceptual framework model was initially evaluated to determine the strength of the relationship between the variables, resulting in a strong positive correlation between the Nasdaq and various macroeconomic variables, including GDP, CPI, trade exports and imports, and a negative correlation with unemployment. Dimension reduction techniques were applied to improve the accuracy of the dataset and mitigate the presence of high multicollinearity. Linear regression and other machine learning models were used to further analyse the data, with the best performing model being the Random Forest Regressor. The analysis suggests that Consumer Price Index, GDP, and trade exports and imports have the most significant impact on Nasdaq pricing. It is recommended to utilize the Random Forest Regressor model for deployment.

5. DISCUSSION

The discussion section of this study aims to interpret and analyse the findings in relation to the research questions and objectives. This section will also examine the findings in the context of the existing literature on the topic, and draw conclusions about the overall significance of the results. Overall, the discussion section will provide valuable insight into the factors that drive stock returns on the Nasdaq Stock Exchange and inform investment decisions in the stock market.

- CPI results is similar with Parmar, 2013; Venkatraja, 2014; Kumar, 2014; Asmy at el. 2010 and Adam 2008 i.e. when the CPI is rising, it typically indicates inflation, which can lead to higher prices for goods and services and higher profits for companies. This in turn can lead to an increase in stock prices, as investors anticipate that the company will continue to make a profit. However, if the inflation rate is too high, it can erode the purchasing power of consumers and make it difficult for companies to maintain profitability, leading to a decrease in stock prices. Hence it is stated that the Consumer Price Index has a positive and significant effect on stock returns for listed companies in our research.

- When unemployment is high, consumer spending and economic growth tend to decrease, which can lead to lower profits for companies and a decrease in stock prices. Additionally, high unemployment can indicate a weaker economy, which can lead to a decrease in investor confidence and a decrease in demand for stocks. Conversely, low unemployment can indicate a strong economy, which can lead to an increase in investor confidence and an increase in demand for stocks. As stated previously that the unemployment rate has a negative and significant effect on stock returns as scene in the linear regression model in this research.

- The IPI is a measure of the change in the total inflation-adjusted value of output produced by manufacturers, mines, and utilities. When the IPI is declining, it typically indicates a decrease in industrial activity, which can lead to lower profits for companies and a decrease in stock prices which can lead to a decrease in investor confidence and a decrease in demand for stocks. Hence stated that the Industrial Production Index has a negative and significant effect on stock returns as observed through a linear regression model. The results are similar to (Patel, 2012).

- Trade Exports & imports , Monetary & Fiscal Policy, Government Spending and Financial Regulations have a minor impact on stock returns and its statistically significant variables as seen in the regression model. Trade Imports and Fiscal policy have a positive relation meaning with one unit increase in these factors may lead to decrease in stock prices. On the other hand an increase in unit of Trade Exports, Monetary Policy, Government spending and financial regulations tends to decrease the stock prices as per the regression coefficient derived in LR model 3.

To conclude this study has practical implications for investors and policymakers in understanding the factors that drive stock returns on the Nasdaq Stock Exchange and making informed investment decisions. However, the study has its own limitations, and it is important that this analysis should be considered in light of the ongoing impact of the COVID-19 pandemic on the economy. It is recommended to consider these findings alongside other factors and conduct further research to fully understand the relationship between macroeconomic variables and stock prices. In summary, this study provides a valuable contribution to the literature on the relationship between microeconomic variables and stock prices on the Nasdaq Stock Exchange.

6. CONCLUSION

The analysis of the Nasdaq Stock Exchange revealed that the Consumer Price Index has a positive and significant impact on stock returns for listed companies during the period 2002 -2022. Conversely, the Industrial Production Index and Unemployment rate have a negative and significant effect on stock returns as observed through our linear regression model. This highlights the importance of monitoring and controlling macroeconomic variables in order to maintain stability in stock returns and promote growth in stock market investments.

It is hypothesized that the stock market serves as a leading indicator of the real economy, as market trends tend to precede economic cycles by a period of six to twelve months. This theory is supported by our study of the correlation between the Nasdaq market price index and various macroeconomic variables, as illustrated through graphical representation and scatter plots. Additionally, statistical analysis through the calculation of Pearson correlation coefficients confirms the significance of these relationships. Our findings indicate a strong positive and significant relationship between the US stock market and CPI and GDP, and a strong negative and significant relationship with unemployment and IPI, which aligns with previous research.

