Major Research Project

Econometric Model for Predicting Indexes – Using ARMA, VAR AND VECM Approach

Submitted by

Divik Mehrotra

2K21/DMBA/46

Under the Guidance of

Mr. Dhiraj Kumar Pal

Assistant Professor

Delhi School of Management

DELHI SCHOOL OF MANAGEMENT

Delhi Technological University

Bawana Road Delhi 110042

CERTIFICATE FROM THE INSTITUTE

This is to certify that **Divik Mehrotra, 2K21/DMBA/46** has submitted the major research report titled **"Econometric Model for Predicting Indexes – Using ARMA, VAR AND VECM Approach"** under the guidance of Mr. Dhiraj Kumar Pal as a part of Master of Business Administration (MBA) curriculum of Delhi School of Management, Delhi Technological University, New Delhi during the academic year 2022-23.

Signature of the Guide	Signature of Head of Department
Mr.Dhiraj Kumar Pal	Dr. Archana Singh

DECLARATION

I, **Divik Mehrotra** student of Delhi School of Management, Delhi Technological University hereby declare that the "Econometric Model for Predicting Indexes – Using ARMA, VAR AND VECM Approach" submitted in partial fulfilment of the requirements for the award of the degree of Master of Business Administration (MBA) is the original work conducted by me. I also confirm that neither I nor any other person has submitted this project report to any other institution or university for any other degree or diploma. I further declare that the information collected from various sources has been duly acknowledged in this project.

Divik Mehrotra

2K21/DMBA/46

ACKNOWLEDGEMENT

Before we get into the crux of the matter, I'd want to express my gratitude to those who have been a part of this project since its start. This project's writing has been one of the most major academic obstacles I've faced, and it would not have been accomplished without the help, patience, and advice of the people involved.

It gives me immense pleasure in presenting this project report on "Econometric Model for Predicting Indexes – Using ARMA, VAR AND VECM Approach". I want to express my gratitude to Mr. Dhiraj Pal Sir, Assistant Professor, Delhi School of Management who have aided me since the beginning of this project and for his invaluable advice and timely ideas.

Finally, I would like to thank our department's head, Dr. Archana Singh Delhi School of Management DTU for her recommendations regarding the structure of the report which led to the creation of this report. The success of this project is the product of my hard work and perseverance, as well as the assistance of my mentors and HOD.

Divik Mehrotra

2K21/DMBA/46

EXECUTIVE SUMMARY

A financial market is a place where purchase and sale of financial products ranging from derivatives, stocks, bonds, currencies, and commodities are conducted. These financial markets offer individuals, corporations, and organisations a place to invest, borrow, and lend money, manage risk, and make money. In a market for financial instruments, both buyers and sellers come together to exchange financial securities at a price determined by supply and demand in the market. The choice is an exchange, where trading occurs through a centralised platform, or an over-the-counter (OTC) market, in which deals occur directly between buyers and sellers, can be used to organise the market.

An index is a statistical indicator of how well a specific area of the financial market has performed. The index measures the overall performance of a collection of stocks or other financial instruments that are either comparable in character to one another or that are part of the same sector or industry. By monitoring the price alterations of the underlying assets over time, the value of an index is determined. Stock market indices, which assess a set of equities that are traded on a stock exchange, are the most often used indexes. Indices act as benchmarks for evaluating how well portfolios, mutual funds, and other financial instruments perform. The index may be used by investors to compare the returns on their investments to the returns on the whole market.

Indices act as benchmarks for evaluating how well portfolios, The financial market in India is a system that allows people, businesses, and organisations to purchase and sell financial items such bonds, derivatives, stocks, commodities, and currencies. Primary market and secondary market are the two general divisions of the Indian financial market, which is governed by Securities and Exchange Board of India (SEBI).

Companies that want to raise money from the public issue new securities, such as bonds, debentures, and shares, on the main market. On the other hand, investors exchange these assets on the secondary market. The National Stock Exchange (NSE) and the Bombay Stock Exchange (BSE) are the two main stock exchanges in India. The BSE Sensex, which monitors the performance of 30 sizable, well-known firms listed on the Bombay Stock Exchange, is the most popular index in India. The performance of the top 50 firms listed on the National Stock Exchange is tracked by the NSE's own index, the Nifty 50. These indices are valued according to the market capitalization of the firms which make up the index. The index functions as a benchmark for the performance of the index-companies and reflects the market's general mood.

The objectives of this report are: -

- To determine the factor that can affect the Indexes in question namely Nifty 50, S&P 500 and Nikkei 225 using already existing literature published.
- To make all the data take as stationary so that time series analysis can be done.
- To create a suitable model using Vector Auto Regressive Model (VAR) in order to create a predictive model and the gain an equation for the following indexes
- Selecting suitable lag parameter so that optimal model can be created and performing various test in order to make sure all the assumptions of the time series analysis and regression are met.
- Performing various test in order to establish that the assumptions of these models are met like Heteroskedasticity, normality etc.
- Creating a ARMA model for the same parameter and evaluating which of the 2 models is more optimal.

The Secondary Data was collected from various Yahoo finance and government websites. On the basis of the information gained the model is created using various parameters found and supported by literature review.

The models are created and suitable lag are selected which are used to create the model equation. The models were tested for in order to confirms that none of the assumptions of the models are violated. The model created in VAR though had a great value of Square but violated a condition of Cointegration which led us to using VECM model in order to correct the VAR model. ARMA model on the other had good R square value which was lower than VAR but had no violation of any assumption.

The interpretation of this study is based on the assumption that the information taken is correct and the faults of the methods used are minimized but not completely removed.

Contents

List of Figures

List of Table

Chapter-1 Introduction

1.1 Financial Market

The financial market is a platform on which different financial instruments are traded between buyers and sellers, including bonds, stocks, commodities, currencies, and derivatives. Financial markets provide an essential infrastructure for businesses, people, and governments to invest, manage risks, raise money, and protect themselves from unforeseen events.

Types of Players in Financial Market

- **Investors:** These are either institutional buyers or private people of financial securities. Retail investors, fund managers, and institutional investors are among them.
- **Issuers:** These are individuals or organizations who sell securities on the stock market to raise money. Governments, municipalities, and businesses are among examples.
- **Intermediaries:** Financial firms are referred to as intermediaries as they facilitate market trade. Brokers, investment banks, and dealers are some examples.
- **Regulators:** They are governmental organisations that keep an eye on the market to maintain stability, openness, and fairness. Examples include the Financial Conduct Authority (FCA) and the Securities and Exchange Commission (SEC) in the US in the UK. the Indian Securities and Exchange Board (SEBI)(*Key Players in the Capital Markets*, n.d.)

Types of Financial Markets

- **Primary Market:** This market is where securities that have recently been issued are sold. Organizations acquire money in the primary market to finance their expansion goals or operations.
- **Secondary Market:** This is the market where investors exchange pre-existing securities. In the secondary market investors have access to liquidity, which enables them to purchase and sell assets whenever they choose.
- **Money Market:** Commercial papers and treasury bills are examples in money market of short-term debt securities that are traded on. It offers a venue for individuals and companies to temporarily invest extra cash.
- **Capital Market:** Long-term securities, such as stocks and bonds, are exchanged on the capital market. The capital market to raise long-term cash for investments or growth offers a venue for people and businesses.

- Over the Counter (OTC) Market: In a decentralised market known as the over-the-counter (OTC) market, securities can be exchanged directly between two persons without the use of a broker. There are no actual sites, and all commerce is done online.
- **Forex Market:** The market where players may sell, purchase, speculate, and hedge on the rate of exchange between currency pairings is referred to as the forex (foreign exchange) market. Since cash is a highly liquid asset, the foreign exchange market is the most liquid market. Every day, transactions on the currency market go over $6.6 trillion, which is far greater than on the stock markets and futures put together.
- **Commodities Market:** Producers and consumers exchange tangible products such as energy goods (such as oil, carbon credits, and gas), agricultural goods (such as livestock, corn, and soybeans), precious metals (such as silver, gold, and platinum), or "soft" goods (such as cotton, sugar, and coffee) on commodities markets. These places, where tangible products are traded for cash, are referred to as commodities markets.

1.2 Indexes

An index is a statistical metric used in financial market to track the return on investment of a collection of assets or securities. It offers a quick overview of a sector's or market's performance as a whole, that could be used as an indicating baseline to assess how well investment portfolios are operating. An index can be built in a number of different methods, including market-cap weighted, price-weighted, and equal-weighted.

Types of Indexes

- **Price-weighted Index:** In the context of a price-weighted index, each security is assigned a weight depending on its price. Greater priced securities will be given a greater weight in the index. The Dow Jones Industrial Average (DJIA) is the most prevalent illustration of a price-weighted index. Thirty large-cap firms that are listed on the NYSE and NASDAQ are monitored by the DJIA for their performance.
- **Market-cap Weighted Index:** The weights of the constituent stocks in a market-cap weighted index are determined by their market capital. The index will be heavier on the equities with a bigger market cap. The S&P 500 is the most prevalent illustration of a market-cap weighted index.
- The performance of 500 large-cap businesses listed on the NYSE and NASDAQ is tracked by the S&P 500 index.
- Equal-weighted Index: All securities, regardless of their prices or market size, are assigned equal weights in an equal-weighted index. The Russell 2000 is the most popular illustration of an equal-weighted index. The performance of 2000 small-cap firms listed on the NYSE and NASDAQ is tracked by the Russell 2000.(*Market Index*, n.d.)

Other applications for indexes include:

- **Investment Portfolios:** Indexes may be a useful tool for building investment portfolios that are designed to closely mirror the performance of a specific market or industry.
- **Risk management:** By showing the overall performance of a certain market or sector, indices may be used to manage risk.
- **Research:** Indexes may be used as a research tool to study different marketplaces and industries.

Examples of Indexes we will predict

- **Sensex S&P BSE:** The Bombay Stock Exchange (BSE) in India uses the S&P BSE Sensex as its benchmark index. The performance of Thirty large-cap businesses trading on the BSE is tracked by this market capitalization-weighted index.
- **Nikkei 225:** The Nikkei 225 serves as the benchmark index for Tokyo Stock Exchange in Japan. This price-weighted index measures the returns of 225 large-cap companies traded on the Tokyo Stock Exchange.
- **Nifty 50:** In India the National Stock Exchange (NSE) uses the Nifty 50 as its benchmark index. This index, which is based on the value of market caps, monitors the performance of 50 large-cap businesses listed on the NSE.

1.3 Econometrics/ Econometric Model

Econometrics is a subfield of economics that uses statistical techniques to examine economic data and evaluate economic ideas. To quantify and comprehend economic events, it blends statistical analysis, economic theory, and computer programming. Economic theories may be put to the test using actual data and econometric models offer a mechanism to quantify the correlations between various economic variables.

Need for Econometrics

Since economic data are frequently intricate and challenging to interpret without the use of quantitative methods, econometrics is required. Econometrics offers a method for comprehending the connections between various economic factors, forecasting future financial results, and assessing the success of investment programmes. Economists can evaluate the strength of these linkages and discover the causal links between various economic variables. For example, consider a case in which a government wishes to assess how new regulations would affect the economy. Statistics regarding economic activities prior to and thereafter the policy was adopted could be analysed by the government using econometric methods, and the resulting data could be compared to data on comparable economies which weren't subject to the programme. The government can use econometric models to determine the causal impact of the regulation on the economy and assess the effectiveness of the programme.

What is an Econometric Model

A mathematical depiction of the connections between economic factors is called an econometric model. It describes the relationship between a dependent variable and a number of independent variables. An econometric model, for instance, would look at the connection among a business's revenue and its marketing spending. The model would outline the relationship between variations in advertising spending and sales. To forecast future economic trends and assess the success of economic policy, models based on econometrics can be utilised. A tax cut's effect on economic growth, for instance, may be assessed using an econometric model, as could the effect of an increase or decrease in rate of interest on the economy.

Different Types of Econometric Models

There are multiple kinds of econometric models, all of which have unique advantages and disadvantages. The following constitute a few of the most typical econometric model types:

- **Time Series Models:** It examine data that has been collected across time. They are employed to identify patterns and trends within financial data and forecast upcoming monetary developments. To analyse data on economic variables like inflation, GDP, and unemployment, time series models can be utilised.
- **Cross-Sectional Models:** It examine data at one particular time point. They are used to examine the connections between various economic factors, such educational attainment and income. Cross-sectional models may be used to pinpoint the causes of financial disparity and assess how well measures aimed at eradicating it are working.
- **Panel Data Models:** It examine time series data and cross-sectional. They are employed to examine how various economic factors relate to one another over time as well as across various groupings. Panel data models may be used to analyse how policies affect various population groups and to pinpoint the elements that fuel local economic expansion.

Limitation of Econometric and Econometric Models

Despite being helpful, econometrics has significant drawbacks. Its reliance on the presumption that the economic ties under study would remain constant throughout time is one of its limitations. The econometric model might not hold up if its root economic linkages alter. For instance, an econometric model which was correct before to the arrival of the innovation might cease to be accurate if the advent of a new technology alters how a certain sector runs.

Econometrics' inability to study correlations between factors that are not visible and quantifiable is another drawback. Therefore, significant intangible aspects that influence economic results could not be adequately accounted for by econometric models. For instance, an econometric model which examines the link between income and education.

1.4 Regression

Regression analysis is used to statistically examine the relationship between the dependent variable and a number of independent factors. Using the values of a different variable as a guide, one can utilise this strategy to forecast the values of another variable. Numerous disciplines, including economics, finance, marketing, psychology, and a number of others fields, employ regression analysis extensively.

Linear regression, commonly referred to as ordinary least-squares regression (OLS) or normal regression, is the most widely used variant of this approach. An analysis of linear regression establishes the linear relationship between two variables on the foundation of a line of greatest fit. As a result, the slope of the straight line utilised for linear regression shows how altering one variable impact altering another. The y-intercept in a linear regression connection represents the result of a single variables then the outcome of the other variable is zero. Alternative non-linear regression models exist, but they are far more difficult.

Regression and Econometrics

Data analysis in economics and finance is done using a set of mathematical methods called econometrics. The use of observable data in an analysis of the income impact is an example of how econometrics is applied. For instance, an economist would postulate that as one's income rises, so will their expenditure. If the data support the existence of such an association, a regression analysis can be performed to determine the magnitude of the association among consumption and income and if it is statistically significant i.e., whether it is unlikely that the association is the result of pure chance.

Linear Regression

To find the line of best fit in linear regression models, the least-squares approach is widely utilised. The least-squares approach results from reducing the sum of squares that a mathematical problem generates. A square is made by the mean of the observed data set or squaring the distance between a finding and the slope of regression.

Following this process, a regression model is created, which is frequently done nowadays using software. The general form of each regression model is as follows:

Straight-line regression:

$$y = a + bx + u$$

Where x is the independent variable

y is the dependent variable

u is error term or regression residue

b is slope/gradient of the line

a is the y intercept

Logistic Regression

Logistic regression is used when the variable that is dependent is categorised or binary. It is used to determine the chance that an event will occur based on the outcomes of the independent variables. The logistic regression equation reads as follows

$$ln[P\ /\ 1 - P] = \beta_0 + \beta_1 x_1 + \cdots\cdots \beta_n x_n$$

where p is the likelihood that an event will occur, x1, x2,..., xn denotes the number of independent variables, and B0, B1, B2, …, Bn are the coefficients.

Polynomial Regression

When there is a curved or nonlinear connection between the variable that is dependent and the independent variable, polynomial regression is performed. A polynomial expression is applied to estimate the connection between the variables.

$$f(x) = C0 + C1X + C2\ X^2 \cdots CnX^n$$

where c is a collection of coefficients and n is the polynomial's degree.

Multiple Regression

When a number of independent variables has an impact on the dependent variable, multiple regression is performed. It is used to determine the influence of each independent variable on the dependent variable while taking into consideration the effects of all other independent variables.

$$y = B1X1 + B2X2 + \ldots + BnXn + C.$$

where c is the error or constant term and b is the regression coefficient

Assumptions of Regression

The regression model is based on a number of presumptions, all of which must be true in order to produce accurate results. Any of these presumptions that are broken might result in skewed figures and inaccurate conclusions. The seven regression assumptions are as follows:

Linearity: the variable that is dependent and the variables that are independent have a straight-line connection. In other words, regardless of the level at which the independent variable is present, its impact on the variable that is dependent remains constant.

Independence: There is no interdependence between the observations. In other words, the dependent variable's value for a single observation is independent of its value from every other observation.

Homoscedasticity: Throughout every level of the independent variable(s), the variance of residuals (the discrepancy among the actual and projected values of the variable that is the dependent variable) remains constant. In other words, regardless of the value of the independent variable(s), the dispersion of the residuals around the line of regression is the same.

Normality: The residuals have a normal distribution. This presumption is crucial because it enables us to draw inferences about the population using inferential statistics (such as hypothesis testing).

No Multicollinearity: No multicollinearity exists because the independent variables have low correlations with one another. Multicollinearity can result in exaggerated standard errors and unstable estimations of the regression coefficients.

No Autocorrelation: There is no correlation between the residuals. Whenever the residuals of nearby observations are associated, autocorrelation arises. Incorrect standard errors and skewed estimations of the coefficients of regression may result from this.

Additivity: Each independent variable has an additional impact on the dependent variable. In other words, the values of the other independent variables have no bearing on the impact of a variation in a single independent variable on the variable that is dependent.

1.5 Time Series Analysis

A statistical method called the analysis of time series is used to examine data that evolves over time. It is employed to spot cycles, trends, and patterns in the data. In domains where data is gathered over time, such as engineering, economics, finance and others, time series analysis is frequently employed. Modelling the time series' underlying structure and generating forecasts based on the model constitute the analysis.

Time series may be divided into three categories: difference, trend and stationery.

Stationary Series

The mean and variance of a time series that is stationary remain constant across time. In simple terms, the time series' statistical characteristics don't change over time. This indicates that there are no patterns or trends in the data, and the spread of the data stays constant throughout time.

Trend Stationary Series

The mean of a pattern in stationary time series is constant, while the variance varies with time. This indicates that while the time series' statistical characteristics vary with time, the mean does not. This indicates that the data lacks cycles but has a trend.

Difference Stationary Series

A distinction while the mean of stationary time series changes with time, the variance remains constant. This indicates that while the time series' statistical characteristics vary with time, the variance does not. The data therefore exhibits cycles but no pattern.

The following are the fundamental methods for time series analysis:

Time Series Plotting

Plotting the data through time is the initial stage in time series analysis. This makes it easier to spot cycles, trends, and patterns in the data. Finding aberrations and values that are absent in the data is also helpful.

Trend evaluation

The long-term pattern in the data is found via trend analysis. The data may be used to fit a line or curve, and the gradient of the line can be used to determine this. An upward trend is shown by a positive slope, and a downward trend is indicated by a negative slope.

Seasonality Research

Seasonality analysis is utilised to find the seasonal trends in the data, Plotting the data across time and searching for patterns that recur on a regular interval basis might help with this.

Analysis of autocorrelation

The relationship between the data at various moments in time is found via autocorrelation analysis. Plotting the data's partial autocorrelation function (PACF) and autocorrelation function (ACF) might be used to do this.

Time series breakdown

The time series is divided by its residual, trend, and seasonal elements using time series decomposition. Methods like exponential smoothing or moving averages can be used for this.

Forecasting

To anticipate upcoming trends in the time series, forecasting is utilised. ARIMA (autoregressive integrated moving average), seasonal ARIMA or Exponential smoothing, ar e some techniques that may be used for this.

Time series Forecasting

To anticipate future behaviour, time series forecasting works by spotting patterns and trends in historical data. The main presumption is that a time series' past behaviour and future behaviour are connected. Finding the best model that accurately reflects the important patterns and trends and using it to provide precise forecasts is the difficult part.

Periodicity and seasonality in the study of time series:

Periodicity describes a pattern's propensity to repeat itself at random intervals. For instance, there may be long-lasting boom and bust cycles on the stock market that extend for a couple of decades. Different methods, such as ARIMA (autoregressive integrated moving average) models, which include seasonal and periodic data, can be used to capture these periodic patterns.

Contrarily, seasonality is the propensity for specific patterns to repeat themselves periodically. For instance, in many places, the need for cooling systems tends to increase in the summer and decrease in the winter. Several methods, including Fourier analysis, which breaks down a data set into its frequency components, can be used to record this seasonal pattern.

Various time series forecasting models include:

(AR) Auto Regressive Model

Models called autoregressive (AR) presume that a time series' future values rely on its previous values. The quantity of historical data utilised to forecast future values determines the model's ranking.

(MA) Moving Average Models:

These models rely on the premise that a time series' future values are a weighted mean of its previous numbers, with their weights being smaller as the time lag gets longer.

(ARIMA) Autoregressive Integrated Moving Average Model:

The AR and MA models are combined in the ARIMA (Autoregressive Integrated Moving Average) model, which additionally includes a differencing step to keep the time series stationary.

(SARIMA) Seasonal ARIMA Model:

SARIMA stands for seasonal ARIMA models, which expands the ARIMA framework to include seasonal elements like quarterly or monthly trends.

Exponential Smoothing Model:

Models of the exponential smoothing family are based on the idea that a time series' future values are a weighted mean of its previous numbers, with the weights eroding exponentially with time.

Neural Network Model:

Models based on neural networks: This model learns the intricate patterns in the time series using a recurrent neural network (RNN) or MLP (multilayer perceptron).

Vector Autoregressive Model:

When numerous time series are connected and capable of predicting one another, vector autoregression models (VAR) are applied.

(BSTS) Bayesian Structural Time Series Models:

These models use Bayesian statistics to calculate the time series model's parameters and generate forecasts.

Prophet model:

This non-linear model, created by Facebook, incorporates trend and seasonality components to generate predictions.

Time Series Regression:

A statistical technique called time series regression is used to examine the connection over time among a variable that is dependent and a number of independent variables. In other words, it entails forecasting the future outcomes of the variable that is dependent based on historical data. The independent variables are the variables that may affect the variable that is dependent, which might include economic indicators or interest rates. The variable that is dependent is often a continuous variable that varies over time, such as stock prices.

The time series regression equation is:

$$y = B0 + B1x1 + B2x2 + \dots + Bkxk + \varepsilon$$

If the dependent variable is y, the independent variables are x1, x2,..., xk, ε is the error term, the regression coefficients are B0, B1, B2, ..., Bk.

The following are the time series regression's underlying presumptions:

Linearity: The variable that is dependent and the variables that are not dependent have a straight-line connection.

Stationarity: The dependent variable's mean and variance remain stable across time.

No autocorrelation: The variation terms at different time points do not exhibit any correlation.

Normality: The distribution of the error terms is normal.

The ARMA Model

The statistical model known as ARMA, or autoregressive moving average, is used to examine time series data. It combines two models: the autoregressive (AR) model and the moving average (MA) model.

The dependent variable's value is predicted by the AR model based on its past values, whereas the dependent variable's value is predicted by the MA model based on its past mistakes. To accurately represent moving average and the autocorrelation features of time series data, the ARMA framework combines these two models.

The ARMA equation is:

$$Y_t = \mu + \sum_{k=1}^{p} \phi_k (Y_{t-k} - \mu) + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \qquad t \in \mathbb{Z}, \tag{1}$$

where μ is the process mean, $\{\epsilon_t\}$ is a white noise process with mean 0 and variance σ^2, $\phi_p \neq 0$ and $\theta_q \neq 0$. □

Alternatively, the model (1) may be written as

$$Y_t = c + \sum_{k=1}^{p} \phi_k Y_{t-k} + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \qquad t \in \mathbb{Z},$$

where the constant c is given by

$$c = \left(1 - \sum_{k=1}^{p} \phi_k\right) \mu.$$

Another way of writing the model (1) is as

$$\phi(L)(Y_t - \mu) = \theta(L)\epsilon_t, \tag{2}$$

where $\phi(z)$ is the AR characteristic polynomial,

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p,$$

and $\theta(z)$ is the MA characteristic polynomial,

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \ldots + \theta_q z^q.$$

where t is the error term at time t, yt is the dependent variable at time t, c is the constant term, 1, 2,..., p are the autoregressive coefficients, p is the order of the autoregressive process, t is the error term at time t, 1, 2,..., q are the moving average coefficients, and q is the order of the moving average process.

The following are ARMA's presumptions:

Stationarity: The dependent variable's mean and variance remain stable across time.

No autocorrelation: The error terms at different time points do not exhibit any correlation.

Normality: The distribution of the error terms is normal.

Invertibility: A finite autoregressive process can be created by inverting the moving average process.

ARIMA Model

The famous time series analysis method known as ARIMA (AutoRegressive Integrated Moving Average) is used to model and forecast data that displays trends, seasonality, and other intricate patterns. It consists of the Auto Regressive (AR), Integrated (I), and Moving Average (MA) time series models.

The MA component represents the reliance among the present data and a residual error from a model of moving averages applied to lagged data, whereas the AR component depicts the dependence between the current data and a lagged observation. By comparing the time series, the I component corrects the data's non-stationarity. ARIMA can capture a variety of relationships and time series patterns through the integration of these three models.

The basic equation for ARIMA(p,d,q) is: $[1 - \phi_1 * L - \phi_2 * L^2 \ldots - \phi_P * L^P][1 - L]^d y_t = C + [1 + \sigma_1 * L + \sigma_2 * L^2 \ldots + \sigma_q * L^q + error$

where:

L is the lag operator

Yt is the time series to be modeled

p is the order of the AR model

q is the order of the MA model

theta1 to thetaq are the moving average coefficients c is a constant term

d is the degree of differencing (the number of times the series is differenced to make it stationary)

phi1 to phip are the autoregressive coefficients

et is the white noise error term

The ARIMA model bases its predictions on the idea that the time series is stable, or that its statistical characteristics, such mean and variance, remain constant across time. The model contains differencing to make the time series stationary if it is not. The ARIMA model also requires a normal distribution with constant variance for the residuals.

Several statistical tests may be used to verify the ARIMA model's assumptions. For instance, the test known as the Augmented Dickey-Fuller and the Ljung-Box test may be used to check for stationarity and autocorrelation, respectively, in the residuals.

1.6 Objectives of the Study

- To determine the factor that can affect the Indexes in question namely Nifty 50, S&P 500 and Nikkei 225 using already existing literature published.
- To make all the data take as stationary so that time series analysis can be done.
- To create a suitable model using Vector Auto Regressive Model (VAR) in order to create a predictive model and the gain an equation for the following indexes
- Selecting suitable lag parameter so that optimal model can be created and performing various test in order to make sure all the assumptions of the time series analysis and regression are met.
- Performing various test in order to establish that the assumptions of these models are met like Heteroskedasticity, normality etc.
- Creating a ARMA model for the same parameter and evaluating which of the 2 models is more optimal.

1.7 Scope of Study

According to the existing study, the outcomes of the interactions between numerous significant factors are diverse and ambiguous. The time span of the investigation and the time series modelling method employed by the research projects could be the causes of these outcomes. Therefore, it is essential to routinely check the connection using sophisticated methodologies. The study uses vector autocorrelation and cointegration techniques with more current data to investigate the degree of correlations between the stock market returns, cost of crude oil, and gold prices. The study's goal is to systematically validate the association. In this study we would be also creating a ARMA model as well.

Chapter-2 Literature Background

2.1 Abstract

Researchers, decision-makers, and entrepreneurs have all been drawn to the dynamic and intricate interaction between economic variables. In this study, the dynamic relationship between the price of gold, stock returns, currency rates, and oil is tested. Since all of these factors have undergone considerable changes over time, it is imperative to regularly validate the link. Vector autoregressive and cointegration approaches were used in the study to try and represent the dynamic and steady relationship between these variables.

Keywords

ARIMA, Time Series Analysis, Regression, Unit root, Granger Causality Test, Augmented Dickey-Fuller Test.

2.2 Introduction

One of the earliest metals that people dug out was gold. Gold has a hybrid character that makes it analogous to money. It is a resource utilised in numerous sectors as well as serving as a method of trade and an instrument of value throughout history. The Bretton Woods system fixed the value of one troy ounce to the US$35 for the following World War II. The US officially abolished 1971, the automatic conversion of the US dollar to gold and switched to what is now a fiat currency system, the system was in place. The Swiss Franc, which separated from gold in 2000, was the most recent currency.

A relatively minor variation in the actual cost of gold throughout a period of 172 years was shown between the prices of gold between 1833 and 2005, which were $445 and $20.65 per ounce, respectively. Gold's price fell to as low as $257 in September 2001, following a 20-year decline. Early in the 1980s, there were certain days when the value of gold exceeded $800, and there was a near 20-year price impasse. The first occasion since 1982, gold breached the $500 barrier in December 2005.

One ounce of gold was once only worth 7.7 barrels of crude oil in 2005. Since the two commodities' prices have been correlated throughout the initial 40 years, which is the smallest increase. Over the next 40 years, for every ounce of gold there have been 15.2 barrels of crude oil, on average. An ounce of gold could only purchase little more than eight barrels of crude oil between 1975 and 1980, when the Organisation of Petroleum Exporting Nations dramatically hiked the price of crude oil for the first period.

After the dollar/gold convertibility broke down in 1973, the dollar continued to fall. As a result, oil prices surged four-fold to about $12 per barrel in 1974, leading to a dramatic rise in U.S. petrol costs and a subsequent reduction in consumer demand. The same time frame saw a 15% increase in gold prices. The major causes of the 1980–1982 recession were a falling currency and record oil prices. Due to a doubling in oil prices to $29 per barrel and a more moderate 30% gain in gold, the ratio of gold to oil decreased in January 1979 Due to a rise in oil to $29 a barrel and a less dramatic increase in gold of 30%, the gold/oil ratio decreased in January 1979 from 15.3 to 11.4 in August 1979.

The oil/gold ratio briefly increased in fall 1979 from 12.5 to 21 in winter 1980. This is because gold prices increased by $400 between September 1979 and January 1980 as a consequence of the Soviet Union's takeover of Afghanistan.

After falling from a high of 16.9 in February 1983 despite rather stable prices for the fuel and the metal, the oil/gold ratio dropped out at 10.6 in autumn 1985. This occurred at the same time as the fed funds rate peaked at 8%. Oil prices rose about a mere $21 per barrel to $31 per barrel around two weeks on August 2, 1990 following Iraq's tragic invasion of Kuwait, before rising in October to a high $40 per barrel.

Due to the decision taken by OPEC to boost supply and a decrease in Asian demand for oil during the 1997–1998 economic crisis, oil prices fell precipitously in December 1998. Due to an error by OPFC, oil prices fell to their lowest prices since the end of the 1986 oil glut in December 1998, when they were reduced by 50% to $11 per barrel. A second time, the decline in the oil/gold ratio to a nine-year bottom of 1 M in 1999 served as a premonition of an economic slump.

Oil prices started their multi-year bull market at the start of the second Iraq War in March 2003, climbed in March 2003 from $30 per barrel to in March 2005 greater than $50 per barrel. Compared to the previous two years, oil has increased by more than 100%, while gold has increased by 54%. Oil concluded at $61 that year a barrel. The oil/gold ratio fell in August 2005 to 6.7%, the lowest level in its 35-year history, as a result of changes in oil prices.

The price of oil started their multi-year bull market at the start of the subsequent Iraq War in March 2003, climbed in March 2003 from $30 per barrel to in March 2005 greater than $50 per barrel . Compared to the previous two years, oil has increased by more than 100%, while gold has increased by 54%. Oil concluded at $61 a barrel. The oil/gold ratio fell in August 2005 to 6.7%, the least level in its 35-year history, as a result of changes in oil prices.

2.3 GOLD AND HIGH OIL PRICES: OVERVIEW

Politics may have played a role in the 1980 gold price increase. The Soviet Union's invasion of Afghanistan, which started around Christmas 1979, shocked the world at that point in time. 1978 saw the signing of a "bilateral treaty of cooperation" among the Afghanistan and Soviet Union, but by the following year, things had changed. This move, which took place in the middle of the Cold War, it inflicted a serious blow to the United States, which had been suffering from inflation, high unemployment, and oil prices.

It didn't feel at all certain how the power and the economy of America would develop in the future. Gold was just a safe refuge in moments of anxiety and conflict, reflecting that dread. The start of a 22-year gold downtrend followed this all-time top, but the purchasing frenzy swiftly passed.

Gold's price increased dramatically from $255 in 2000 to nearly $1400 per ounce in 2010. Inflation numbers decreased sharply in 2009 and 2010, occasionally even approaching bearish levels. Compared to its 2007 highs, the equity market has drastically declined. As of late 2010, the indicators remain ambiguous about trend. The world economy is still in an unstable state as it emerges from recession.

In July 2008, the price of oil reached a record-high level of $145 per barrel. As a result, petrol reached four dollars a gallon. A lot of news outlets attributed this to a combination of increasing demand from India and China and falling supplies from oil resources in Iraq and Nigeria. In actuality, throughout that period, worldwide supply increased while global demand decreased. Beginning with 86.66 million barrels of crude per day (bpd) in the fourth quarter of 2007 to 85.73 million bpd during the initial quarter of 2008, the use of oil dropped. Supply climbed about 85.49 to 86.17 million bpd at the same period.

2.4 A REAL TERMS COMPARISON OF GOLD AND OIL FROM 1900 TO 2010

In actual terms, gold reached in 1980 a record-breaking high of $1537.94. The highest actual oil price ever was $96.91 in the year 2008. The year 2010 saw the next-highest gold price recorded at $1208.55. The second-greatest oil price in the previous 110 years occurred in 1980, when it reached $95.89. Over time, there were changes in the true price of oil. With an average price of $12.17, the years 1960–1969 had the lowest oil prices between 1900 and 2010. With a mean value of $240.18, the actual gold price was likewise at the lowest point in history during the years 1960 to 1969.The decade 1920–1929 had the second lowest actual gold prices, with a mean price per ounce of $254.84.

In the forty-year span between 1930 and 1969, the value of oil and gold fluctuated erratically. Mean oil and gold prices were displayed declining trend from the 1940s and 1969. The mean gold price in dollars declined by 46.8% throughout this thirty-year period. When contrasted to the ten years before that, the mean value of oil climbed by 4.3 percent between 1940 and 1949. Mean oil prices fell by 2.79 percent and 18.8 percent, accordingly, throughout the 1950s and 1960s. In comparison to the prior decade, the mean actual gold price climbed by 2.04 and 1.76 times, respectively, all throughout the 1970s and 1980s. The mean actual price of gold fell by 41.79 percent throughout the 1990s.

The true price of oil fell by 52% within the same time frame. The average cost for crude oil climbed about 23.5% between the years 2000 and 2011, while the value of gold shot up by almost 107 percent.

The value of gold has varied greatly throughout the past century and decade. The biggest swings in the price of gold occurred between 2000 and 2011. Relative to the earlier analytical periods, the swings in the value of gold doubled after 1970. Actual price of gold showed the least fluctuation between and 1960 and 1969 between 1920 and 1929.In the decade that followed, gold price swings were 872-fold greater than they were in the early 1960s. In the years 1950–1959, prices for oil remained very consistent. Oil price variations became more pronounced in the 1970s and 1980s.

2.5 A REAL TERMS COMPARISON OF GOLD AND OIL FROM 2011 TO 2023

Among of the most significant commodities throughout the global economy are gold and oil, together with oil serving as the predominant form of energy and gold serving as a refuge asset for investors. These commodities' actual prices underwent considerable variations between 2011 and 2023. In 2011, the cost of gold rose to a record unparalleled record high of 1,900 dollars per ounce. The following was brought on by a variety of things, such as the weakened US currency, political instability, and inflationary worries. Nevertheless, there have been falls in the value of gold subsequently, with some variations. The value of gold in 2023 is projected to be approximately 1,700 dollars per ounce, this continues to be quite high.

In terms of oil, the cost was nearly one hundred dollars per barrel in 2011. The cost for crude oil nevertheless began to significantly decline in the period that followed as a result of an abundance and insufficient demand. Oil prices decreased to a trough of about thirty dollars per barrel in 2016. After that point, the cost of oil has been gradually increasing, reaching a high of almost eighty dollars per barrel in 2023. It is essential to take inflation into account in actual terms. The pace with which costs are generally increasing and, as a result, the buying power of money is decreasing is known as inflation. In 2011 The cost of gold was almost $2,175 in 2023 USD once inflation was taken into account. This indicates that during the previous twelve years, the actual cost of gold has dropped by almost 22%. In contrast, in 2011 the cost of oil was almost $110 in 2023, which indicates that during the same time span, the cost of oil fell by about 27% in real-world terms.

The COVID-19 epidemic constitutes a major single element that has recently impacted the cost of both oil and gold. Oil consumption has decreased as a result of the epidemic, therefore has increased pricing pressure. The epidemic has also increased uncertainty regarding the economy, that has pushed up the value of gold because buyers look for safe haven assets. The move favouring renewable energy sources has additionally had an impact on oil prices. The need for oil is anticipated to decline as more nations make investments in energy from renewable sources, that might further squeeze prices.

Finally, it can be said that between 2011 and 2023, the actual prices of both gold and oil saw considerable fluctuations. In actual terms, the cost of gold has dropped by approximately 22%, whereas the cost of oil has fallen by about 27%. The COVID-19 epidemic and the move closer to energy from renewable sources are two variables that have recently and probably will keep continuing to have an impact on the cost of both products.

Chapter-3 Literature Review

Understanding the connections or interactions between various indices of economic activity has been the subject of extensive research. Studies have looked into how the prices of oil and gold relate to stock values. Rates of interest (Hondroyiannis & Papapetrou, 2001), exchange rates (Amoateng & Kargar, 2004), manufacturing output (Marion and Flood, 2006), and price inflation (Moore, 1990) are only a few examples of economic indicators. Using data specific to the United Kingdom, El-Sharif et al., (2005) discovered favourable, frequently significant connections between the price of oil and stock values in the gas and oil industry. Strong support for the claim that oil price risk affects stock price returns in developing nations was provided by Basher & Sadorsky (2006). Numerous research has made an effort to statistically represent the factors that affect the value of gold. These investigations often employ three basic methodologies.

Table 3-1 Literature Review Approach

Approaches	*Perspectives*	*Studies*
1	Models variation in the price of gold in terms of variation in main macroeconomic variables	Ariovich, 1983; Dooley, Isard and Taylor, 1995; Kaufmannand Winters, 1989; Sherman, 1982, 1983, 1986; Sjaastad and Scacciallani, 1996).
2	Focuses on speculation and the rationality of gold price movements	(Baker and Van Tassel, 1985; Chua, Sick and Woodward, 1990; Diba and Grossman, 1984; Koutsoyiannis, 1983; Pindyck, 1993)
3	Gold as a hedge against inflation with particular emphasis on short-run and long-run relationships	Chappell and Dowd, 1997; Ghosh *et al.*, 2004; Kolluri, 1981; Laurent, 1994; Mahdavi andZhou, 1997; Moore, 1990; Ranson, 2005a, b).

Source: Self Analysis

Applying everyday dollar-based indexes of stocks dataset, Janabiet al. (2010) investigate how well the Gulf Cooperation Council (GCC) equities markets are efficient in terms of information with respect to the gold and oil price spikes over the years 2006–2008. The influence of the gold and oil prices on the financial performance across the six unique GCC equity markets is investigated as well in the study. According to the research, the equities markets in the GCC are efficient in terms of information when it comes to the oil and gold price indices.

The research conducted by Zang et al. (Zhang & Wei, 2010) examines the causation and cointegration connection that exists between the prices crude oil of and gold. The analysis discovers a strong positive association between the price of crude oil and gold over the sample time frame, with recurrent trends between the two prices. The report goes on to say that the irregularity of the value of gold is caused by a long-term balance among both markets and a linear Granger shift in crude oil price. The impact of the price of crude oil appears to be greater compared to the impact of the gold price in terms of the overall effective cost among the two markets.

According to Laughlin's analysis from 1997 (Laughlin, 1887), gold's value has increased regardless of whether other assets have increased or decreased in value relative to it. The analysis by Ashraf looks at five examples when a lower level of oil-gold ratio corresponded with decreasing (or negative) yield propagates, a topping a dropping dollar, fed funds rate, and ultimately declining growth.

Pravit (2009) forecasts the price of gold using a combination of auto regressive integrated moving average (ARIMA) and multiple regression. According to the study's findings, ARIMA (1, 1, 1) was the model that is most useful for predicting the short-term price of gold. The study revealed indicates the change in the price of Thai gold is influenced by the Japanese Yen, Australian Dollar, Canadian Dollar, US Dollar, Oil Prices, EU Pond, and Gold Future Prices using a multiple regression model.

The Larry et al. (1997) analysis lends credence to the idea that the global gold market was efficient between 1991 and 2004. The analysis also reveals that the cost of gold in every other currency is significantly impacted by the actual increases or decreases in both the Japanese yen and the euros versus the dollar. The research also reveals even world's three largest gold producers Russia, Australia, and South Africa seem to have virtually no effect on the cost of gold globally. The key findings of the research conducted by Ismail et al. (2009) show that a number of factors, including the USD/EUR exchange rate, the money supply (M1), the inflation rate, the S&P Poor Index, the NYSE Index, and the US dollar index, affect gold prices.