7. RECOMMENDATIONS

Based on the findings of this study, some recommendations for further research and practical applications include:

- Conduct further statistical measures and significance testing to confirm the validity of the model, which suggests a correlation between macroeconomic variables and stock prices on the Nasdaq Stock Exchange.
- Monitor the impact of the COVID-19 pandemic on the economy and how it affects the distribution of data, outliers and the relationship between macroeconomic variables and stock prices.
- Investigate the effect of other macroeconomic variables such as interest rate and inflation on stock prices, as they may also have a significant impact.
- Utilize the Random Forest Regressor model 1 for prediction and analysis of stock prices on the Nasdaq Stock Exchange, as it has been shown to have high performance and low error rates.
- Consider the impact of Consumer Price Index, GDP, trade exports and imports, and unemployment rate when making investment decisions in the stock market.
- Use other machine learning models and techniques to improve the accuracy of predictions and analysis of stock prices on the Nasdaq Stock Exchange.
- Investigate the effect of other global and local factors on stock prices, such as political stability, natural disasters, and technological advancements.
- Consider the results of this study in the context of other stock markets and economies around the world and investigate whether similar relationships exist between macroeconomic variables and stock prices in other regions.

REFERENCES

- *United States Full Year GDP Growth - 2021 Data - 2022 Forecast - 1950-2020 Historical*. (n.d.). Tradingeconomics.com. https://tradingeconomics.com/united-states/full-year-gdp-growth
- *Sharing insights elevates their impact*. (n.d.). S&P Global. https://www.spglobal.com/marketintelligence/en/mi/products/us-monthly-gdp-index.html
- *Economic Policy Uncertainty Index*. (2012). Policyuncertainty.com. https://www.policyuncertainty.com/us_monthly.html
- Trading Economics. (2022). *Crude oil*. Tradingeconomics.com. https://tradingeconomics.com/commodity/crude-oil
- *Consumer Price Index Data from 1913 to 2019*. (2019, May 10). US Inflation Calculator. https://www.usinflationcalculator.com/inflation/consumer-price-index-and-annual-percent-changes-from-1913-to-2008/
- *U.S. Sector Indices, Charts, Prices.* (n.d.). Barchart.com. Retrieved January 24, 2023, from https://www.barchart.com/stocks/indices/us-sectors?viewName=main
- Masdah, S. R., Djirimu, M., Anam, H., Tombolotutu, A., Laturama, Y., & Rahman, F. (2022). Determinant Factors Effect of the Indonesia's Banking Sector Share Prices in 2005 QII - 2019 QIV Period. *JOURNAL of ECONOMICS, FINANCE and MANAGEMENT STUDIES*, *05*(12). https://doi.org/10.47191/jefms/v5-i12-40
- Hwang, T. J. (2013). Stock Market Returns and Clinical Trial Results of Investigational Compounds: An Event Study Analysis of Large Biopharmaceutical Companies. *PLoS ONE*, *8*(8), e71966. https://doi.org/10.1371/journal.pone.0071966
- Al-Abdallah, S., Ibrahim, N., & Aljarayesh, A. (2017). Licensed under Creative Common INFLUENCE OF INTEREST RATE, EXCHANGE RATE AND INFLATION ON COMMON STOCK RETURNS OF AMMAN STOCK EXCHANGE, JORDAN. *International Journal of Economics, Commerce and Management United Kingdom*, *10*. https://ijecm.co.uk/wp-content/uploads/2017/10/51037.pdf
- Pradhan, R., Arvin, M., Samadhan, B., & Taneja, S. (n.d.). *THE IMPACT OF STOCK MARKET DEVELOPMENT ON INFLATION AND ECONOMIC GROWTH OF 16 ASIAN COUNTRIES: A*

PANEL VAR APPROACH. Retrieved March 20, 2022, from https://www.usc.es/economet/reviews/aeid13116.pdf