In his 2004 study, Max proposes an economic theory for the nominal values of gold and oil. A VAR system featuring previously unknown structural fractures is used for verifying the mathematical model. Studies using US data show that monetary considerations are the primary driver of actual gold and oil prices. Additionally, money Granger produces inflation, and this in turn Granger causes fluctuations in the trajectory of production growth.

The research conducted by Mu Lan et al. (2010) examines the effects of changes in the prices of crude oil, gold, and different currencies' conversion rates on the value of the stock metrics of the Germany, Japan, US, China, and Taiwan, as well as the long-term and immediate correlations between these variables, using every day data and using the time series method. The findings demonstrate that there are cointegrations between changes in the prices of gold, oil, and other currencies as well as the stock exchanges in Germany, China, Taiwan, and. Japan

Moore (1990) employed the most prevalent inflation signals to evaluate the link between those enabling signals as well as the price of gold of the New York Market from 1970 in order to investigate if the value of gold were impacted by other market conditions along with inflation. According to empirical findings, from 1970 to 1988, there was an adverse relationship between the stock and bond markets and gold prices, meaning that whenever gold rates rose, the markets declined.

Making use of monthly, weekly, and quarterly information, the article by Ai Han et al. (2008) presents a periodic approach to investigate the connection that exists between the rate of exchange of the US dollar versus the Australian dollar and the price of gold. The time-series approach uses period sample information to demonstrate changeable instability. Multi-model evaluation is added to the existing ILS technique, and computational frameworks are offered. According to the empirical evidence, both the short- and long-term connections between the rate of exchange and the value of gold are accurately represented by the ILS estimations.

According to Eric et al. (2006), who used cointegration methods, there is indeed an ongoing correlation among the level of the US dollar and the price of gold. The idea of a one percentage point rise in the overall US prices causes a one percent increase in the cost of gold is supported by the fact that the US value base and overall value of gold fluctuate in tandem through a manner that is statistically significant in the long-term connection fluctuations US inflation volatility, in US inflation, and credit risk all showed favourable correlations with gold price fluctuations. The study also discovered the opposite relationship among changes in the trade-weighted exchange rate and changes in the price of gold for the gold leasing rate as well as the US dollar.

Chapter-4 Research Methodology

4.1 Data Collection and Variables Selected

This study has taken daily gold prices in dollar from Nasdaq CMX. The crude oil daily prices were taken from Yahoo Finance Similarly the data for Nifty 50 Index Nikkei was taken from Yahoo Finance but the data of Investing.com. In the data series some of the data was missing due to holidays and events these data have been removed from the series.

The data then collected is from 10 April 2018 to 6 April 2023 constitution 1121 observations after filtering. The variation in the index is calculated using Log normal returns in excel using Ln function of 2 successive days given by $Ln(P_t / P_{t-1})$. It is done to make the data series Stationary. The stationarity of the data was confirmed using Jarque-Bera test is normal.

Software Used were EViews, Excel

4.2 Stationarity of the Data

A stationary time series is important for an analysis of regression that utilises the time series since it is challenging to find relevant data or features within a nonstationary time series. Consequently, a nonstationary time series could end up in an spurious regression. The majority of market time series, nonetheless, are nonstationary in real life. Time series must thus be rendered stationary following differencing. After undergoing differencing, there is still beneficial data or traits to be found in the respective series. If the variance, mean, and covariances of a series of time value are constant and the covariances rely on the separation between two time periods, the time series is considered to have become stationary. The sequence for integration & the stationary nature of the parameters are both tested using the unit root test. Stationary behaviour is frequently tested using the Dicky-Fuller unit root test (DF), the Augmented Dicky-Ful ler unit root test (ADF) (Dicky and Fuller, 1979), and the Phillips-Perron unit root analysis (PP) (Perron and Phillips, 1988).

Every single one of the variables that make up the model must be stationary in order for the VAR estimate to work. As used to describe the impulse response function, the initial variance data series has at least two benefits. First off, rather than emphasising the true change, it concentrates upon the growth or decline tendency. The impulse response component will identify an increase or decrease of the pattern since the initial difference series of data is the rise or drop between two successive dates. Secondly, it gathers additional data on the fluctuations to gold prices since the current data only indicates changes over a single day, but the first difference data reveals changes over the previous two days.

The ADF test was conducted in EViews by selecting the Sheet with the respective log returns then click on the view button and select unit root test and choose ADF test with test for unit root in level option and lag length of with maximum lag of 21 and to include the intercept in the test and check for the probability is the value of probability is less than 0.05 then the null hypothesis which in this case is that the series is stationary is accepted.

Figure 4.1 Stationarity of Nifty_DLOG

Augmented Dickey-Fuller Unit Root Test on NIFTY_DLOG

Null Hypothesis: NIFTY_DLOG has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=21)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-33.48737	0.0000
Test critical values:	1% level	-3.435973	
	5% level	-2.863911	
	10% level	-2.568083	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(NIFTY_DLOG)
Method: Least Squares
Date: 04/16/23 Time: 21:27
Sample (adjusted): 4/11/2018 4/06/2023
Included observations: 1120 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
NIFTY_DLOG(-1)	-1.001522	0.029907	-33.48737	0.0000
C	0.000470	0.000376	1.249026	0.2119

R-squared	0.500760	Mean dependent var	1.71E-07
Adjusted R-squared	0.500314	S.D. dependent var	0.017810
S.E. of regression	0.012590	Akaike info criterion	-5.910061
Sum squared resid	0.177209	Schwarz criterion	-5.901095
Log likelihood	3311.634	Hannan-Quinn criter.	-5.906672
F-statistic	1121.404	Durbin-Watson stat	1.999821
Prob(F-statistic)	0.000000		

Source Self Analysis

As it can be seen in Figure 1 that the stationarity of the Nifty50_Dlog is achieved as the probability is zero. Similarly, the stationarity result of Nikkei225_Dlog, S&P500_Dlog Oil_Dlog and gold_Dlog are shown in the following figures 2-5

Figure 4.2 Stationarity of Nikkei225_Dlog

Augmented Dickey-Fuller Unit Root Test on NEKKEI_DLOG

Null Hypothesis: NEKKEI_DLOG has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, maxlag=21)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-21.26963	0.0000
Test critical values:	1% level	-3.435978	
	5% level	-2.863913	
	10% level	-2.568084	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(NEKKEI_DLOG)
Method: Least Squares
Date: 04/16/23 Time: 21:30
Sample (adjusted): 4/12/2018 4/06/2023
Included observations: 1119 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
NEKKEI_DLOG(-1)	-0.901799	0.042398	-21.26963	0.0000
D(NEKKEI_DLOG(-1))	-0.109672	0.029778	-3.683024	0.0002
C	0.000188	0.000383	0.491244	0.6234

R-squared	0.512259	Mean dependent var	-6.60E-06
Adjusted R-squared	0.511385	S.D. dependent var	0.018317
S.E. of regression	0.012804	Akaike info criterion	-5.875451
Sum squared resid	0.182958	Schwarz criterion	-5.861992
Log likelihood	3290.315	Hannan-Quinn criter.	-5.870363
F-statistic	586.0506	Durbin-Watson stat	1.995693
Prob(F-statistic)	0.000000		

Source Self Analysis

Figure 4.3 Stationarity of SP500_Dlog

Augmented Dickey-Fuller Unit Root Test on SP500_DLOG

Null Hypothesis: SP500_DLOG has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=21)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-31.37905	0.0000
Test critical values:	1% level	-3.435973	
	5% level	-2.863911	
	10% level	-2.568083	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(SP500_DLOG)
Method: Least Squares
Date: 04/16/23 Time: 21:31
Sample (adjusted): 4/11/2018 4/06/2023
Included observations: 1120 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SP500_DLOG(-1)	-0.936434	0.029843	-31.37905	0.0000
C	0.000364	0.000359	1.014913	0.3104

R-squared	0.468289	Mean dependent var	-1.00E-05
Adjusted R-squared	0.467813	S.D. dependent var	0.016451
S.E. of regression	0.012001	Akaike info criterion	-6.005803
Sum squared resid	0.161030	Schwarz criterion	-5.996836
Log likelihood	3365.250	Hannan-Quinn criter.	-6.002414
F-statistic	984.6448	Durbin-Watson stat	2.004082
Prob(F-statistic)	0.000000		

Source Self Analysis

Figure 4.4 Stationarity of Crude Oil_Dlog

Augmented Dickey-Fuller Unit Root Test on OIL_DLOG

Null Hypothesis: OIL_DLOG has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=21)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-33.98961	0.0000
Test critical values:	1% level	-3.435988	
	5% level	-2.863918	
	10% level	-2.568087	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(OIL_DLOG)
Method: Least Squares
Date: 04/16/23 Time: 21:32
Sample (adjusted): 4/11/2018 4/06/2023
Included observations: 1117 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
OIL_DLOG(-1)	-0.986992	0.029038	-33.98961	0.0000
C	0.000428	0.001100	0.389378	0.6971

R-squared	0.508874	Mean dependent var	-0.000389
Adjusted R-squared	0.508434	S.D. dependent var	0.052435
S.E. of regression	0.036763	Akaike info criterion	-3.766839
Sum squared resid	1.506976	Schwarz criterion	-3.757853
Log likelihood	2105.779	Hannan-Quinn criter.	-3.763442
F-statistic	1155.294	Durbin-Watson stat	2.050997
Prob(F-statistic)	0.000000		

Source Self Analysis

Figure 4.5 Stationarity of Gold_Dlog

Augmented Dickey-Fuller Unit Root Test on GOLD_DLOG

Null Hypothesis: GOLD_DLOG has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=21)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-34.28257	0.0000
Test critical values:	1% level	-3.435973	
	5% level	-2.863911	
	10% level	-2.568083	

*MacKinnon (1996) one-sided p-values.

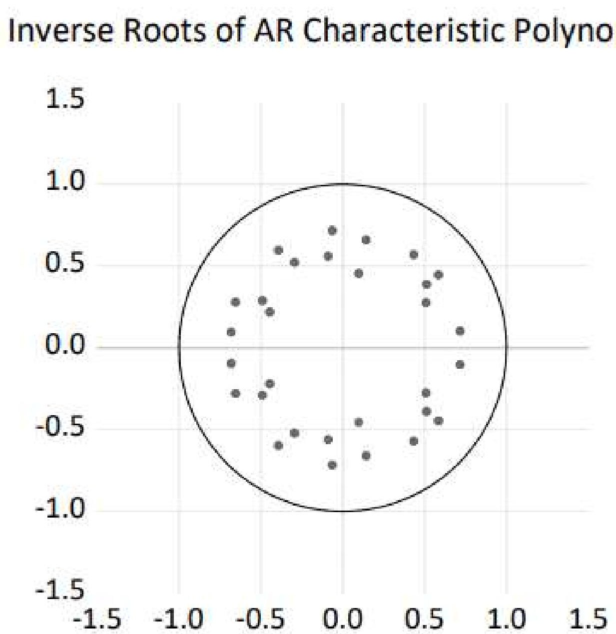
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(GOLD_DLOG)
Method: Least Squares
Date: 04/16/23 Time: 21:34
Sample (adjusted): 4/11/2018 4/06/2023
Included observations: 1120 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GOLD_DLOG(-1)	-1.024943	0.029897	-34.28257	0.0000
C	0.000375	0.000316	1.185100	0.2362

R-squared	0.512492	Mean dependent var	-1.43E-06
Adjusted R-squared	0.512056	S.D. dependent var	0.015131
S.E. of regression	0.010569	Akaike info criterion	-6.259930
Sum squared resid	0.124893	Schwarz criterion	-6.250964
Log likelihood	3507.561	Hannan-Quinn criter.	-6.256541
F-statistic	1175.295	Durbin-Watson stat	2.001241
Prob(F-statistic)	0.000000		

Source Self Analysis

Figure 4.6 Unit Root AR Polynomial



Source Self Analysis

To test that whether all the root lies within the unit root area this also used to identify whether the data is stationary or not. As shown in Figure 6 all the root lies in unit area and hence its stationary.

4.3 Vector Auto Regressive Model

It is necessary to analyse how the cost of crude oil, the stock market and currency rate, affect the price of gold. There exists an application of vector autoregression (VAR). The rest of the variables in this framework are thought of being endogenous, but the lag values of all other endogenous variables in the model are used to explain every endogenous variable. Since the model contains no external variables, the VAR greatly increases the model's adaptability by preventing the enactment of beforehand restrictions.

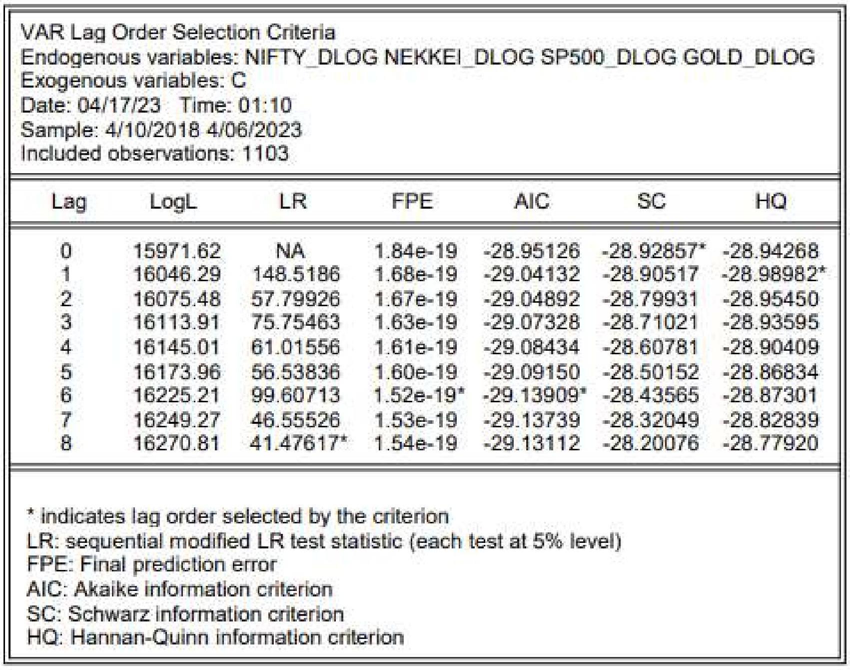
It is necessary to analyse how the the stock market, the cost of crude oil and exchange rate, The vector autoregression (VAR) provides a statistical method for analysing the changing impact of unpredictable shocks upon a group of variables in addition to predicting systems with linked time series. By modelling each endogenous variable within the framework as an indicator of the lagged values of each endogenous variables in the system, the VAR method avoids the necessity for structural modelling.

A VAR's mathematical structure is

$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + Bx_t + \varepsilon_t$$

Where Y_t is the vector of Endogenous Variable, $A_1 \dots A_p$, B are the coefficients matrix that is to be estimated. X_t is the exogenous variable and ε_t is the error term. A p, B, and t are vectors of improvements that may be at the same time associated with one another but have no correlation with their own delayed values as well as the right-side variables. There is no simultaneity issue because only lag values of the endogenous any serial correlation might be eliminated by adding more delayed y's, it should be noted that the belief that variables are included on the right-hand side of each equation, making OLS the best method. Since the changes are not sequentially connected is not restrictive.

Figure 4.7 Lag Order Selection Criteria for VAR



VAR Lag Order Selection Criteria
Endogenous variables: NIFTY_DLOG NEKKEI_DLOG SP500_DLOG GOLD_DLOG
Exogenous variables: C
Date: 04/17/23 Time: 01:10
Sample: 4/10/2018 4/06/2023
Included observations: 1103

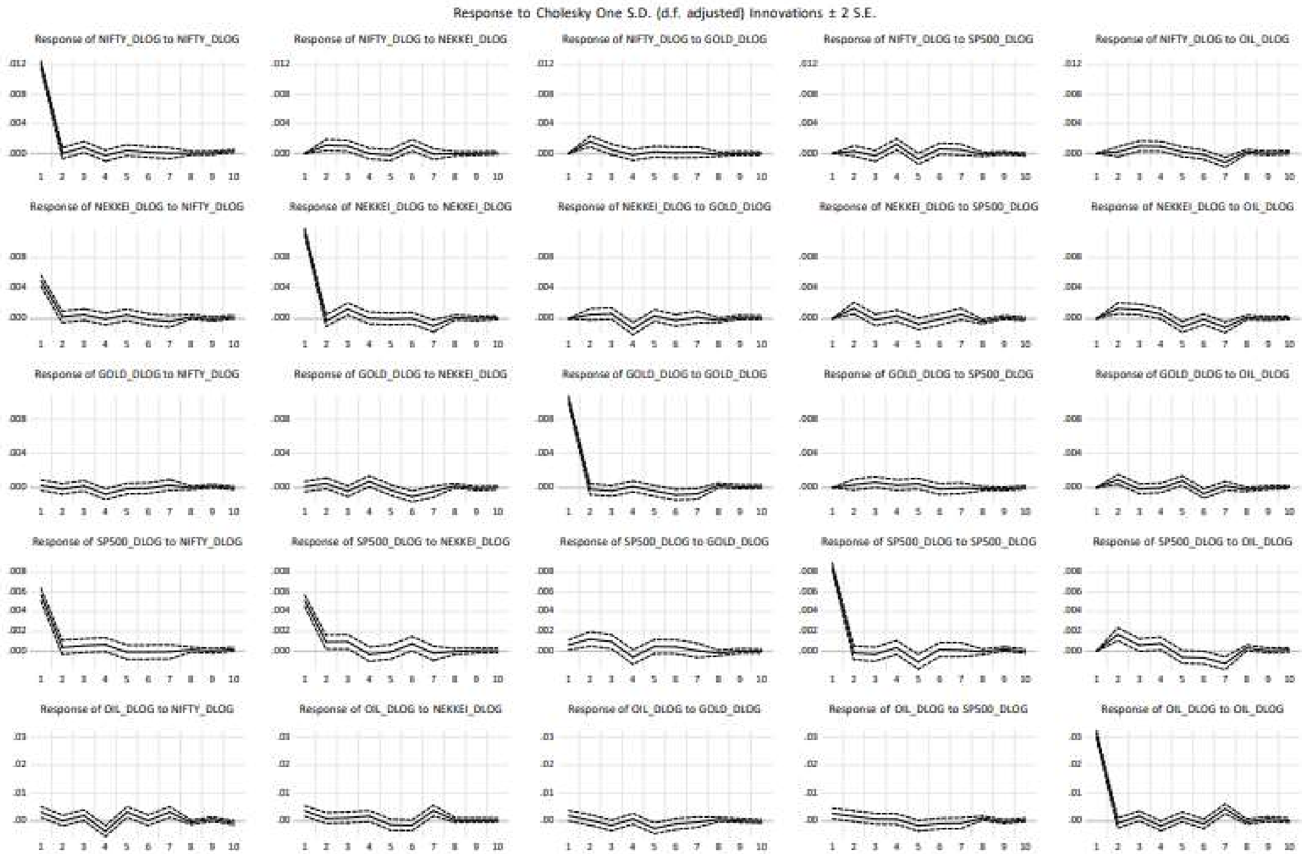
Lag	LogL	LR	FPE	AIC	SC	HQ
0	15971.62	NA	1.84e-19	-28.95126	-28.92857*	-28.94268
1	16046.29	148.5186	1.68e-19	-29.04132	-28.90517	-28.98982*
2	16075.48	57.79926	1.67e-19	-29.04892	-28.79931	-28.95450
3	16113.91	75.75463	1.63e-19	-29.07328	-28.71021	-28.93595
4	16145.01	61.01556	1.61e-19	-29.08434	-28.60781	-28.90409
5	16173.96	56.53836	1.60e-19	-29.09150	-28.50152	-28.86834
6	16225.21	99.60713	1.52e-19*	-29.13909*	-28.43565	-28.87301
7	16249.27	46.55526	1.53e-19	-29.13739	-28.32049	-28.82839
8	16270.81	41.47617*	1.54e-19	-29.13112	-28.20076	-28.77920

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

Source Self Analysis

The series or order of lag to be selected for the model is 6 as the test conducted highlighted that the LR test is significant at this lag level showing a 99.06 value.

Figure 4.8 Impulse Response of all the variables together in VAR



Source Self Analysis

The impulse response of Nifty 50 show that the index is largely not affected by the variables except Nifty 50 itself at one lag but shows slight variations with Nikkei around a lag of 6 while the gold has a slight response at lag of 2. The response of Nifty50 with S&P500 is at lag 4 and oil this response is seen around lag of 3. The impulse response of Nikkei225 show that the index is largely not affected by the variables except itself with one lag but shows good variations with Nifty 50 around a lag of 1 while the gold has a slight negative response at lag of 4. The response of Nikkei with S&P500 is at lag 4 and oil this response is seen around lag of 5. The impulse response of gold show that it is largely not affected by the variables except itself at one lag but shows no variations with Nifty 50 while the gold has a slight response with Nikkei at lag of 4. The response of Gold with S&P500 is not there and oil this response is seen around lag of 5. The impulse response of S&P500 show that it is largely affected by the variables along with itself with one lag its response high at lag 1 for Nifty 20 and Nikkei which having a negative response with gold and oil at lag of 4 and 7 respectively.