- Permatasari, A., & Yuniarsih, N. (2020). ANALYSIS OF FACTORS THAT AFFECT COMPOSITE STOCK PRICE INDEX MOVEMENT IN INDONESIA STOCK EXCHANGE. *Asian Journal of Management Sciences & Education*, *9*(2). http://www.ajmse.leena-luna.co.jp/AJMSEPDFs/Vol.9(2)/AJMSE2020(9.2-09).pdf
- Samontaray, D. P., Nugali, S., & Sasidhar, B. (2014). A Study of the Effect of Macroeconomic Variables on Stock Market: Saudi Perspective. *International Journal of Financial Research*, *5*(4). https://doi.org/10.5430/ijfr.v5n4p120
- *Arouri, M.E.H. and Rault, C. (2010) Oil Prices and Stock Markets What Drives What in the Gulf Corporation Council countries. Economie Internationale, 56, 41-56. - References - Scientific Research Publishing*. (n.d.). Www.scirp.org. Retrieved January 24, 2023, from https://www.scirp.org/(S(351jmbntvnsjt1aadkposzje))/reference/referencespapers.aspx?referenceid=1834897
- Domingos P (2012) A few useful things to know about machine learning. Commun ACM 55(10):78–87
- Rastogi, S. (2014). The financial crisis of 2008 and stock market volatility - analysis and impact on emerging economies pre and post crisis. *Afro-Asian J. Of Finance and Accounting*, *4*(4), 443. https://doi.org/10.1504/aajfa.2014.067017
- Hosseini, S. M., Ahmad, Z., & Lai, Y. W. (2011). The Role of Macroeconomic Variables on Stock Market Index in China and India. *International Journal of Economics and Finance*, *3*(6). https://doi.org/10.5539/ijef.v3n6p233
- Chee, R., Chia, J., & Shiok Ye Lim. (2015). MALAYSIAN STOCK PRICE AND MACROECONOMIC VARIABLES: AUTOREGRESSIVE DISTRIBUTED LAG (ARDL) BOUNDS TEST. *Kajian Malaysia*, *33*(1), 85–103. http://web.usm.my/km/33(Supp.1)2015/Art.6%20(85-103).pdf
- Forson, J. A., & Janrattanagul, J. (2014). Selected Macroeconomic Variables and Stock Market Movements: Empirical evidence from Thailand. *Contemporary Economics*, *8*(2), 157–174. https://doi.org/10.5709/ce.1897-9254.138
- Wikipedia Contributors. (2022, August 28). *Lehmann–Scheffé theorem*. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Lehmann%E2%80%93Scheff%C3%A9_theorem

- Hastie, T., & Tibshirani, R. (1990). Exploring the Nature of Covariate Effects in the Proportional Hazards Model. *Biometrics*, *46*(4), 1005–1016. https://doi.org/10.2307/2532444
- Vychytilová, J., Pavelková, D., Pham, H., & Urbánek, T. (2019). Macroeconomic factors explaining stock volatility: multi-country empirical evidence from the auto industry. *Economic Research-Ekonomska Istraživanja*, *32*(1), 3327–3341. https://doi.org/10.1080/1331677x.2019.1661003
- Kenton, W. (2021, October 6). *Analysis Of Variance - ANOVA*. Investopedia. https://www.investopedia.com/terms/a/anova.asp
- Sartorius. (2020, August 18). *What Is Principal Component Analysis (PCA) and How It Is Used?* Sartorius. https://www.sartorius.com/en/knowledge/science-snippets/what-is-principal-component-analysis-pca-and-how-it-is-used-507186
- Glen, S. (2021). *Correlation in Statistics: Correlation Analysis Explained*. Statistics How To. https://www.statisticshowto.com/probability-and-statistics/correlation-analysis/
- Kenton, W. (2021, August 4). *Understanding the Durbin Watson Statistic*. Investopedia. https://www.investopedia.com/terms/d/durbin-watson-statistic.asp#:~:text=The%20Durbin%20Watson%20statistic%20is
- Mujtaba, H. (2020, September 29). *An Introduction to Grid Search CV | What is Grid Search*. GreatLearning. https://www.mygreatlearning.com/blog/gridsearchcv/

APPENDIX

Research Proposal

TABLE OF CONTENTS

ABSTRACT

The purpose of this study is to investigate the impact of macroeconomic variables on the stock prices of the NASDAQ index. The research focuses on the correlation between unemployment rate, industrial production index, gross domestic product, consumer price index, trade exports and imports, oil price, monetary and fiscal policy, tax, government spending, national security, financial regulations, and trade policy and the stock prices of NASDAQ. The study uses a dataset of 251 observations spanning from year 2000 to 2021. The data is analysed using various statistical techniques such as histograms, box plots, linear regression, decision tree regression, random forest regression, and artificial neural networks.