Figure 4.9 VAR Estimates

Vector Autoregression Estimates

Vector Autoregression Estimates
Date: 04/16/23 Time: 17:43
Sample (adjusted): 4/18/2018 4/06/2023
Included observations: 1107 after adjustments
Standard errors in ( ) & t-statistics in [ ]

	NIFTY_DLOG	NEKKEI_DL	GOLD_DLOG	SP500_DLOG	OIL_DLOG
NIFTY_DLOG(-1)	-0.053969	-0.017512	-0.043538	0.000554	-0.081292
	(0.03532)	(0.03641)	(0.03022)	(0.03394)	(0.09281)
	[-1.52817]	[-0.48097]	[-1.44066]	[0.01632]	[-0.87588]
NIFTY_DLOG(-2)	0.049677	0.002663	0.010794	0.030231	0.096429
	(0.03525)	(0.03634)	(0.03016)	(0.03387)	(0.09263)
	[1.40944]	[0.07329]	[0.35787]	[0.89245]	[1.04104]
NIFTY_DLOG(-3)	-0.059227	-0.017842	-0.105712	0.056349	-0.393833
	(0.03523)	(0.03632)	(0.03014)	(0.03385)	(0.09257)
	[-1.68137]	[-0.49128]	[-3.50697]	[1.66447]	[-4.25432]
NIFTY_DLOG(-4)	0.075353	0.066927	-0.017502	0.066619	0.319053
	(0.03569)	(0.03679)	(0.03054)	(0.03430)	(0.09378)
	[2.11162]	[1.81915]	[-0.57316]	[1.94250]	[3.40212]
NIFTY_DLOG(-5)	-0.023402	0.007198	0.022077	-0.039619	0.150988
	(0.03601)	(0.03713)	(0.03082)	(0.03461)	(0.09464)
	[-0.64986]	[0.19387]	[0.71643]	[-1.14474]	[1.59544]
NIFTY_DLOG(-6)	-0.039474	-0.043784	0.029664	-0.018120	0.067736
	(0.03573)	(0.03684)	(0.03058)	(0.03434)	(0.09390)
	[-1.10476]	[-1.18857]	[0.97018]	[-0.52768]	[0.72136]
NEKKEI_DLOG(-1)	0.080110	-0.103099	0.020362	0.079966	0.002608
	(0.03742)	(0.03857)	(0.03202)	(0.03596)	(0.09833)
	[2.14111]	[-2.67275]	[0.63597]	[2.22384]	[0.02652]
NEKKEI_DLOG(-2)	0.106864	0.092786	-0.065664	0.106616	0.048022
	(0.03790)	(0.03907)	(0.03243)	(0.03642)	(0.09960)
	[2.81961]	[2.37462]	[-2.02462]	[2.92701]	[0.48213]
NEKKEI_DLOG(-3)	-0.057261	-0.005756	0.032625	-0.038768	0.116537
	(0.03819)	(0.03937)	(0.03268)	(0.03670)	(0.10036)
	[-1.49946]	[-0.14621]	[0.99834]	[-1.05631]	[1.16121]
NEKKEI_DLOG(-4)	-0.008973	0.029709	-0.050982	0.028479	-0.030695
	(0.03804)	(0.03922)	(0.03255)	(0.03656)	(0.09997)
	[-0.23588]	[0.75753]	[-1.56616]	[0.77898]	[-0.30704]
NEKKEI_DLOG(-5)	0.065831	-0.002840	-0.078631	0.062794	-0.063342
	(0.03796)	(0.03914)	(0.03249)	(0.03649)	(0.09977)
	[1.73406]	[-0.07256]	[-2.42041]	[1.72107]	[-0.63489]
NEKKEI_DLOG(-6)	-0.006188	-0.101387	-0.053842	0.019094	0.274577
	(0.03770)	(0.03887)	(0.03226)	(0.03623)	(0.09908)
	[-0.16413]	[-2.60842]	[-1.66887]	[0.52696]	[2.77126]
GOLD_DLOG(-1)	0.153289	0.037247	-0.025354	0.111700	0.024448
	(0.03561)	(0.03671)	(0.03047)	(0.03422)	(0.09358)
	[4.30501]	[1.01464]	[-0.83208]	[3.26407]	[0.26127]
GOLD_DLOG(-2)	0.050898	0.051079	-0.040020	0.093904	-0.180904
	(0.03567)	(0.03678)	(0.03053)	(0.03428)	(0.09375)
	[1.42679]	[1.38885]	[-1.31099]	[2.73897]	[-1.92965]

Vector Autoregression Estimates

GOLD_DLOG(-3)	-0.036299	-0.131400	0.005037	-0.052297	0.027360
	(0.03564)	(0.03674)	(0.03050)	(0.03425)	(0.09365)
	[-1.01860]	[-3.57651]	[0.16519]	[-1.52697]	[0.29214]
GOLD_DLOG(-4)	0.035940	0.044488	-0.041374	0.055438	-0.209063
	(0.03550)	(0.03660)	(0.03038)	(0.03412)	(0.09329)
	[1.01242]	[1.21557]	[-1.36198]	[1.62492]	[-2.24094]
GOLD_DLOG(-5)	0.021218	-0.004887	-0.080452	0.077618	-0.132627
	(0.03567)	(0.03677)	(0.03052)	(0.03428)	(0.09374)
	[0.59485]	[-0.13289]	[-2.63570]	[2.26415]	[-1.41483]
GOLD_DLOG(-6)	0.032898	0.024195	-0.073352	0.042266	-0.073435
	(0.03575)	(0.03686)	(0.03059)	(0.03436)	(0.09395)
	[0.92021]	[0.65644]	[-2.39764]	[1.23011]	[-0.78161]
SP500_DLOG(-1)	0.033450	0.147094	0.028286	-0.038950	0.208657
	(0.04245)	(0.04377)	(0.03633)	(0.04080)	(0.11156)
	[0.78796]	[3.36092]	[0.77866]	[-0.95468]	[1.87032]
SP500_DLOG(-2)	-0.066382	-0.018525	0.066990	-0.071101	0.092803
	(0.04258)	(0.04390)	(0.03644)	(0.04092)	(0.11190)
	[-1.55895]	[-0.42199]	[1.83845]	[-1.73741]	[0.82930]
SP500_DLOG(-3)	0.096509	0.011030	0.046662	-0.003289	0.090678
	(0.04258)	(0.04390)	(0.03643)	(0.04092)	(0.11189)
	[2.26668]	[0.25127]	[1.28070]	[-0.08037]	[0.81040]
SP500_DLOG(-4)	-0.088281	-0.073751	0.052327	-0.137505	-0.168022
	(0.04260)	(0.04392)	(0.03646)	(0.04094)	(0.11196)
	[-2.07222]	[-1.67916]	[1.43535]	[-3.35840]	[-1.50074]
SP500_DLOG(-5)	0.054054	0.003212	-0.003927	0.016665	-0.122281
	(0.04280)	(0.04413)	(0.03663)	(0.04114)	(0.11249)
	[1.26283]	[0.07280]	[-0.10720]	[0.40512]	[-1.08704]
SP500_DLOG(-6)	0.077943	0.099301	0.026630	0.021080	-0.020129
	(0.04247)	(0.04378)	(0.03634)	(0.04081)	(0.11160)
	[1.83537]	[2.26807]	[0.73279]	[0.51648]	[-0.18036]
OIL_DLOG(-1)	0.006660	0.044356	0.029826	0.054351	-0.022490
	(0.01080)	(0.01113)	(0.00924)	(0.01038)	(0.02838)
	[0.61683]	[3.98460]	[3.22803]	[5.23760]	[-0.79257]
OIL_DLOG(-2)	0.022303	0.035301	-0.006277	0.015728	0.045868
	(0.01095)	(0.01129)	(0.00937)	(0.01052)	(0.02877)
	[2.03691]	[3.12714]	[-0.66990]	[1.49465]	[1.59403]
OIL_DLOG(-3)	0.026030	0.016398	-0.003023	0.016462	-0.057507
	(0.01068)	(0.01101)	(0.00914)	(0.01027)	(0.02808)
	[2.43634]	[1.48876]	[-0.33065]	[1.60326]	[-2.04816]
OIL_DLOG(-4)	0.000529	-0.033860	0.024428	-0.017569	0.023228
	(0.01056)	(0.01089)	(0.00904)	(0.01015)	(0.02775)
	[0.05014]	[-3.11037]	[2.70344]	[-1.73124]	[0.83704]
OIL_DLOG(-5)	-0.002742	-0.006111	-0.020183	-0.022866	-0.007089
	(0.01056)	(0.01089)	(0.00904)	(0.01015)	(0.02776)
	[-0.25956]	[-0.56121]	[-2.23291]	[-2.25245]	[-0.25539]
OIL_DLOG(-6)	-0.040716	-0.029931	0.013344	-0.042880	0.160093
	(0.01034)	(0.01066)	(0.00885)	(0.00994)	(0.02718)
	[-3.93686]	[-2.80717]	[1.50779]	[-4.31413]	[5.89024]

Vector Autoregression Estimates

C	0.000273 (0.00036) [0.75116]	0.000135 (0.00037) [0.36020]	0.000439 (0.00031) [1.41004]	0.000227 (0.00035) [0.64877]	2.91E-05 (0.00096) [0.03045]
R-squared	0.098692	0.097639	0.074308	0.104572	0.115407
Adj. R-squared	0.073563	0.072480	0.048499	0.079607	0.090744
Sum sq. resids	0.155474	0.165253	0.113850	0.143603	1.073770
S.E. equation	0.012020	0.012393	0.010286	0.011552	0.031590
F-statistic	3.927372	3.880914	2.879118	4.188680	4.679306
Log likelihood	3339.160	3305.398	3511.630	3383.121	2269.547
Akaike AIC	-5.976803	-5.915805	-6.288402	-6.056226	-4.044349
Schwarz SC	-5.836521	-5.775524	-6.148120	-5.915944	-3.904067
Mean dependent	0.000406	0.000194	0.000365	0.000348	0.000147
S.D. dependent	0.012489	0.012868	0.010545	0.012042	0.033129
Determinant resid covariance (dof adj.)	1.40E-19				
Determinant resid covariance	1.21E-19				
Log likelihood	16254.46				
Akaike information criterion	-29.08666				
Schwarz criterion	-28.38525				
Number of coefficients	155				

Source Self Analysis

VAR is calculated by selecting the variable in the order Nifty 50dlog, Nikkei 225dlog, S&P 500dlog, Golddlog and Oildlog. The following coefficient are calculated with 6 lag which was explained earlier as to why we use 6 lags. From this we create a rudimentary equation for all the variables with its coefficient. These rudimentary equations can be seen in figure 10 to 14.

Figure 4.10 VAR Nifty 50 Rudimentary Equation

```
Equation: NIFTY_DLOG = C(1)*NIFTY_DLOG(-1) + C(2)*NIFTY_DLOG(-2) +
    C(3)*NIFTY_DLOG(-3) + C(4)*NIFTY_DLOG(-4) + C(5)*NIFTY_DLOG(
    -5) + C(6)*NIFTY_DLOG(-6) + C(7)*NEKKEI_DLOG(-1) + C(8)
    *NEKKEI_DLOG(-2) + C(9)*NEKKEI_DLOG(-3) + C(10)
    *NEKKEI_DLOG(-4) + C(11)*NEKKEI_DLOG(-5) + C(12)
    *NEKKEI_DLOG(-6) + C(13)*SP500_DLOG(-1) + C(14)*SP500_DLOG(
    -2) + C(15)*SP500_DLOG(-3) + C(16)*SP500_DLOG(-4) + C(17)
    *SP500_DLOG(-5) + C(18)*SP500_DLOG(-6) + C(19)*GOLD_DLOG(
    -1) + C(20)*GOLD_DLOG(-2) + C(21)*GOLD_DLOG(-3) + C(22)
    *GOLD_DLOG(-4) + C(23)*GOLD_DLOG(-5) + C(24)*GOLD_DLOG(-6)
    + C(25)*OIL_DLOG(-1) + C(26)*OIL_DLOG(-2) + C(27)*OIL_DLOG(-3)
    + C(28)*OIL_DLOG(-4) + C(29)*OIL_DLOG(-5) + C(30)*OIL_DLOG(-6)
    + C(31)
```

Source Self Analysis

The correct/ significant coefficients are selected on the bases on the probability the value of probability should be greater than 1.67 (90%) in case there are 2 or more variable satisfying this condition then the coefficient with the greatest t value is chosen

Figure 4.11 VAR Nikkei 225 Rudimentary Equation

```
Equation: NEKKEI_DLOG = C(32)*NIFTY_DLOG(-1) + C(33)*NIFTY_DLOG(
        -2) + C(34)*NIFTY_DLOG(-3) + C(35)*NIFTY_DLOG(-4) + C(36)
        *NIFTY_DLOG(-5) + C(37)*NIFTY_DLOG(-6) + C(38)*NEKKEI_DLOG(
        -1) + C(39)*NEKKEI_DLOG(-2) + C(40)*NEKKEI_DLOG(-3) + C(41)
        *NEKKEI_DLOG(-4) + C(42)*NEKKEI_DLOG(-5) + C(43)
        *NEKKEI_DLOG(-6) + C(44)*SP500_DLOG(-1) + C(45)*SP500_DLOG(
        -2) + C(46)*SP500_DLOG(-3) + C(47)*SP500_DLOG(-4) + C(48)
        *SP500_DLOG(-5) + C(49)*SP500_DLOG(-6) + C(50)*GOLD_DLOG(
        -1) + C(51)*GOLD_DLOG(-2) + C(52)*GOLD_DLOG(-3) + C(53)
        *GOLD_DLOG(-4) + C(54)*GOLD_DLOG(-5) + C(55)*GOLD_DLOG(-6)
        + C(56)*OIL_DLOG(-1) + C(57)*OIL_DLOG(-2) + C(58)*OIL_DLOG(-3)
        + C(59)*OIL_DLOG(-4) + C(60)*OIL_DLOG(-5) + C(61)*OIL_DLOG(-6)
        + C(62)
```

Source Self Analysis

Figure 4.12 VAR S&P 500 Rudimentary Equation

```
Equation: SP500_DLOG = C(63)*NIFTY_DLOG(-1) + C(64)*NIFTY_DLOG(
        -2) + C(65)*NIFTY_DLOG(-3) + C(66)*NIFTY_DLOG(-4) + C(67)
        *NIFTY_DLOG(-5) + C(68)*NIFTY_DLOG(-6) + C(69)*NEKKEI_DLOG(
        -1) + C(70)*NEKKEI_DLOG(-2) + C(71)*NEKKEI_DLOG(-3) + C(72)
        *NEKKEI_DLOG(-4) + C(73)*NEKKEI_DLOG(-5) + C(74)
        *NEKKEI_DLOG(-6) + C(75)*SP500_DLOG(-1) + C(76)*SP500_DLOG(
        -2) + C(77)*SP500_DLOG(-3) + C(78)*SP500_DLOG(-4) + C(79)
        *SP500_DLOG(-5) + C(80)*SP500_DLOG(-6) + C(81)*GOLD_DLOG(
        -1) + C(82)*GOLD_DLOG(-2) + C(83)*GOLD_DLOG(-3) + C(84)
        *GOLD_DLOG(-4) + C(85)*GOLD_DLOG(-5) + C(86)*GOLD_DLOG(-6)
        + C(87)*OIL_DLOG(-1) + C(88)*OIL_DLOG(-2) + C(89)*OIL_DLOG(-3)
        + C(90)*OIL_DLOG(-4) + C(91)*OIL_DLOG(-5) + C(92)*OIL_DLOG(-6)
        + C(93)
```

Source Self Analysis

Figure 4.13 VAR Gold Rudimentary Equation

```
Equation: GOLD_DLOG = C(94)*NIFTY_DLOG(-1) + C(95)*NIFTY_DLOG(
        -2) + C(96)*NIFTY_DLOG(-3) + C(97)*NIFTY_DLOG(-4) + C(98)
        *NIFTY_DLOG(-5) + C(99)*NIFTY_DLOG(-6) + C(100)*NEKKEI_DLOG(
        -1) + C(101)*NEKKEI_DLOG(-2) + C(102)*NEKKEI_DLOG(-3) + C(103)
        *NEKKEI_DLOG(-4) + C(104)*NEKKEI_DLOG(-5) + C(105)
        *NEKKEI_DLOG(-6) + C(106)*SP500_DLOG(-1) + C(107)
        *SP500_DLOG(-2) + C(108)*SP500_DLOG(-3) + C(109)
        *SP500_DLOG(-4) + C(110)*SP500_DLOG(-5) + C(111)
        *SP500_DLOG(-6) + C(112)*GOLD_DLOG(-1) + C(113)*GOLD_DLOG(
        -2) + C(114)*GOLD_DLOG(-3) + C(115)*GOLD_DLOG(-4) + C(116)
        *GOLD_DLOG(-5) + C(117)*GOLD_DLOG(-6) + C(118)*OIL_DLOG(-1)
        + C(119)*OIL_DLOG(-2) + C(120)*OIL_DLOG(-3) + C(121)*OIL_DLOG(
        -4) + C(122)*OIL_DLOG(-5) + C(123)*OIL_DLOG(-6) + C(124)
Observations: 1108
```

Source Self Analysis

Figure 4.14 VAR Crude Oil Rudimentary Equation

Equation: OIL_DLOG = C(125)*NIFTY_DLOG(-1) + C(126)*NIFTY_DLOG(-2) + C(127)*NIFTY_DLOG(-3) + C(128)*NIFTY_DLOG(-4) + C(129)*NIFTY_DLOG(-5) + C(130)*NIFTY_DLOG(-6) + C(131)*NEKKEI_DLOG(-1) + C(132)*NEKKEI_DLOG(-2) + C(133)*NEKKEI_DLOG(-3) + C(134)*NEKKEI_DLOG(-4) + C(135)*NEKKEI_DLOG(-5) + C(136)*NEKKEI_DLOG(-6) + C(137)*SP500_DLOG(-1) + C(138)*SP500_DLOG(-2) + C(139)*SP500_DLOG(-3) + C(140)*SP500_DLOG(-4) + C(141)*SP500_DLOG(-5) + C(142)*SP500_DLOG(-6) + C(143)*GOLD_DLOG(-1) + C(144)*GOLD_DLOG(-2) + C(145)*GOLD_DLOG(-3) + C(146)*GOLD_DLOG(-4) + C(147)*GOLD_DLOG(-5) + C(148)*GOLD_DLOG(-6) + C(149)*OIL_DLOG(-1) + C(150)*OIL_DLOG(-2) + C(151)*OIL_DLOG(-3) + C(152)*OIL_DLOG(-4) + C(153)*OIL_DLOG(-5) + C(154)*OIL_DLOG(-6) + C(155)

Source Self Analysis

Figure 4.15 VAR Coefficients with value greater than 1.67

	A	B	C	D	E	F	G	H	I
1	co	coval	lo	to	P	equn	What	Lag	Represents
5	C(4)	0.073443	0.035602	2.062917	0.0392	Nifty	Nifty	-4	Nifty(-4)impacts Nifty equation.
8	C(7)	0.081552	0.037367	2.182437	0.0291	Nifty	Nekki	-1	Nekki(-1)impacts Nifty equation.
9	C(8)	0.106627	0.037893	2.813873	0.0049	Nifty	Nekki	-2	Nekki(-2)impacts Nifty equation.
16	C(15)	0.096652	0.04257	2.270422	0.0232	Nifty	S&P500	-3	S&P500(-3)impacts Nifty equation.
17	C(16)	-0.08634	0.042528	-2.0301	0.0424	Nifty	S&P500	-4	S&P500(-4)impacts Nifty equation.
20	C(19)	0.151733	0.03555	4.268176	0	Nifty	Gold	-1	Gold(-1)impacts Nifty equation.
27	C(26)	0.022274	0.010948	2.034652	0.0419	Nifty	OIL	-2	OIL(-2)impacts Nifty equation.
28	C(27)	0.025004	0.010607	2.357269	0.0184	Nifty	OIL	-3	OIL(-3)impacts Nifty equation.
31	C(30)	-0.0405	0.010337	-3.91806	0.0001	Nifty	OIL	-6	OIL(-6)impacts Nifty equation.
39	C(38)	-0.10266	0.038514	-2.66559	0.0077	Nekkei	Nekki	-1	Nekki(-1)impacts Nekkei equation.
40	C(39)	0.092714	0.039056	2.373889	0.0176	Nekkei	Nekki	-2	Nekki(-2)impacts Nekkei equation.
44	C(43)	-0.10064	0.038727	-2.59879	0.0094	Nekkei	Nekki	-6	Nekki(-6)impacts Nekkei equation.
45	C(44)	0.146989	0.043745	3.360168	0.0008	Nekkei	S&P500	-1	S&P500(-1)impacts Nekkei equation.
50	C(49)	0.098453	0.043619	2.257121	0.024	Nekkei	S&P500	-6	S&P500(-6)impacts Nekkei equation.
53	C(52)	-0.13158	0.036716	-3.58364	0.0003	Nekkei	Gold	-3	Gold(-3)impacts Nekkei equation.
57	C(56)	0.044115	0.011081	3.981089	0.0001	Nekkei	OIL	-1	OIL(-1)impacts Nekkei equation.
58	C(57)	0.035292	0.011283	3.127767	0.0018	Nekkei	OIL	-2	OIL(-2)impacts Nekkei equation.
60	C(59)	-0.03398	0.01087	-3.12599	0.0018	Nekkei	OIL	-4	OIL(-4)impacts Nekkei equation.
62	C(61)	-0.02987	0.010654	-2.80322	0.0051	Nekkei	OIL	-6	OIL(-6)impacts Nekkei equation.
70	C(69)	0.081576	0.035916	2.271287	0.0232	S&P500	Nekki	-1	Nekki(-1)impacts S&P500 equation.
71	C(70)	0.106351	0.036422	2.919969	0.0035	S&P500	Nekki	-2	Nekki(-2)impacts S&P500 equation.
79	C(78)	-0.13533	0.040877	-3.31075	0.0009	S&P500	S&P500	-4	S&P500(-4)impacts S&P500 equation.
82	C(81)	0.109961	0.034169	3.218107	0.0013	S&P500	Gold	-1	Gold(-1)impacts S&P500 equation.
83	C(82)	0.09337	0.034278	2.723904	0.0065	S&P500	Gold	-2	Gold(-2)impacts S&P500 equation.
86	C(85)	0.079783	0.034202	2.332685	0.0197	S&P500	Gold	-5	Gold(-5)impacts S&P500 equation.
88	C(87)	0.053464	0.010334	5.173724	0	S&P500	OIL	-1	OIL(-1)impacts S&P500 equation.
92	C(91)	-0.0238	0.010102	-2.35633	0.0185	S&P500	OIL	-5	OIL(-5)impacts S&P500 equation.
93	C(92)	-0.04264	0.009936	-4.2917	0	S&P500	OIL	-6	OIL(-6)impacts S&P500 equation.

Source Self Analysis

Eviews churn 92 coefficients for all independent variable, it becomes essential to choose only the statistically significant out of the bunch in order to make the equations not only precise but also to reduce over fitting, the excels given below show the statistical significance and t-values of each coefficient. Following were chosen:

For **NIFTY** equation:

C(4) for the nifty coefficient: 0.073443; It represents NIFTY at lag 4

C(8) for the Nikkei coefficient: 0.106627 It represents the Nikkei at lag 2

C(15) for the S&P coefficient: 0.0966 It represents the S&P at lag 3

C(19) for the Gold coefficient: 0.15 It represents the Gold at lag 1

C(27) for the OIL coefficient: 0.025 It represents the OIL at lag 3

For **Nikkei** equation:

C(39) for the Nikkei coefficient: -0.0927; It represents the Nikkei at lag 2

C(44) for the S&P coefficient: 0.1469 It represents the S&P at lag 1

C(52) for the Gold coefficient: -0.13158 It represents the Gold at lag 3

C(56) for the OIL coefficient: 0.044115 It represents the OIL at lag 1

For **S&P 500** equation:

C(70) for the Nikkei coefficient: 0.106351 It represents the Nikkei at lag 2

C(78) for the S&P coefficient: -0.135 It represents the S&P500 at lag 4

C(81) for the Gold coefficient: 0.109961 It represents the Gold at lag 1

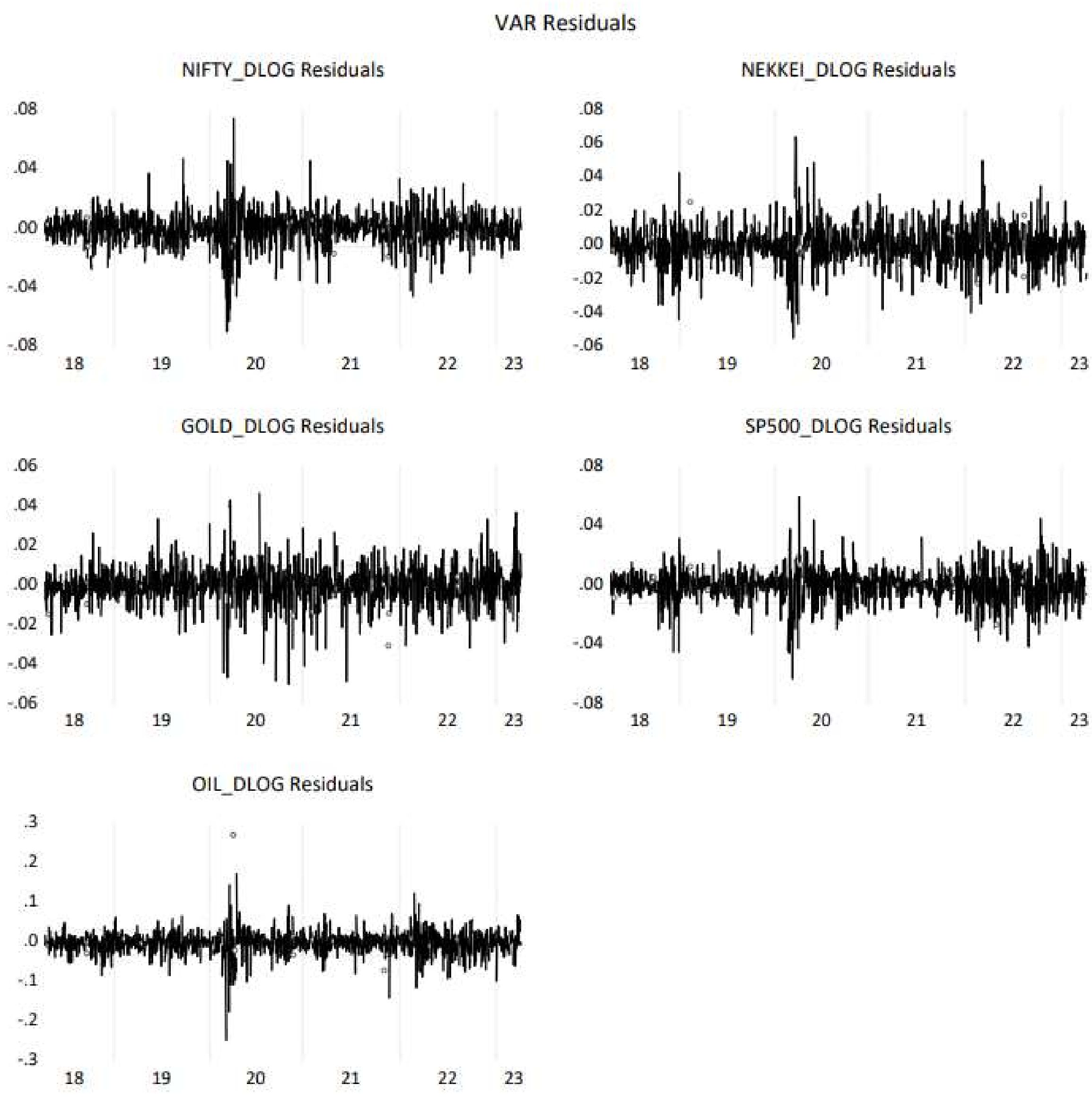
C(87) for the OIL coefficient: 0.05346 It represents the OIL at lag 1

$$\boldsymbol{Nifty_t = 0.073443 * (Nifty_{t-4}) +. 0.106627 * Nikkei_{t-2} +. 0.0966 * S\&P500_{t-3} + 0.15 * Gold_{t-1} + 0.025 * Oil_{t-3}}$$

$$Nikkei_t = -0.0927 * Nikkei_{t-2} + 0.1469 * S\&P500_{t-4} - 0.13158 * Gold_{t-3} + 0.044115 * Oil_{t-1}$$

$$S\&P500_t = 0.106351 * Nikkei_{t-2} - 0.135 * S\&P500_{t-4} + 0.109961 * Gold_{t-3} + 0.044115 * Oil_{t-1}$$

Figure 4.16 VAR Residue showing normal distribution



Source Self Analysis

Figure 4.17 VAR Test for Heteroskedasticity

VAR Residual Heteroskedasticity Tests (Levels and Squares)
Date: 04/16/23 Time: 18:00
Sample: 4/10/2018 4/06/2023
Included observations: 1107

Joint test:

Chi-sq	df	Prob.
4390.388	900	0.0000

Individual components:

Dependent	R-squared	F(60,1046)	Prob.	Chi-sq(60)	Prob.
res1*res1	0.487006	16.55017	0.0000	539.1157	0.0000
res2*res2	0.364090	9.981453	0.0000	403.0478	0.0000
res3*res3	0.150009	3.076689	0.0000	166.0600	0.0000
res4*res4	0.289412	7.100329	0.0000	320.3788	0.0000
res5*res5	0.438161	13.59573	0.0000	485.0445	0.0000
res2*res1	0.369907	10.23456	0.0000	409.4875	0.0000
res3*res1	0.302659	7.566388	0.0000	335.0434	0.0000
res3*res2	0.328231	8.518044	0.0000	363.3516	0.0000
res4*res1	0.392111	11.24516	0.0000	434.0672	0.0000
res4*res2	0.209663	4.624757	0.0000	232.0965	0.0000
res4*res3	0.189268	4.069872	0.0000	209.5198	0.0000
res5*res1	0.486085	16.48924	0.0000	538.0957	0.0000
res5*res2	0.309596	7.817591	0.0000	342.7230	0.0000
res5*res3	0.182659	3.896006	0.0000	202.2040	0.0000
res5*res4	0.253929	5.933529	0.0000	281.0997	0.0000

Source Self Analysis

Figure 4.18 VAR Johansen Cointegration Test

Johansen Cointegration Test

Date: 04/16/23 Time: 18:04
Sample (adjusted): 4/19/2018 4/06/2023
Included observations: 1105 after adjustments
Trend assumption: Linear deterministic trend
Series: NIFTY_DLOG NEKKEI_DLOG GOLD_DLOG SP500_DLOG OIL_DLOG
Lags interval (in first differences): 1 to 6

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.208526	865.2896	69.81889	0.0000
At most 1 *	0.164150	606.8759	47.85613	0.0000
At most 2 *	0.136526	408.7427	29.79707	0.0000
At most 3 *	0.115248	246.5382	15.49471	0.0000
At most 4 *	0.095762	111.2328	3.841465	0.0000

Trace test indicates 5 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None *	0.208526	258.4137	33.87687	0.0000
At most 1 *	0.164150	198.1332	27.58434	0.0000
At most 2 *	0.136526	162.2045	21.13162	0.0000
At most 3 *	0.115248	135.3054	14.26460	0.0000
At most 4 *	0.095762	111.2328	3.841465	0.0000

Max-eigenvalue test indicates 5 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegrating Coefficients (normalized by b'*S11*b=I):

NIFTY_DLOG	NEKKEI_DLOG	GOLD_DLOG	SP500_DLOG	OIL_DLOG
0.115885	-268.0809	-187.7082	210.8641	22.51572
189.0452	30.72339	-83.61162	-222.6909	-28.41376
-117.5870	148.2357	-204.6490	38.61028	-12.28126
48.57958	-56.73700	1.102000	199.4423	-66.67534
-152.2010	-28.61557	34.05239	32.14166	-37.58186

Unrestricted Adjustment Coefficients (alpha):

D(NIFTY_DL	-0.000329	-0.002136	0.000789	-0.001598	0.003084
D(NEKKEI_D	0.003363	0.000993	-0.002356	-0.000964	0.002438
D(GOLD_DLO	0.002705	0.001570	0.003022	-0.000790	-2.53E-05
D(SP500_DL	-0.000812	0.002308	-0.000877	-0.002564	0.002151
D(OIL_DLOG)	-0.002506	0.005995	0.002024	0.006528	0.006199

1 Cointegrating Equation(s): Log likelihood 15978.88

Normalized cointegrating coefficients (standard error in parentheses)

NIFTY_DLOG	NEKKEI_DLOG	GOLD_DLOG	SP500_DLOG	OIL_DLOG
1.000000	-2313.344	-1619.786	1819.605	194.2944
	(160.237)	(149.753)	(172.577)	(43.8365)

Johansen Cointegration Test

Adjustment coefficients (standard error in parentheses)

D(NIFTY_DL	-3.82E-05
	(4.4E-05)
D(NEKKEI_D	0.000390
	(4.5E-05)
D(GOLD_DLO	0.000313
	(3.8E-05)
D(SP500_DL	-9.41E-05
	(4.3E-05)
D(OIL_DLOG)	-0.000290
	(0.00011)

2 Cointegrating Equation(s): Log likelihood 16077.95

Normalized cointegrating coefficients (standard error in parentheses)

NIFTY_DLOG	NEKKEI_DLOG	GOLD_DLOG	SP500_DLOG	OIL_DLOG
1.000000	0.000000	-0.556039	-1.050071	-0.136642
		(0.10501)	(0.08557)	(0.03091)
0.000000	1.000000	0.699952	-0.787023	-0.084048
		(0.06362)	(0.05185)	(0.01873)

Adjustment coefficients (standard error in parentheses)

D(NIFTY_DL	-0.403924	0.022654
	(0.07062)	(0.10079)
D(NEKKEI_D	0.188199	-0.870957
	(0.07271)	(0.10378)
D(GOLD_DLO	0.297100	-0.676947
	(0.06125)	(0.08743)
D(SP500_DL	0.436169	0.288576
	(0.06843)	(0.09767)
D(OIL_DLOG)	1.132944	0.855895
	(0.18322)	(0.26152)

3 Cointegrating Equation(s): Log likelihood 16159.05

Normalized cointegrating coefficients (standard error in parentheses)

NIFTY_DLOG	NEKKEI_DLOG	GOLD_DLOG	SP500_DLOG	OIL_DLOG
1.000000	0.000000	0.000000	-1.097376	-0.113005
			(0.07391)	(0.02720)
0.000000	1.000000	0.000000	-0.727474	-0.113802
			(0.05146)	(0.01894)
0.000000	0.000000	1.000000	-0.085076	0.042510
			(0.05856)	(0.02155)

Adjustment coefficients (standard error in parentheses)

D(NIFTY_DL	-0.496680	0.139587	0.079021
	(0.08299)	(0.11476)	(0.10811)
D(NEKKEI_D	0.465232	-1.220197	-0.232125
	(0.08412)	(0.11632)	(0.10957)
D(GOLD_DLO	-0.058251	-0.228974	-1.257485
	(0.06915)	(0.09562)	(0.09007)
D(SP500_DL	0.539247	0.158632	0.138859
	(0.08037)	(0.11114)	(0.10469)
D(OIL_DLOG)	0.894959	1.155910	-0.445066
	(0.21533)	(0.29778)	(0.28050)

4 Cointegrating Equation(s): Log likelihood 16226.70

Normalized cointegrating coefficients (standard error in parentheses)

NIFTY_DLOG	NEKKEI_DLOG	GOLD_DLOG	SP500_DLOG	OIL_DLOG
1.000000	0.000000	0.000000	0.000000	-0.464095

Johansen Cointegration Test

				(0.03934)
0.000000	1.000000	0.000000	0.000000	-0.346547
				(0.03041)
0.000000	0.000000	1.000000	0.000000	0.015291
				(0.01977)
0.000000	0.000000	0.000000	1.000000	-0.319936
				(0.03160)

Adjustment coefficients (standard error in parentheses)

D(NIFTY_DL	-0.574305	0.230247	0.077260	0.118086
	(0.08421)	(0.11569)	(0.10718)	(0.13595)
D(NEKKEI_D	0.418398	-1.165499	-0.233187	0.204602
	(0.08583)	(0.11792)	(0.10924)	(0.13857)
D(GOLD_DLO	-0.096618	-0.184165	-1.258356	0.179962
	(0.07056)	(0.09694)	(0.08980)	(0.11391)
D(SP500_DL	0.414701	0.304091	0.136034	-1.230292
	(0.08030)	(0.11032)	(0.10220)	(0.12964)
D(OIL_DLOG)	1.212099	0.785516	-0.437872	-0.483128
	(0.21567)	(0.29630)	(0.27449)	(0.34817)

Source Self Analysis

Figure 4.19 VAR Granger Causality test

VAR Granger Causality/Block Exogeneity Wald Tests
Date: 04/16/23 Time: 18:01
Sample: 4/10/2018 4/06/2023
Included observations: 1107

Dependent variable: NIFTY_DLOG

Excluded	Chi-sq	df	Prob.
NEKKEI_DLOG	18.05774	6	0.0061
GOLD_DLOG	23.21555	6	0.0007
SP500_DLOG	20.01030	6	0.0028
OIL_DLOG	27.57183	6	0.0001
All	109.4007	24	0.0000

Dependent variable: NEKKEI_DLOG

Excluded	Chi-sq	df	Prob.
NIFTY_DLOG	5.274911	6	0.5091
GOLD_DLOG	18.15334	6	0.0059
SP500_DLOG	21.16231	6	0.0017
OIL_DLOG	45.51307	6	0.0000
All	95.09836	24	0.0000

Dependent variable: GOLD_DLOG

Excluded	Chi-sq	df	Prob.
NIFTY_DLOG	17.06757	6	0.0090
NEKKEI_DLOG	17.22000	6	0.0085
SP500_DLOG	6.804450	6	0.3393
OIL_DLOG	24.82572	6	0.0004
All	66.45427	24	0.0000

Dependent variable: SP500_DLOG

Excluded	Chi-sq	df	Prob.
NIFTY_DLOG	8.837077	6	0.1830
NEKKEI_DLOG	18.00883	6	0.0062
GOLD_DLOG	28.59127	6	0.0001
OIL_DLOG	55.90810	6	0.0000
All	111.1902	24	0.0000

Dependent variable: OIL_DLOG

Excluded	Chi-sq	df	Prob.
NIFTY_DLOG	37.77684	6	0.0000
NEKKEI_DLOG	10.28203	6	0.1133
GOLD_DLOG	10.96160	6	0.0896
SP500_DLOG	8.430912	6	0.2082
All	83.94923	24	0.0000

Source Self Analysis

Figure 4.20 No Auto Correlation at lag 6

VAR Residual Serial Correlation LM Tests
Date: 04/16/23 Time: 17:45
Sample: 4/10/2018 4/06/2023
Included observations: 1107

Null hypothesis: No serial correlation at lag h

Lag	LRE* stat	df	Prob.	Rao F-stat	df	Prob.
1	108.0267	25	0.0000	4.367628	(25, 3965.2)	0.0000
2	70.63651	25	0.0000	2.842459	(25, 3965.2)	0.0000
3	41.96095	25	0.0181	1.682445	(25, 3965.2)	0.0181
4	56.77225	25	0.0003	2.280563	(25, 3965.2)	0.0003
5	56.43447	25	0.0003	2.266898	(25, 3965.2)	0.0003
6	38.52454	25	0.0411	1.543992	(25, 3965.2)	0.0411
7	34.70212	25	0.0937	1.390127	(25, 3965.2)	0.0937

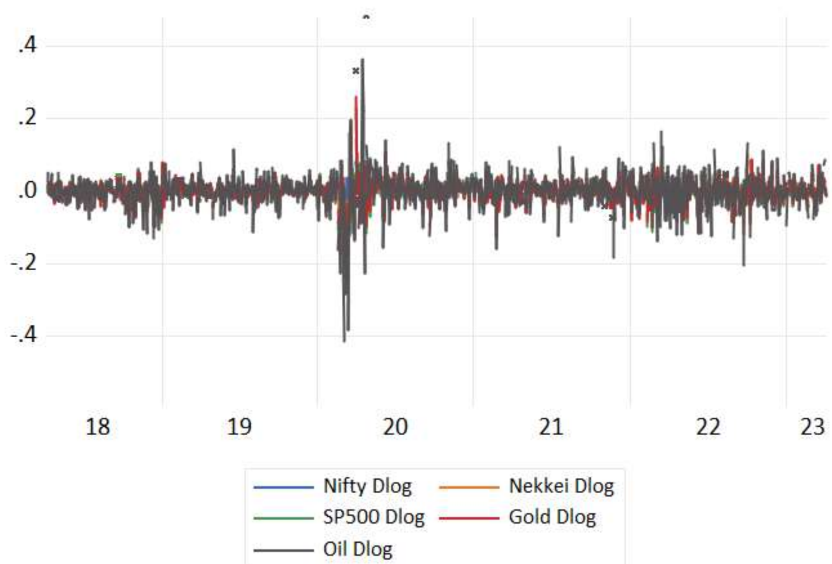
Null hypothesis: No serial correlation at lags 1 to h

Lag	LRE* stat	df	Prob.	Rao F-stat	df	Prob.
1	108.0267	25	0.0000	4.367628	(25, 3965.2)	0.0000
2	124.2277	50	0.0000	2.504109	(50, 4846.8)	0.0000
3	158.7717	75	0.0000	2.134851	(75, 5067.1)	0.0000
4	203.7713	100	0.0000	2.058640	(100, 5136.9)	0.0000
5	251.2472	125	0.0000	2.034880	(125, 5157.5)	0.0000
6	287.6229	150	0.0000	1.943293	(150, 5157.7)	0.0000
7	322.6656	175	0.0000	1.870389	(175, 5148.1)	0.0000

*Edgeworth expansion corrected likelihood ratio statistic.