INTRODUCTION

The stock market is a crucial indicator of the overall economic health of a country. Changes in the stock market can have a significant impact on individual investors, as well as on the economy as a whole. One of the most important factors that can affect stock market performance is macroeconomic variables. These macroeconomic variables can have a direct or indirect impact on the stock market. This research proposal aims to provide insights that can be useful for investors in making informed decisions. This study is significant because understanding the relationship between macroeconomic variables and stock returns can help investors to make better investment decisions and can also assist policymakers in formulating economic policies that support the growth of the stock market.

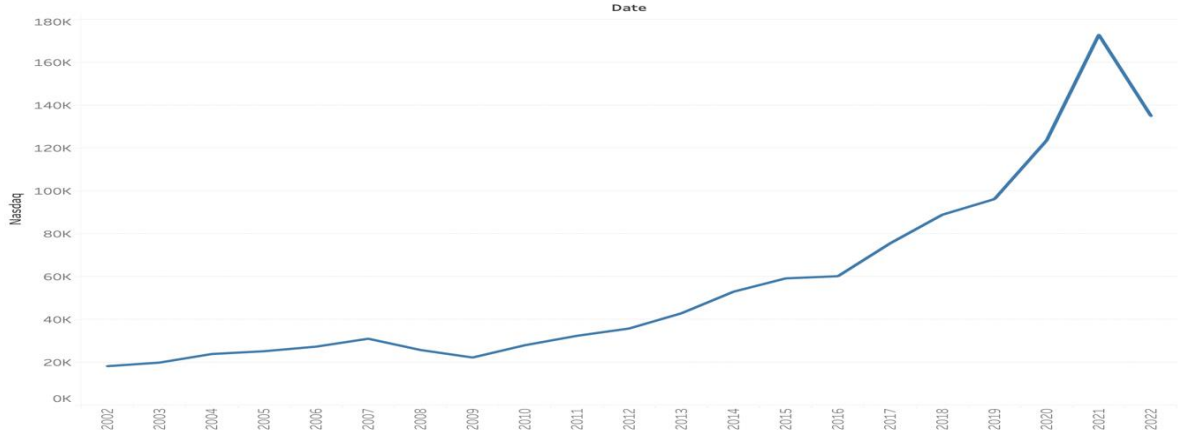
AIMS & OBJECTIVES

The stock market is a barometer of the overall health of an economy and is closely tied to the performance of macroeconomic variables. The objective of this research proposal is to examine the impact of macroeconomic variables on the NASDAQ stock price index. The study will focus on analysing the relationship between the NASDAQ stock price index and various macroeconomic variables such as Unemployment rate (U_R), Industrial Production Index (I_P_I), Real Gross Domestic Product (Real_GDP), Consumer Price Index (C_P_I), Trade Exports (T_E), Trade Imports (T_I), Oil Price (O_P), Monetary Policy index (M_P), Fiscal Policy Index (F_P), TAX index (TAX), Government Spendings (GOV_S), National Security (N_S), Financial Regulations (F_R), and Trade Policy (T_P)

SCOPE OF STUDY

The findings of this study will be of importance to investors as they can use the information to inform their investment decisions. The study will provide a deeper understanding of the relationship between macroeconomic variables and stock prices, which can help investors to identify potential risks and opportunities in the stock market. Additionally, the results of this research can be used by investors to make more informed decisions on when to buy and sell stocks. Furthermore, this study will be a valuable resource for financial analysts and economists as it will provide a comprehensive analysis of the factors that drive the stock market, which can be used to improve forecasting models and inform policy decisions. This research will help to provide a better understanding of the stock market and its relationship to the broader economy, which will be of great importance to investors and policy makers alike.

Nasdaq stock return from 2002 to 2022



Literature Review

Previous Literature

The study conducted by Krisna & Wirawati (2013) found that inflation has a positive and significant effect on stock returns, which differs from the findings of Tangjitprom (2013) and Azar (2013) that the effect of inflation on negative stock returns is not significant. The exchange rate, often used as a reference for investors and economists, was also studied and research by Al-abdallah & Aljarayesh (2017) found it does not significantly influence the stock index, while other studies such as Amansyah (2013), Liauw (2013), Silim (2012), and Sisbintari (2009) found that it has a significant effect on stock price index. Additionally, research by Gunawan & Wibowo (2012) found that the Rupiah/US $ exchange rate has no significant effect on positive stock returns, differing from the findings of Prasetiono (2010) which found that interest rates have a negative and non-significant effect on stock return. The impact of GDP on the movement of the Stock Price Index also varied in research, with Adam et al (2013) and Sisbintari (2009) finding it has a significant effect and Kewal (2012) finding the opposite. Furthermore, Pícha (2017) found that money supply affects stock market and Suwandy (2014) found that world oil prices have a significant effect on stock return.

Research Questions

1. Does macroeconomic factor influence the drops and rise of Nasdaq stock prices?
2. What is the trend of Nasdaq's stock prices in comparison with the macroeconomic variables?
3. Which macroeconomic variables are the most significant indicator of stock returns?
4. Does the increase in the unemployment of US residents have a direct impact on the stock returns?
5. Does world oil prices effect the stock returns?
6. Is there a dependency of stock return based on industrial productions in the United States of America?
7. Does the Consumer price index effect the stock returns?
8. Is there an impact on stock return prices based on the increase or decrease of united states international trade exports & imports?
9. Does the monetary or fiscal policies have any positive or negative impact on stock prices?
10. Does the National security or Financial regulations of US have an impact on stock exchange market?
11. Does the US Government spendings have an influence on Nasdaq stock exchange?

Data Collection

- Research Question : Does macroeconomic factor influence the drops and rise of Nasdaq stock prices?

There are several ways , one of which would be to gather as much data that is available on macroeconomic variables for a specific period that we would want to analyse , it could be daily, monthly, quarterly or yearly data depending on the type of analysis we want to run. This can be collected through sources such as tradingeconomics.com, insights.ceicdata.com, world bank, etc. Then collecting the data for the chosen stock exchange which in this study is Nasdaq stock exchange taken from yahoo finance website.

- Research Question: What is the trend of Nasdaq's stock prices in comparison with the macroeconomic variables?

To address this particular research question on the Nasdaq stock prices trend we have collected the data for Nasdaq composite index which includes 3700 stocks. The data chosen could be either the opening stock, closing stock , adjusted closing stock . The data being used should be any one of the earlier mentioned as our dependent variable is going to only one hence we have gone ahead and extracted the monthly adjusted closing prices data.

- Research Question: Which macroeconomic variables are the most significant indicator of stock returns?

To address this particular research question the data of United States of Americas macroeconomic factors need to be collected. This can be obtained through various sources like Office for national statistics, world bank , trade economics etc. The data collected is to be of the same frequency as the data extracted for stock prices i.e. either daily, monthly , quarterly, yearly depending on the availability . All macroeconomic indicators selected in this study have been chosen in such a way that we have the monthly data available to us for all the factors.

Proposed Methodology

Methods

In this research proposal, we propose to use linear regression and machine learning techniques such as decision tree regressor, random forest regressor and artificial neural network (ANN) regressor to analyse the impact of macroeconomic variables on stock returns. We will use historical data of 14 macroeconomic variables including unemployment rate, industrial production index, gross domestic product, consumer price index, trade exports and imports, oil price, monetary policy index, fiscal policy index, tax index, government spending, national security, financial regulations, and trade policy. These variables will be collected over a period of 251 months from January 2002 to December 2022. The data will be split into training and testing sets and the models will be trained and tested on these sets to evaluate their performance. We will also use principal component analysis to identify the underlying patterns in the data and to reduce the dimensionality of the data. Finally, we will use the best performing model to make predictions about future stock returns and to analyse the impact of specific macroeconomic variables on stock returns. The results of this study will be useful for investors as it will provide insights into the factors that drive stock returns and can help them make informed investment decisions.

Data Analysis

How data analysis will address each research question

Data analysis conducted in the study will address the research question i.e. influence of macroeconomic variable on stock prices by:

Firstly analysing the data of macroeconomic variables and stock prices to determine the trends and patterns. In this particular analysis we will make use of histogram, box plots and dual axis time series plots to compare all the macroeconomic variables trend with Nasdaq stock price and to see that if factors like Unemployment rate, industrial production index, consumer price index, international trade exports and imports have positive or negative trends.

Next we will look at the correlation matrices wherein we will see if there is significant positive relation between the Industrial production index , consumer price index , trade exports ,imports and Nasdaq .We will also perform hypothesis (ANOVA) testing to understand the significance of variables.