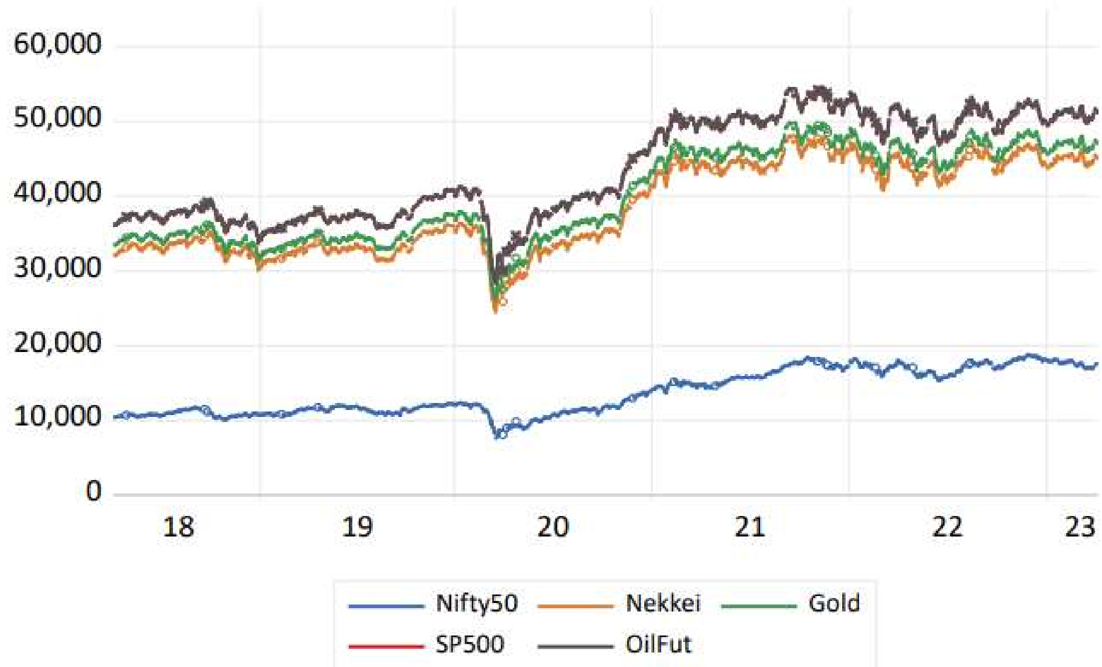
Source Self Analysis

Figure 4.21 Distribution of Dlog data (Stationary Data)



Source Self Analysis

Figure 4.22 Chart representing data of all variables



Source Self Analysis

4.4 Limitations

The model had a very poor square score while it showed full impulse response however considerable cointegration that existed between the variable as per the Johansen Cointegration test there were 5 cointegrating variable coupled with the fact the R square was very poor of the model, F statistic was low AIC was negative it represented a poorly fitted model. Thus, in order to improve this model VECM (Vector Error Correction Model) was adopted

However, this exercise was not in full futility as it gave a rudimentary and informed us about the lags that can be utilised in VECM model moreover it also told us how variables interact with each other that is by looking at the impulse responses we could identify which were the major impact factors of each and every dependent variable.

4.5 Vector Error Correction Model

A model of statistics called the Vector Error Correction Model (VECM) is employed to examine long-term correlations among factors which are non-stationary, implying they're likely to go through cycles or patterns over time. The Vector Autoregression (VAR) model, that implies that every one of the system's variables are stationary, is extended by the VECM.

The idea of a correction of errors phrase, that reflects short-term departures from the long-term optimum connection between the variables, is introduced by the VECM. The framework can reflect each short-run dynamic alongside equilibrium over time connections because to the error correction phrase, which measures how quickly the framework adapts to changes in the long-run equilibrium.

It is possible to write the fundamental formula pertaining to a VECM having p lags as follows:

$$\Delta yt = \alpha + \Pi yt\text{-}1 + \Gamma 1\Delta yt\text{-}1 + \Gamma 2\Delta yt\text{-}2 + \ldots + \Gamma p\Delta yt\text{-}p + \varepsilon t$$

Where

Δyt is first difference of non-stationary variable yt in vector form

Π is the long-term equilibrium matrix of relationship between variables

A is the constant term

Et is the white noise vector terms with zero and constant variance matrix

Γ1, Γ2, ..., Γp are short term dynamics that capture the adjustment of speed of the deviations in the system from long-term equilibrium.(Maysami & Koh, 2000)

Using either the Maximum Likelihood (ML) approach or the OLS (Ordinary Least Squares) technique, the VECM calculates the coefficients of Π and Γ matrices. The calculated parameters may be used to analyse the causal connections between the variables and forecast how the framework will behave in the years to come.

VAR model helped in creating a rudimentary model which is stable, however it was a poorly fit as the R^2 was low and there was cointegration among all 5-regressing equation.

VECM or Vector Error Correction Method is an improvement over VAR, which provides for an error correction coefficient which compensates for cointegration.

$$\begin{bmatrix} \Delta Y_t \\ \Delta X_t \end{bmatrix} = \begin{bmatrix} \varphi_1 \\ \varphi_2 \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} [Y_{t-1} - \beta_0 - \beta_1 X_{t-1}] + \sum_{i=1}^{\kappa} \begin{bmatrix} \delta_{1j} & \rho_{1j} \\ \delta_{2j} & \rho_{2j} \end{bmatrix} \begin{bmatrix} \Delta Y_{t-i} \\ \Delta X_{t-i} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

Following were the specification of the model:

Few important points

Period chosen: 4/13/2018 to 4/06/2022

Dependent Variable: NIFTY 50, Nikkei 250 and S&P500

Independent Variable: Depending on the dependent variable other two indices and Gold and OIL

From the VAR model we know that the VECM model needs to be made at the lags of 6 using the AIC.

Figure 4.23 VECM Estimates

Vector Error Correction Estimates

Vector Error Correction Estimates
Date: 04/17/23 Time: 01:16
Sample (adjusted): 4/19/2018 4/06/2023
Included observations: 1105 after adjustments
Standard errors in ( ) & t-statistics in [ ]

Cointegrating Eq:	CointEq1	CointEq2	CointEq3	CointEq4
NIFTY_DLOG(-1)	1.000000	0.000000	0.000000	0.000000
NEKKEI_DLOG(-1)	0.000000	1.000000	0.000000	0.000000
SP500_DLOG(-1)	0.000000	0.000000	1.000000	0.000000
GOLD_DLOG(-1)	0.000000	0.000000	0.000000	1.000000
OIL_DLOG(-1)	-0.464095	-0.346547	-0.319936	0.015291
	(0.03939)	(0.03046)	(0.03165)	(0.01980)
	[-11.7818]	[-11.3781]	[-10.1101]	[0.77244]
C	-0.000275	-0.000100	-0.000267	-0.000389

Error Correction:	D(NIFTY_DL	D(NEKKEI_D	D(SP500_DL	D(GOLD_DL	D(OIL_DLOG)
CointEq1	-0.574305	0.418398	0.414701	-0.096618	1.212099
	(0.08433)	(0.08595)	(0.08041)	(0.07066)	(0.21597)
	[-6.81031]	[4.86766]	[5.15707]	[-1.36741]	[5.61226]
CointEq2	0.230247	-1.165499	0.304091	-0.184165	0.785516
	(0.11585)	(0.11809)	(0.11048)	(0.09707)	(0.29671)
	[1.98739]	[-9.86978]	[2.75256]	[-1.89719]	[2.64740]
CointEq3	0.118086	0.204602	-1.230292	0.179962	-0.483128
	(0.13614)	(0.13876)	(0.12982)	(0.11407)	(0.34866)
	[0.86741]	[1.47449]	[-9.47714]	[1.57770]	[-1.38568]
CointEq4	0.077260	-0.233187	0.136034	-1.258356	-0.437872
	(0.10733)	(0.10940)	(0.10234)	(0.08993)	(0.27487)
	[0.71986]	[-2.13160]	[1.32918]	[-13.9931]	[-1.59300]
D(NIFTY_DLOG(-1))	-0.402585	-0.374760	-0.363333	0.051415	-1.207423
	(0.07950)	(0.08103)	(0.07581)	(0.06661)	(0.20361)
	[-5.06387]	[-4.62471]	[-4.79263]	[0.77185]	[-5.93007]
D(NIFTY_DLOG(-2))	-0.306309	-0.330412	-0.297948	0.064935	-0.978765
	(0.07385)	(0.07527)	(0.07042)	(0.06188)	(0.18913)
	[-4.14787]	[-4.38962]	[-4.23106]	[1.04945]	[-5.17511]
D(NIFTY_DLOG(-3))	-0.306415	-0.304609	-0.199474	-0.040069	-1.250526
	(0.06752)	(0.06883)	(0.06439)	(0.05658)	(0.17294)
	[-4.53784]	[-4.42576]	[-3.09791]	[-0.70821]	[-7.23115]
D(NIFTY_DLOG(-4))	-0.159326	-0.181834	-0.083642	-0.054281	-0.778439
	(0.05986)	(0.06102)	(0.05709)	(0.05016)	(0.15332)
	[-2.66145]	[-2.97998]	[-1.46521]	[-1.08217]	[-5.07730]
D(NIFTY_DLOG(-5))	-0.118583	-0.121392	-0.078956	-0.033026	-0.487797
	(0.05031)	(0.05128)	(0.04798)	(0.04216)	(0.12885)
	[-2.35694]	[-2.36713]	[-1.64573]	[-0.78342]	[-3.78567]
D(NIFTY_DLOG(-6))	-0.083752	-0.091412	-0.037755	-0.003392	-0.281950
	(0.03613)	(0.03682)	(0.03445)	(0.03027)	(0.09253)
	[-2.31820]	[-2.48237]	[-1.09592]	[-0.11204]	[-3.04722]

Vector Error Correction Estimates

D(NEKKEI_DLOG(-1))	-0.118002 (0.10554) [-1.11806]	0.077105 (0.10758) [0.71674]	-0.202806 (0.10064) [-2.01512]	0.204327 (0.08843) [2.31056]	-0.714138 (0.27030) [-2.64200]
D(NEKKEI_DLOG(-2))	0.004461 (0.09566) [0.04664]	0.169755 (0.09750) [1.74109]	-0.089553 (0.09121) [-0.98178]	0.137866 (0.08015) [1.72015]	-0.625010 (0.24498) [-2.55126]
D(NEKKEI_DLOG(-3))	-0.037009 (0.08534) [-0.43366]	0.175435 (0.08699) [2.01679]	-0.119569 (0.08138) [-1.46927]	0.168672 (0.07151) [2.35883]	-0.499251 (0.21857) [-2.28419]
D(NEKKEI_DLOG(-4))	-0.052108 (0.07348) [-0.70911]	0.196995 (0.07490) [2.63007]	-0.097674 (0.07007) [-1.39389]	0.117226 (0.06157) [1.90390]	-0.456037 (0.18820) [-2.42316]
D(NEKKEI_DLOG(-5))	0.014648 (0.05938) [0.24667]	0.199237 (0.06053) [3.29172]	-0.034650 (0.05663) [-0.61191]	0.041687 (0.04976) [0.83785]	-0.502160 (0.15208) [-3.30190]
D(NEKKEI_DLOG(-6))	0.004745 (0.03906) [0.12149]	0.083626 (0.03981) [2.10047]	-0.015640 (0.03725) [-0.41991]	-0.007806 (0.03273) [-0.23851]	-0.162942 (0.10004) [-1.62883]
D(SP500_DLOG(-1))	-0.088996 (0.12389) [-0.71834]	-0.048725 (0.12628) [-0.38585]	0.193545 (0.11814) [1.63828]	-0.153772 (0.10381) [-1.48134]	0.720957 (0.31729) [2.27220]
D(SP500_DLOG(-2))	-0.147507 (0.11105) [-1.32829]	-0.046971 (0.11319) [-0.41497]	0.130664 (0.10589) [1.23390]	-0.088795 (0.09305) [-0.95430]	0.750677 (0.28441) [2.63943]
D(SP500_DLOG(-3))	-0.056924 (0.09685) [-0.58774]	-0.040032 (0.09872) [-0.40552]	0.123152 (0.09236) [1.33344]	-0.045061 (0.08115) [-0.55527]	0.784057 (0.24805) [3.16092]
D(SP500_DLOG(-4))	-0.171237 (0.08198) [-2.08865]	-0.140436 (0.08357) [-1.68055]	-0.031665 (0.07818) [-0.40504]	0.007363 (0.06869) [0.10719]	0.525293 (0.20997) [2.50176]
D(SP500_DLOG(-5))	-0.119571 (0.06476) [-1.84648]	-0.132725 (0.06600) [-2.01083]	-0.014359 (0.06175) [-0.23254]	0.002309 (0.05426) [0.04255]	0.352080 (0.16585) [2.12293]
D(SP500_DLOG(-6))	-0.058867 (0.04386) [-1.34217]	-0.037371 (0.04471) [-0.83594]	-0.004837 (0.04182) [-0.11565]	0.030096 (0.03675) [0.81894]	0.250412 (0.11233) [2.22928]
D(GOLD_DLOG(-1))	0.051999 (0.09748) [0.53343]	0.244666 (0.09936) [2.46240]	-0.039572 (0.09296) [-0.42570]	0.234629 (0.08168) [2.87259]	0.492435 (0.24966) [1.97243]
D(GOLD_DLOG(-2))	0.092651 (0.08727) [1.06165]	0.274380 (0.08895) [3.08453]	0.042579 (0.08322) [0.51165]	0.191444 (0.07312) [2.61812]	0.320608 (0.22351) [1.43443]
D(GOLD_DLOG(-3))	0.037974 (0.07696) [0.49341]	0.116554 (0.07845) [1.48576]	-0.028283 (0.07339) [-0.38538]	0.197986 (0.06449) [3.07019]	0.301655 (0.19711) [1.53039]
D(GOLD_DLOG(-4))	0.056544 (0.06630) [0.85283]	0.156112 (0.06758) [2.31003]	0.017308 (0.06322) [0.27375]	0.157874 (0.05555) [2.84184]	0.130900 (0.16980) [0.77089]

Vector Error Correction Estimates

D(GOLD_DLOG(-5))	0.055705	0.126154	0.077043	0.077953	-0.026881
	(0.05314)	(0.05417)	(0.05067)	(0.04453)	(0.13610)
	[1.04824]	[2.32903]	[1.52034]	[1.75071]	[-0.19751]
D(GOLD_DLOG(-6))	0.076389	0.135558	0.103051	0.005537	-0.170950
	(0.03681)	(0.03752)	(0.03510)	(0.03084)	(0.09427)
	[2.07540]	[3.61330]	[2.93610]	[0.17953]	[-1.81350]
D(OIL_DLOG(-1))	-0.116505	-0.078288	-0.031308	-0.001696	-0.385916
	(0.02713)	(0.02765)	(0.02587)	(0.02273)	(0.06948)
	[-4.29468]	[-2.83129]	[-1.21027]	[-0.07463]	[-5.55460]
D(OIL_DLOG(-2))	-0.081358	-0.033820	-0.006517	-0.007468	-0.310863
	(0.02560)	(0.02609)	(0.02441)	(0.02145)	(0.06555)
	[-3.17860]	[-1.29631]	[-0.26703]	[-0.34823]	[-4.74220]
D(OIL_DLOG(-3))	-0.055080	-0.014946	0.014136	-0.009233	-0.301063
	(0.02286)	(0.02330)	(0.02180)	(0.01916)	(0.05855)
	[-2.40920]	[-0.64139]	[0.64841]	[-0.48198]	[-5.14173]
D(OIL_DLOG(-4))	-0.029132	-0.027290	0.012829	0.014176	-0.232991
	(0.01931)	(0.01968)	(0.01841)	(0.01618)	(0.04945)
	[-1.50893]	[-1.38677]	[0.69685]	[0.87633]	[-4.71203]
D(OIL_DLOG(-5))	-0.010488	-0.017826	0.003972	-0.006710	-0.223737
	(0.01529)	(0.01558)	(0.01458)	(0.01281)	(0.03916)
	[-0.68593]	[-1.14380]	[0.27242]	[-0.52374]	[-5.71366]
D(OIL_DLOG(-6))	-0.025894	-0.030311	-0.024555	0.005936	-0.063994
	(0.01092)	(0.01113)	(0.01042)	(0.00915)	(0.02798)
	[-2.37054]	[-2.72241]	[-2.35743]	[0.64859]	[-2.28751]
C	3.24E-05	-4.19E-05	-1.81E-05	-3.97E-05	-0.000520
	(0.00037)	(0.00038)	(0.00035)	(0.00031)	(0.00095)
	[0.08751]	[-0.11102]	[-0.05132]	[-0.12798]	[-0.54882]
R-squared	0.512921	0.541668	0.509523	0.542851	0.594244
Adj. R-squared	0.497444	0.527104	0.493937	0.528325	0.581351
Sum sq. resids	0.161928	0.168233	0.147243	0.113682	1.062116
S.E. equation	0.012302	0.012539	0.011731	0.010308	0.031506
F-statistic	33.14028	37.19271	32.69259	37.37037	46.08975
Log likelihood	3309.654	3288.552	3362.179	3505.100	2270.477
Akaike AIC	-5.926976	-5.888782	-6.022043	-6.280724	-4.046112
Schwarz SC	-5.768364	-5.730170	-5.863432	-6.122112	-3.887500
Mean dependent	9.40E-05	2.77E-05	2.15E-05	-3.94E-05	-0.000303
S.D. dependent	0.017353	0.018234	0.016490	0.015008	0.048693

Determinant resid covariance (dof adj.)	1.42E-19
Determinant resid covariance	1.21E-19
Log likelihood	16226.70
Akaike information criterion	-29.01665
Schwarz criterion	-28.13296
Number of coefficients	195

Source Self Analysis

One of the first things we observe is how the value of R^2 is improved over VAR.