Finally we will build multiple machine learning models such as Linear Regression, Decision Tree Regression, Random Forest Regression and Artificial Neural Networks to select the best model to be used for deployment and have a detailed understanding of the actual impact of these independent variables on stock prices also to know the significant factor affecting the stock prices . We will also build a LR model after performing Principal component analysis to reduce the dimensions and multicollinearity between variables.

LIMITATIONS OF THE STUDY

One limitation that may affect our research paper is the impact of the COVID-19 pandemic on the economy and stock market. The pandemic has caused unprecedented disruptions and changes to economic activity, which may make it difficult to accurately interpret the relationship between macroeconomic variables and stock returns. Additionally, the pandemic may have caused drastic changes in certain macroeconomic variables, such as unemployment rates and consumer prices, which may not be representative of long-term trends. This could affect the generalizability of our findings and make it difficult to make predictions about future market conditions. Additionally, the pandemic has caused a high level of uncertainty and volatility in the stock market, which could make it difficult to obtain accurate data and make predictions. This could make it difficult to draw clear conclusions about the relationship between macroeconomic variables and stock returns.

ETHICAL IMPLICATIONS

In this research proposal, we must consider the ethical implications and risks associated with the collection and analysis of the data. As we will be analysing financial data, it is important to ensure that the data is collected and handled in a manner that is compliant with laws and regulations related to financial data privacy. Additionally, as the study may include sensitive information, it is important to have proper measures in place to protect participant confidentiality if needed which should not be as we are not conducting a primary research. Another ethical consideration is the potential impact of our research findings on the financial markets. It is crucial to ensure that the findings of this study are not used for insider trading or other unethical practices.

References

- Masdah, S. R., Djirimu, M., Anam, H., Tombolotutu, A., Laturama, Y., & Rahman, F. (2022). Determinant Factors Effect of the Indonesia's Banking Sector Share Prices in 2005 QII - 2019 QIV Period. *JOURNAL of ECONOMICS, FINANCE and MANAGEMENT STUDIES*, *05*(12). https://doi.org/10.47191/jefms/v5-i12-40
- Hwang, T. J. (2013). Stock Market Returns and Clinical Trial Results of Investigational Compounds: An Event Study Analysis of Large Biopharmaceutical Companies. *PLoS ONE*, *8*(8), e71966. https://doi.org/10.1371/journal.pone.0071966
- Al-Abdallah, S., Ibrahim, N., & Aljarayesh, A. (2017). Licensed under Creative Common INFLUENCE OF INTEREST RATE, EXCHANGE RATE AND INFLATION ON COMMON STOCK RETURNS OF AMMAN STOCK EXCHANGE, JORDAN. *International Journal of Economics, Commerce and Management United Kingdom*, *10*. https://ijecm.co.uk/wp-content/uploads/2017/10/51037.pdf
- Pradhan, R., Arvin, M., Samadhan, B., & Taneja, S. (n.d.). *THE IMPACT OF STOCK MARKET DEVELOPMENT ON INFLATION AND ECONOMIC GROWTH OF 16 ASIAN COUNTRIES: A PANEL VAR APPROACH*. Retrieved March 20, 2022, from https://www.usc.es/econmet/reviews/aeid13116.pdf
- Permatasari, A., & Yuniarsih, N. (2020). ANALYSIS OF FACTORS THAT AFFECT COMPOSITE STOCK PRICE INDEX MOVEMENT IN INDONESIA STOCK EXCHANGE. *Asian Journal of Management Sciences & Education*, *9*(2). http://www.ajmse.leena-luna.co.jp/AJMSEPDFs/Vol.9(2)/AJMSE2020(9.2-09).pdf
- *Sharing insights elevates their impact*. (n.d.). S&P Global. https://www.spglobal.com/marketintelligence/en/mi/products/us-monthly-gdp-index.html
- *Economic Policy Uncertainty Index*. (2012). Policyuncertainty.com. https://www.policyuncertainty.com/us_monthly.html
- Trading Economics. (2022). *Crude oil*. Tradingeconomics.com. https://tradingeconomics.com/commodity/crude-oil

- *Consumer Price Index Data from 1913 to 2019*. (2019, May 10). US Inflation Calculator. https://www.usinflationcalculator.com/inflation/consumer-price-index-and-annual-percent-changes-from-1913-to-2008/
- *U.S. Sector Indices, Charts, Prices*. (n.d.). Barchart.com. Retrieved January 24, 2023, from https://www.barchart.com/stocks/indices/us-sectors?viewName=main

----------xxxxxxxxxxxxxxxxxx----------