Figure 4.24 VECM Nifty 50 Coefficient

	A	B	C	D	E	F	G	H	I
1	co	coval	lo	to	P	equn	What	Lag	Represents
2	C(1)	-0.57273	0.08413	-6.80769	0	Nifty	Coint Erro	-1	Coint Error(-1)impacts Nifty equation.
3	C(2)	0.23219	0.115624	2.008144	0.0447	Nifty	E	-1	E(-1)impacts Nifty equation.
4	C(3)	0.119571	0.135989	0.879267	0.3793	Nifty	E	-1	E(-1)impacts Nifty equation.
5	C(4)	0.079033	0.107119	0.737808	0.4607	Nifty	E	-1	E(-1)impacts Nifty equation.
6	C(5)	-0.40373	0.079377	-5.08621	0	Nifty	Nifty	-1	Nifty(-1)impacts Nifty equation.
7	C(6)	-0.30737	0.073732	-4.16871	0	Nifty	Nifty	-2	Nifty(-2)impacts Nifty equation.
8	C(7)	-0.30779	0.06734	-4.5707	0	Nifty	Nifty	-3	Nifty(-3)impacts Nifty equation.
9	C(8)	-0.1613	0.059477	-2.71196	0.0067	Nifty	Nifty	-4	Nifty(-4)impacts Nifty equation.
10	C(9)	-0.11966	0.050163	-2.38542	0.0171	Nifty	Nifty	-5	Nifty(-5)impacts Nifty equation.
11	C(10)	-0.08538	0.035705	-2.39118	0.0168	Nifty	Nifty	-6	Nifty(-6)impacts Nifty equation.
12	C(11)	-0.11922	0.105419	-1.1309	0.02581	Nifty	Nikkei	-1	Nikkei(-1)impacts Nifty equation.
13	C(12)	0.00335	0.095543	0.035064	0.04461	Nifty	Nikkei	-2	Nikkei(-2)impacts Nifty equation.
14	C(13)	-0.03747	0.085292	-0.43937	0.6604	Nifty	Nikkei	-3	Nikkei(-3)impacts Nifty equation.
15	C(14)	-0.05288	0.073409	-0.72029	0.4714	Nifty	Nikkei	-4	Nikkei(-4)impacts Nifty equation.
16	C(15)	0.013206	0.059163	0.22321	0.8234	Nifty	Nikkei	-5	Nikkei(-5)impacts Nifty equation.
17	C(16)	0.004384	0.039025	0.112342	0.9106	Nifty	Nikkei	-6	Nikkei(-6)impacts Nifty equation.
18	C(17)	-0.09087	0.123681	-0.73471	0.4625	Nifty	SP500	-1	SP500(-1)impacts Nifty equation.
19	C(18)	-0.14953	0.110799	-1.34955	0.1772	Nifty	SP500	-2	SP500(-2)impacts Nifty equation.
20	C(19)	-0.05882	0.096605	-0.6089	0.05426	Nifty	SP500	-3	SP500(-3)impacts Nifty equation.
21	C(20)	-0.17244	0.081851	-2.10679	0.0352	Nifty	SP500	-4	SP500(-4)impacts Nifty equation.
22	C(21)	-0.11987	0.064721	-1.85215	0.0641	Nifty	SP500	-5	SP500(-5)impacts Nifty equation.
23	C(22)	-0.06018	0.043623	-1.3796	0.1678	Nifty	SP500	-6	SP500(-6)impacts Nifty equation.
24	C(23)	0.049692	0.097138	0.511567	0.0309	Nifty	Gold	-1	Gold(-1)impacts Nifty equation.
25	C(24)	0.090336	0.086893	1.039619	0.2986	Nifty	Gold	-2	Gold(-2)impacts Nifty equation.
26	C(25)	0.035742	0.076571	0.466774	0.6407	Nifty	Gold	-3	Gold(-3)impacts Nifty equation.
27	C(26)	0.054875	0.066041	0.830929	0.4061	Nifty	Gold	-4	Gold(-4)impacts Nifty equation.
28	C(27)	0.054781	0.05303	1.033018	0.3016	Nifty	Gold	-5	Gold(-5)impacts Nifty equation.
29	C(28)	0.075479	0.036666	2.058533	0.0396	Nifty	Gold	-6	Gold(-6)impacts Nifty equation.
30	C(29)	-0.11496	0.026623	-4.31802	0	Nifty	OIL	-1	OIL(-1)impacts Nifty equation.
31	C(30)	-0.07979	0.025044	-3.18582	0.0015	Nifty	OIL	-2	OIL(-2)impacts Nifty equation.
32	C(31)	-0.05397	0.022552	-2.39309	0.0167	Nifty	OIL	-3	OIL(-3)impacts Nifty equation.
33	C(32)	-0.02818	0.019035	-1.48027	0.1389	Nifty	OIL	-4	OIL(-4)impacts Nifty equation.
34	C(33)	-0.00996	0.015183	-0.6562	0.5117	Nifty	OIL	-5	OIL(-5)impacts Nifty equation.
35	C(34)	-0.02535	0.010767	-2.35431	0.0186	Nifty	OIL	-6	OIL(-6)impacts Nifty equation.
36	C(35)	3.65E-05	0.00037	0.098682	0.9214	Nifty	C	c	C(c)impacts Nifty equation.

Source Self Analysis

EViews churn 35 coefficients for each independent variable, it becomes essential to choose only the statistically significant out of the bunch in order to make the equations not only precise but also to reduce over fitting, the excels given below show the statistical significance and t-values of each coefficient. Following were chosen:

For **NIFTY** equation:

C (2) for the error coefficient: .23219; It represents the VECM error

C (6) for the nifty coefficient: -.30737 It represents the NIFTY at lag 2

C (12) for the Nikkei coefficient: 0.00335 It represents the Nikkei at lag 2

C (19) for the S&P coefficient: -0.058822 It represents the S&P500 at lag 3

C (23) for the Gold coefficient: 0.049692 It represents the Gold at lag 1

C (31) for the OIL coefficient: -0.05397 It represents the OIL at lag 3

Figure 4.25 VECM Nikkei Coefficient

37	C(36)	0.420682	0.085755	4.905614	0	Nikkei	E		-1	E(-1)impacts Nikkei equation.
38	C(37)	-1.16269	0.117859	-9.86509	0	Nikkei	E		-1	E(-1)impacts Nikkei equation.
39	C(38)	0.20675	0.138617	1.491524	0.1359	Nikkei	E		-1	E(-1)impacts Nikkei equation.
40	C(39)	-0.23062	0.109189	-2.11213	0.0347	Nikkei	E		-1	E(-1)impacts Nikkei equation.
41	C(40)	-0.37641	0.080911	-4.65221	0	Nikkei	Nifty		-1	Nifty(-1)impacts Nikkei equation.
42	C(41)	-0.33194	0.075157	-4.41668	0	Nikkei	Nifty		-2	Nifty(-2)impacts Nikkei equation.
43	C(42)	-0.3066	0.068641	-4.46673	0	Nikkei	Nifty		-3	Nifty(-3)impacts Nikkei equation.
44	C(43)	-0.18469	0.060627	-3.0464	0.0023	Nikkei	Nifty		-4	Nifty(-4)impacts Nikkei equation.
45	C(44)	-0.12295	0.051132	-2.40457	0.0162	Nikkei	Nifty		-5	Nifty(-5)impacts Nikkei equation.
46	C(45)	-0.09377	0.036395	-2.57634	0.01	Nikkei	Nifty		-6	Nifty(-6)impacts Nikkei equation.
47	C(46)	0.075342	0.107456	0.701144	0.4832	Nikkei	Nikkei		-1	Nikkei(-1)impacts Nikkei equation.
48	C(47)	0.168146	0.097389	1.726533	0.0843	Nikkei	Nikkei		-2	Nikkei(-2)impacts Nikkei equation.
49	C(48)	0.174761	0.08694	2.010138	0.0445	Nikkei	Nikkei		-3	Nikkei(-3)impacts Nikkei equation.
50	C(49)	0.195884	0.074827	2.617812	0.0089	Nikkei	Nikkei		-4	Nikkei(-4)impacts Nikkei equation.
51	C(50)	0.19715	0.060306	3.269164	0.0011	Nikkei	Nikkei		-5	Nikkei(-5)impacts Nikkei equation.
52	C(51)	0.083103	0.039779	2.089124	0.0367	Nikkei	Nikkei		-6	Nikkei(-6)impacts Nikkei equation.
53	C(52)	-0.05144	0.126071	-0.40801	0.6833	Nikkei	SP500		-1	SP500(-1)impacts Nikkei equation.
54	C(53)	-0.0499	0.11294	-0.44181	0.6586	Nikkei	SP500		-2	SP500(-2)impacts Nikkei equation.
55	C(54)	-0.04278	0.098472	-0.43446	0.664	Nikkei	SP500		-3	SP500(-3)impacts Nikkei equation.
56	C(55)	-0.14218	0.083433	-1.70415	0.0884	Nikkei	SP500		-4	SP500(-4)impacts Nikkei equation.
57	C(56)	-0.13316	0.065972	-2.01846	0.0436	Nikkei	SP500		-5	SP500(-5)impacts Nikkei equation.
58	C(57)	-0.03927	0.044466	-0.88325	0.3771	Nikkei	SP500		-6	SP500(-6)impacts Nikkei equation.
59	C(58)	0.241326	0.099015	2.437275	0.0148	Nikkei	Gold		-1	Gold(-1)impacts Nikkei equation.
60	C(59)	0.271027	0.088572	3.059953	0.0022	Nikkei	Gold		-2	Gold(-2)impacts Nikkei equation.
61	C(60)	0.113321	0.078051	1.451877	0.1466	Nikkei	Gold		-3	Gold(-3)impacts Nikkei equation.
62	C(61)	0.153696	0.067317	2.283163	0.0225	Nikkei	Gold		-4	Gold(-4)impacts Nikkei equation.
63	C(62)	0.124816	0.054055	2.309076	0.021	Nikkei	Gold		-5	Gold(-5)impacts Nikkei equation.
64	C(63)	0.134241	0.037375	3.591727	0.0003	Nikkei	Gold		-6	Gold(-6)impacts Nikkei equation.
65	C(64)	-0.07605	0.027137	-2.80227	0.0051	Nikkei	OIL		-1	OIL(-1)impacts Nikkei equation.
66	C(65)	-0.03154	0.025528	-1.23557	0.2167	Nikkei	OIL		-2	OIL(-2)impacts Nikkei equation.
67	C(66)	-0.01334	0.022988	-0.58023	0.5618	Nikkei	OIL		-3	OIL(-3)impacts Nikkei equation.
68	C(67)	-0.02591	0.019403	-1.33521	0.1819	Nikkei	OIL		-4	OIL(-4)impacts Nikkei equation.
69	C(68)	-0.01707	0.015477	-1.10271	0.2702	Nikkei	OIL		-5	OIL(-5)impacts Nikkei equation.
70	C(69)	-0.02952	0.010975	-2.68987	0.0072	Nikkei	OIL		-6	OIL(-6)impacts Nikkei equation.
71	C(70)	-3.60E-05	0.000377	-0.0954	0.924	Nikkei	C	c		C(c)impacts Nikkei equation.

Source Self Analysis

For **Nikkei** equation:

C(36) for the error coefficient: .4206; It represents the VECM error

C(41) for the nifty coefficient: -.033 It represents the NIFTY at lag 2

C(49) for the Nikkei coefficient: .1958 It represents the Nikkei at lag 4

C(56) for the S&P coefficient: -.13316 It represents the S&P500 at lag 5

C(62) for the Gold coefficient: .1248 It represents the Gold at lag 5

C(64) for the OIL coefficient: -. -0.076045 It represents the OIL at lag 1

Figure 4.26 VECM S&P 500 Coefficient

72	C(71)	0.419899	0.080261	5.231663	0	S&P500	E		-1	E(-1)impacts S&P500 equation.
73	C(72)	0.310495	0.110308	2.814808	0.0049	S&P500	E		-1	E(-1)impacts S&P500 equation.
74	C(73)	-1.2254	0.129736	-9.44534	0	S&P500	E		-1	E(-1)impacts S&P500 equation.
75	C(74)	0.141875	0.102193	1.388304	0.1651	S&P500	E		-1	E(-1)impacts S&P500 equation.
76	C(75)	-0.3671	0.075727	-4.84763	0	S&P500	Nifty		-1	Nifty(-1)impacts S&P500 equation.
77	C(76)	-0.30144	0.070342	-4.28531	0	S&P500	Nifty		-2	Nifty(-2)impacts S&P500 equation.
78	C(77)	-0.20401	0.064244	-3.17558	0.0015	S&P500	Nifty		-3	Nifty(-3)impacts S&P500 equation.
79	C(78)	-0.09015	0.056742	-1.58874	0.1122	S&P500	Nifty		-4	Nifty(-4)impacts S&P500 equation.
80	C(79)	-0.08251	0.047856	-1.72404	0.0848	S&P500	Nifty		-5	Nifty(-5)impacts S&P500 equation.
81	C(80)	-0.04311	0.034064	-1.26569	0.2057	S&P500	Nifty		-6	Nifty(-6)impacts S&P500 equation.
82	C(81)	-0.20682	0.100572	-2.05642	0.0398	S&P500	Nikkei		-1	Nikkei(-1)impacts S&P500 equation.
83	C(82)	-0.09322	0.09115	-1.02266	0.3065	S&P500	Nikkei		-2	Nikkei(-2)impacts S&P500 equation.
84	C(83)	-0.1211	0.08137	-1.4883	0.1367	S&P500	Nikkei		-3	Nikkei(-3)impacts S&P500 equation.
85	C(84)	-0.1002	0.070033	-1.43079	0.1525	S&P500	Nikkei		-4	Nikkei(-4)impacts S&P500 equation.
86	C(85)	-0.0394	0.056442	-0.69808	0.4852	S&P500	Nikkei		-5	Nikkei(-5)impacts S&P500 equation.
87	C(86)	-0.01683	0.037231	-0.45206	0.6512	S&P500	Nikkei		-6	Nikkei(-6)impacts S&P500 equation.
88	C(87)	0.187371	0.117994	1.587973	0.1124	S&P500	SP500		-1	SP500(-1)impacts S&P500 equation.
89	C(88)	0.124002	0.105704	1.173105	0.02408	S&P500	SP500		-2	SP500(-2)impacts S&P500 equation.
90	C(89)	0.116894	0.092163	1.268344	0.2047	S&P500	SP500		-3	SP500(-3)impacts S&P500 equation.
91	C(90)	-0.03564	0.078088	-0.45643	0.6481	S&P500	SP500		-4	SP500(-4)impacts S&P500 equation.
92	C(91)	-0.01535	0.061745	-0.24866	0.8036	S&P500	SP500		-5	SP500(-5)impacts S&P500 equation.
93	C(92)	-0.00917	0.041617	-0.2203	0.8256	S&P500	SP500		-6	SP500(-6)impacts S&P500 equation.
94	C(93)	-0.04717	0.092671	-0.50904	0.6107	S&P500	Gold		-1	Gold(-1)impacts S&P500 equation.
95	C(94)	0.03495	0.082898	0.421603	0.6733	S&P500	Gold		-2	Gold(-2)impacts S&P500 equation.
96	C(95)	-0.03564	0.073051	-0.48789	0.6256	S&P500	Gold		-3	Gold(-3)impacts S&P500 equation.
97	C(96)	0.011808	0.063004	0.18742	0.8513	S&P500	Gold		-4	Gold(-4)impacts S&P500 equation.
98	C(97)	0.073997	0.050591	1.462643	0.1436	S&P500	Gold		-5	Gold(-5)impacts S&P500 equation.
99	C(98)	0.100054	0.03498	2.860276	0.0042	S&P500	Gold		-6	Gold(-6)impacts S&P500 equation.
100	C(99)	-0.0262	0.025398	-1.03172	0.3022	S&P500	OIL		-1	OIL(-1)impacts S&P500 equation.
101	C(100)	-0.00133	0.023892	-0.05575	0.9555	S&P500	OIL		-2	OIL(-2)impacts S&P500 equation.
102	C(101)	0.017796	0.021515	0.827133	0.4082	S&P500	OIL		-3	OIL(-3)impacts S&P500 equation.
103	C(102)	0.015977	0.01816	0.879782	0.379	S&P500	OIL		-4	OIL(-4)impacts S&P500 equation.
104	C(103)	0.0057	0.014485	0.39348	0.0394	S&P500	OIL		-5	OIL(-5)impacts S&P500 equation.
105	C(104)	-0.02276	0.010272	-2.21566	0.0268	S&P500	OIL		-6	OIL(-6)impacts S&P500 equation.
106	C(105)	-4.62E-06	0.000353	-0.01309	0.9896	S&P500	C	c		C(c)impacts S&P500 equation.

Source Self Analysis

For **S&P 500** equation:

C(71) for the error coefficient: 0.41899; It represents the VECM error

C(75) for the nifty coefficient:-0.301 It represents the NIFTY at lag 2

C(81) for the Nikkei coefficient: -0.206 It represents the Nikkei at lag 1

C(88) for the S&P coefficient: 0.12 It represents the S&P500 at lag 2

C(98) for the Gold coefficient: 0.100 It represents the Gold at lag 6

C(104) for the OIL coefficient: -0.022 It represents the OIL at lag 6

Now the equations:

$$\boldsymbol{Nifty_t = -.30737 * (Nifty_{t-2}) + .004461 * Nikkei_{t-2} + .0966 * S\&P500_{t-3} + .097 * Gold_{t-1} - .053 * Oil_{t-3} + .2329 * \varepsilon_{t-1}}$$

$$\boldsymbol{Nikkei_t = -.33 * (Nifty_{t-2}) + .1958 * Nikkei_{t-4} - .1336 * S\&P500_{t-5} + .1248 * Gold_{t-5} - .7695 * Oil_{t-1} + .4206 * \varepsilon_{t-1}}$$

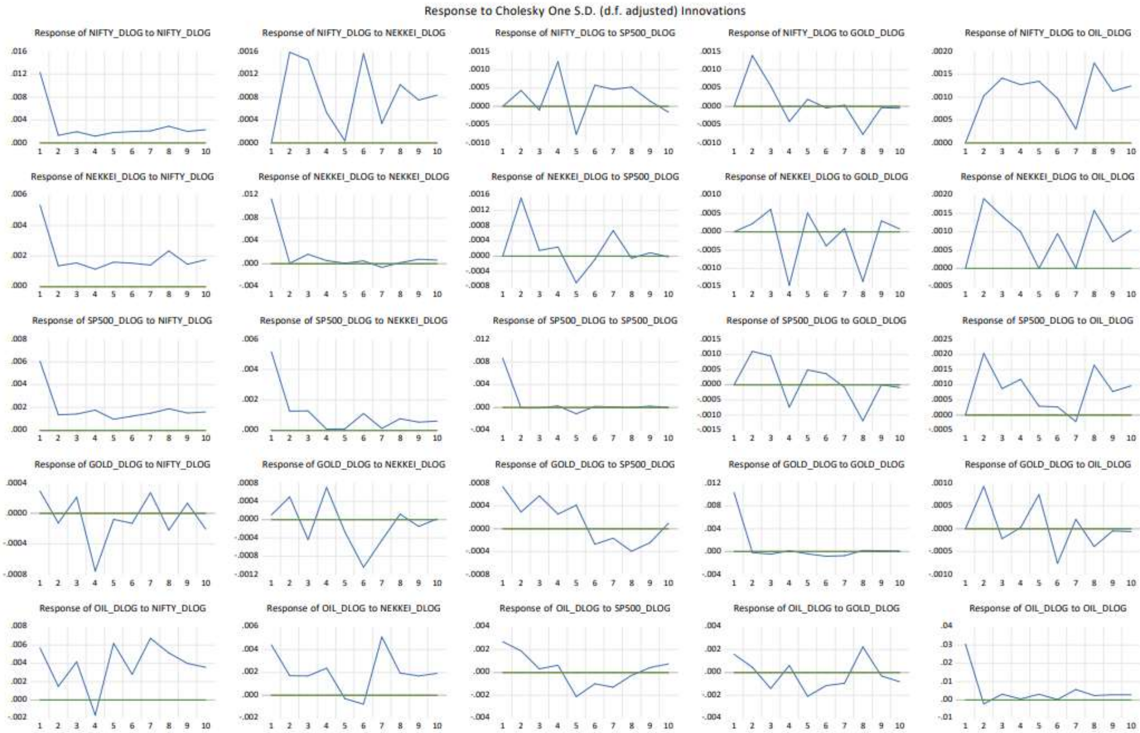
$$\boldsymbol{S\&P500_t = -.301 * (Nifty_{t-2}) + .206 * Nikkei_{t-1} + .12 * S\&P500_{t-2} + .100 * Gold_{t-6} - .022 * Oil_{t-6} + .4189 * \varepsilon_{t-1}}$$

Observations:

1. While the model is more robust and has better curve fitting than VAR, it too isn't good for forecasting utilizing the data.
2. Moreover, the residuals are correlated which leads to curve not being the best fit.
3. The borderline root may make the system unstable in extreme values which are frequent in market shocks
4. The model has a slow reaction or an extreme reaction to the changes as seen in impulse plot.
5. In equation the significance of coefficients is low and thus they are not the best predictors.

While VAR had an R^2 of 9%, here the R^2 has increased to 51% this along with improvements in AIC, Swartz and Darbin Watson, which shows a relatively stable system as represented by the systems, impulse response and unit root representation.

Figure 4.27 Impulse Response of all the variables together in VECM



Source Self Analysis

Here, the impulse response shows that:

1. For Nifty, Nifty at time zero forms the biggest affecting factor, whereas other factors such as Nikkei, S&P500, gold, oil affect at higher lags. (Nikkei-2,3,6; S&P 500-2,4,5,6,7; gold-2;oil-2,3,4.)
2. For Nikkei: Nifty and Nikkei impacts at start while others at later lags
3. For S&P 500 Nifty Nikkei and S&P500 impacts at the initial stage.

These help us select the desired coefficient for the equation at appropriate lag.

Figure 4.28 VECM Normality Test

VEC Residual Normality Tests
Orthogonalization: Cholesky (Lutkepohl)
Null Hypothesis: Residuals are multivariate normal
Date: 04/17/23 Time: 01:44
Sample: 4/10/2018 4/06/2023
Included observations: 1105

Component	Skewness	Chi-sq	df	Prob.*
1	0.217420	8.705806	1	0.0032
2	0.152844	4.302366	1	0.0381
3	-0.143379	3.785987	1	0.0517
4	-0.362944	24.25995	1	0.0000
5	-0.011142	0.022863	1	0.8798
Joint		41.07697	5	0.0000

Component	Kurtosis	Chi-sq	df	Prob.
1	6.380209	526.0634	1	0.0000
2	6.446539	546.9119	1	0.0000
3	4.072764	52.98579	1	0.0000
4	6.489402	560.6000	1	0.0000
5	11.88823	3637.324	1	0.0000
Joint		5323.885	5	0.0000

Component	Jarque-Bera	df	Prob.
1	534.7692	2	0.0000
2	551.2143	2	0.0000
3	56.77178	2	0.0000
4	584.8599	2	0.0000
5	3637.346	2	0.0000
Joint	5364.962	10	0.0000

*Approximate p-values do not account for coefficient estimation

Source Self Analysis

This test represents the normality of the residuals and zero heteroscedasticity in the system.

This means that the system shall easily be BLUE and the confidence intervals can be clearly defined.

Figure 4.29 VECM Heteroskedasticity

VEC Residual Heteroskedasticity Tests (Levels and Squares)
Date: 04/17/23 Time: 01:44
Sample: 4/10/2018 4/06/2023
Included observations: 1105

Joint test:

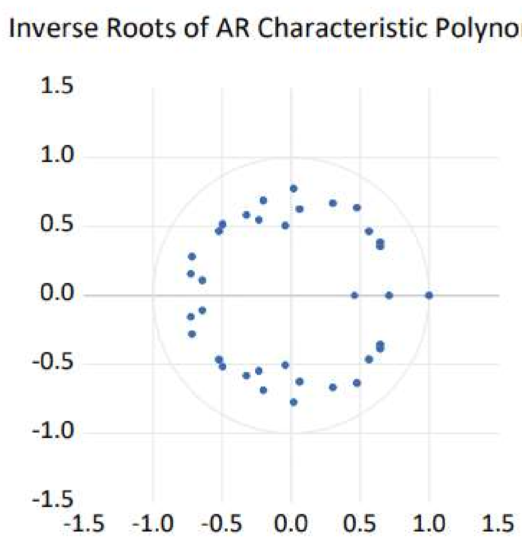
Chi-sq	df	Prob.
4181.434	1020	0.0000

Individual components:

Dependent	R-squared	F(68,1036)	Prob.	Chi-sq(68)	Prob.
res1*res1	0.384460	9.515817	0.0000	424.8285	0.0000
res2*res2	0.431044	11.54236	0.0000	476.3041	0.0000
res3*res3	0.264441	5.477235	0.0000	292.2069	0.0000
res4*res4	0.138900	2.457530	0.0000	153.4843	0.0000
res5*res5	0.472810	13.66374	0.0000	522.4546	0.0000
res2*res1	0.420331	11.04744	0.0000	464.4655	0.0000
res3*res1	0.382239	9.426832	0.0000	422.3743	0.0000
res3*res2	0.257458	5.282469	0.0000	284.4915	0.0000
res4*res1	0.280739	5.946591	0.0000	310.2171	0.0000
res4*res2	0.351406	8.254427	0.0000	388.3035	0.0000
res4*res3	0.208443	4.011952	0.0000	230.3294	0.0000
res5*res1	0.373605	9.086898	0.0000	412.8338	0.0000
res5*res2	0.357511	8.477627	0.0000	395.0495	0.0000
res5*res3	0.340200	7.855484	0.0000	375.9211	0.0000
res5*res4	0.167815	3.072288	0.0000	185.4357	0.0000

Source Self Analysis

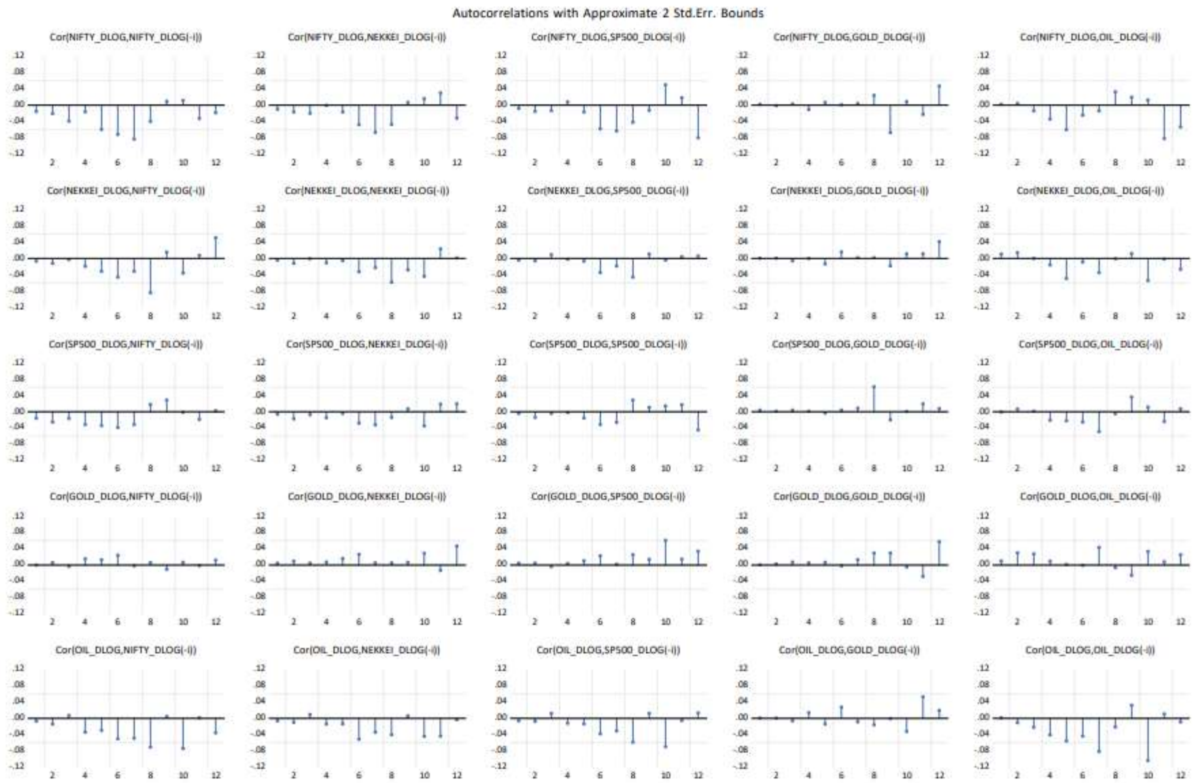
Figure 4.30 VECM Unit Root



Source Self Analysis

Here the stability of the system is represented, VECM forms a very stable system as it can be show with the unit root graph, all roots are within the unit circle

Figure 4.31 Auto Correlation with approximate 2 Standard Deviations



Source Self Analysis

Auto correlation in the residuals is represented, there are certain autocorrelation however it is at higher lags but as seen in impulse at higher lags the impact of the shocks dies down.

It has to be kept in mind that at higher lags the coefficient selected must not interfere with our other variables.

The output below shows the lack of granger causality at a combined level, while there might be a few instances of small granger causality occurring due to one factor, for example nifty has a granger causal relation with OIL but overall, the independent variables don’t granger cause the dependent variable.

This is essential as it makes the VECM model much more reliable.

Figure 4.32 VECM Granger Causality

VEC Granger Causality/Block Exogeneity Wald Tests
Date: 04/17/23 Time: 01:44
Sample: 4/10/2018 4/06/2023
Included observations: 1105

Dependent variable: D(NIFTY_DLOG)

Excluded	Chi-sq	df	Prob.
D(NEKKEI_DLOG)	15.97656	6	0.0139
D(SP500_DLOG)	19.90624	6	0.0029
D(GOLD_DLOG)	8.612327	6	0.1966
D(OIL_DLOG)	32.10264	6	0.0000
All	92.85498	24	0.0000

Dependent variable: D(NEKKEI_DLOG)

Excluded	Chi-sq	df	Prob.
D(NIFTY_DLOG)	26.55848	6	0.0002
D(SP500_DLOG)	10.74739	6	0.0965
D(GOLD_DLOG)	32.52316	6	0.0000
D(OIL_DLOG)	32.00780	6	0.0000
All	137.4008	24	0.0000

Dependent variable: D(SP500_DLOG)

Excluded	Chi-sq	df	Prob.
D(NIFTY_DLOG)	30.02518	6	0.0000
D(NEKKEI_DLOG)	14.59366	6	0.0237
D(GOLD_DLOG)	24.99379	6	0.0003
D(OIL_DLOG)	25.46856	6	0.0003
All	122.3923	24	0.0000

Dependent variable: D(GOLD_DLOG)

Excluded	Chi-sq	df	Prob.
D(NIFTY_DLOG)	14.38915	6	0.0256
D(NEKKEI_DLOG)	15.28287	6	0.0182
D(SP500_DLOG)	7.131542	6	0.3089
D(OIL_DLOG)	15.28943	6	0.0181
All	61.62004	24	0.0000

Dependent variable: D(OIL_DLOG)

Excluded	Chi-sq	df	Prob.
D(NIFTY_DLOG)	73.47851	6	0.0000
D(NEKKEI_DLOG)	15.79288	6	0.0149
D(SP500_DLOG)	12.34238	6	0.0548
D(GOLD_DLOG)	15.32368	6	0.0179
All	164.4340	24	0.0000

Source Self Analysis

4.6 ARMA Model

1. While the previous VECM model was a considerable improvement over VAR model, R value improved from about 9.8% to 51.0% and it led to a much stabler model, however the goodness-of- fit required much more to be desired.
2. It was observed from the model's residuals that we required a model which was much more receptive to the frequent changes of the dependent variable (any index which is being predicted)
3. Thus, we now move to our final and much more responsive model(ARMA) model.
4. Impulse plots and lag length exercises from the previous models give the correct depth of lags.
5. In the ARMA model we utilize 2 AR and 1 MA. i.e 2 lags of Autoregression are utilized and 1 Moving Average is utilized.
6. While the model could have been AR (6) and MA (4) it was not taken in order to prevent overfitting and us leading to non-forecastable model.
7. In order to determine the equation, the equation builder in EViews was utilized and for various outputs, the settings were tweaked in order to get the most optimal equation.
8. Only NIFTY was predicted and while other variables were taken as independent.
9. A small period was left of the complete sample in order to test the forecasted results

Few important points

Period chosen: 4/13/2018 to 4/06/2022

Dependent Variable: NIFTY 50

Independent Variables and their lags:

a. Nikkei 250, lag 2 ie Nikkei(t-2)
b. S&P 500, lag 3 ie S&P500(t-3)
c. Gold, lag 1 ie Gold(t-1), these are the gold spot prices in the market
d. OIL, lag 3 ie OIL(t-1), these are the oil future prices.

Here the prices the lags represent their t minus value, for example a lag of 3 shows prices of the index or commodity 3 trading periods before, so 3 days before.

Now, as discussed AR(2) was used, here AR means autoregressive term which is basically NIFTY(t-2) in order to predict NIFTY(t).

This infact is intuitive as we saw from the impulse plots that nifty is highly affected by nifty itself.

MA parts represents the moving average of dependent variable.

Now the, results:

Figure 4.33 Nifty 50 ARMA coefficients

Dependent Variable: NIFTY_Dlog
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 04/17/23 Time: 01:47
Sample: 4/13/2018 4/06/2022
Included observations: 891
Convergence achieved after 39 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
NEKKEI_DLOG(-2)	753.8904	225.1292	3.348701	0.0008
SP500_DLOG(-3)	837.5047	276.7839	3.025843	0.0026
GOLD_DLOG(-1)	556.5248	295.2673	1.884817	0.0598
OIL_DLOG(-3)	59.67291	49.63233	1.202299	0.2296
C	13694.09	2721.395	5.032010	0.0000
AR(2)	0.997742	0.003235	308.4196	0.0000
MA(1)	0.999051	0.003317	301.2326	0.0000
SIGMASQ	23908.01	628.1346	38.06193	0.0000

R-squared	0.846496	Mean dependent var	12798.72
Adjusted R-squared	0.996469	S.D. dependent var	2613.708
S.E. of regression	155.3210	Akaike info criterion	12.94588
Sum squared resid	21302040	Schwarz criterion	12.98891
Log likelihood	-5759.389	Hannan-Quinn criter.	12.96233
F-statistic	35877.44	Durbin-Watson stat	1.956354
Prob(F-statistic)	0.000000		

Inverted AR Roots	1.00	-1.00
Inverted MA Roots	-1.00	

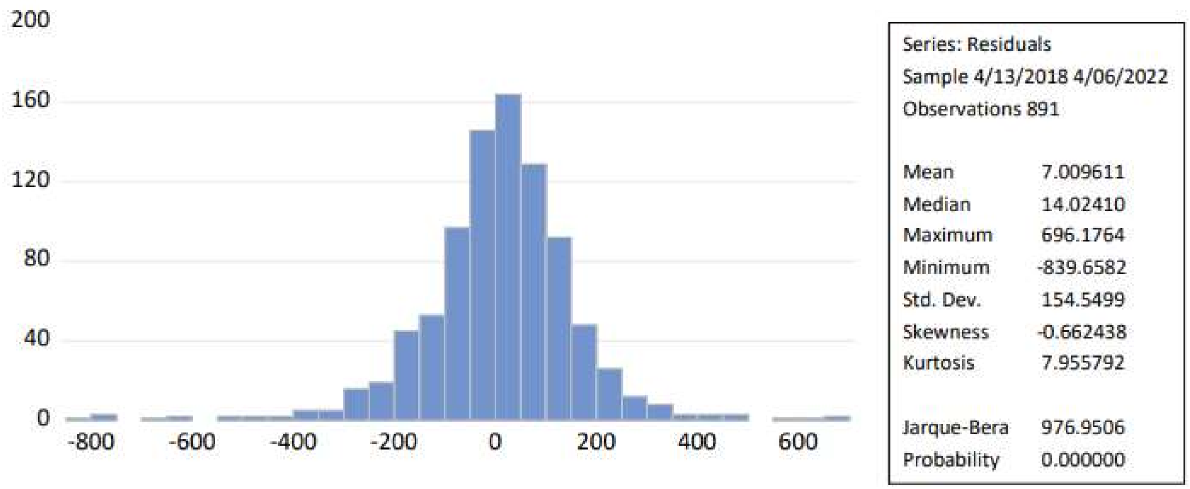
Source Self Analysis

Observations:

- The following results were derived using the EVIEWs equation builder and adding the variables with correct lag in it
- It can be clearly observed that R^2 has considerabley improved from VECM from 51% to 84% representing a better fit model
- Moreover, Durbin Watson being higher than R^2 shows that the model doesn't suffer from the case of Spurious regression
- High AIC and BIC show that model well fitted.

* Moreover, significant and high F-Statistic show that jointly the independent variables were easily able to explain the dependent NIFTY and thus we get a good equation to predict NIFTY.

Figure 4.34Residual Normality Test



Source Self Analysis

Furthermore, the above test show:

1. All assumption of CLRM is followed
2. The residuals are normally distributed.
3. There is no heteroscedasticity in the data as shown by Breusch-Pegan-Godfrey test.
4. Moreover, residuals are not partially or completely autocorrelated
5. Thus, the model seems to be BLUE.

Figure 4.35 ARMA Test for Heteroskedasticity

Heteroskedasticity Test: Breusch-Pagan-Godfrey
Null hypothesis: Homoskedasticity

F-statistic	8.553298	Prob. F(4,886)	0.0000
Obs*R-squared	33.12706	Prob. Chi-Square(4)	0.0000
Scaled explained SS	110.8736	Prob. Chi-Square(4)	0.0000

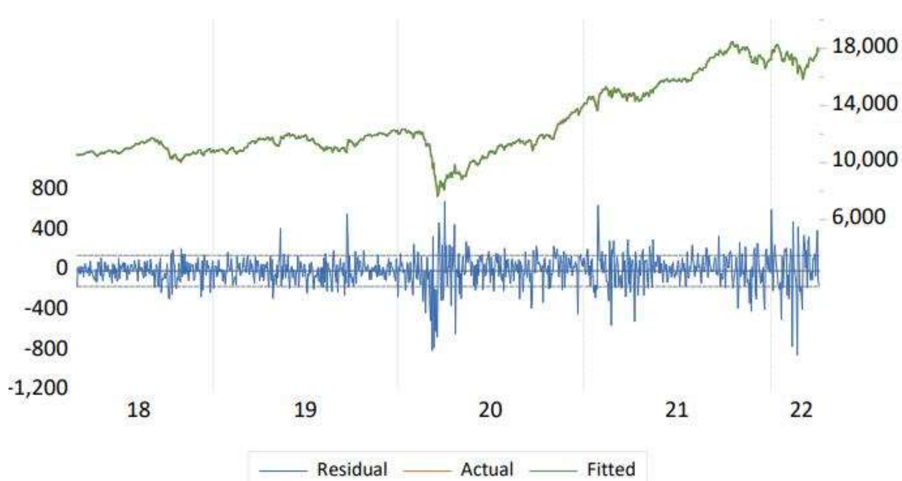
Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 04/17/23 Time: 02:05
Sample: 4/13/2018 4/06/2022
Included observations: 891

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	24573.79	2062.503	11.91454	0.0000
NEKKEI_DLOG(-2)	-646416.9	158178.2	-4.086637	0.0000
SP500_DLOG(-3)	-601081.3	180725.5	-3.325936	0.0009
GOLD_DLOG(-1)	-47360.56	193717.2	-0.244483	0.8069
OIL_DLOG(-3)	-58379.05	52298.33	-1.116270	0.2646

R-squared	0.037180	Mean dependent var	23908.01
Adjusted R-squared	0.032833	S.D. dependent var	62451.37
S.E. of regression	61417.59	Akaike info criterion	24.89438
Sum squared resid	3.34E+12	Schwarz criterion	24.92127
Log likelihood	-11085.44	Hannan-Quinn criter.	24.90465
F-statistic	8.553298	Durbin-Watson stat	1.742559
Prob(F-statistic)	0.000001		

Source Self Analysis

Figure 4.36 ARMA Residual and Actual Fitting



Source Self Analysis

Figure 4.37 ARMA Correlation of Residuals

Correlogram of Residuals

Date: 04/17/23 Time: 02:05
Sample (adjusted): 4/13/2018 4/06/2022
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	0.018	0.018	0.3053	
		2	-0.008	-0.008	0.3565	
		3	-0.028	-0.028	1.0578	0.304
		4	0.071	0.072	5.6361	0.060
		5	0.032	0.029	6.5504	0.088
		6	0.040	0.039	7.9734	0.093
		7	0.011	0.014	8.0791	0.152
		8	-0.073	-0.077	12.905	0.045
		9	-0.003	-0.002	12.913	0.074
		10	0.053	0.047	15.423	0.051
		11	-0.043	-0.054	17.115	0.047
		12	0.026	0.038	17.713	0.060
		13	0.018	0.023	18.005	0.081
		14	0.037	0.033	19.276	0.082
		15	-0.053	-0.047	21.870	0.057
		16	0.083	0.077	28.133	0.014
		17	0.027	0.022	28.799	0.017
		18	-0.004	-0.007	28.810	0.025
		19	0.010	0.011	28.893	0.036
		20	-0.037	-0.048	30.170	0.036
		21	0.070	0.076	34.664	0.015
		22	0.034	0.026	35.693	0.017
		23	0.009	-0.007	35.763	0.023
		24	-0.059	-0.041	38.972	0.014
		25	-0.054	-0.049	41.684	0.010
		26	0.018	-0.002	41.974	0.013
		27	-0.037	-0.041	43.265	0.013
		28	0.031	0.028	44.154	0.015
		29	-0.030	-0.014	44.961	0.016
		30	-0.022	-0.017	45.394	0.020
		31	0.017	0.027	45.657	0.025
		32	0.001	-0.009	45.658	0.033
		33	0.020	0.013	46.029	0.040
		34	-0.014	-0.006	46.215	0.050
		35	0.033	0.015	47.222	0.052
		36	0.014	0.028	47.401	0.063

*Probabilities may not be valid for this equation specification.

Source Self Analysis

ARMA Model

Thus, the ARMA model is

$$\boldsymbol{Nifty_t = 0.99 * (Nifty_{t-1} + Nifty_{t-2}) + 753.89 * Nikkei_{t-2} + 873.5 * S\&P500_{t-3} + 556 * Gold_{t-1} + 59.67 * Oil_{t-3} + .99 * \varepsilon_{t-1} + 1364}$$

4.7 Forecasting, Testing and Final Observations

Now utilizing the equation so generated, forecasting for the period 04/1/2023 to 06/04/2023 was done in order to check the veracity of the equation.

In order to do forecasting we utilize the set of values as left during the equation building exercise.

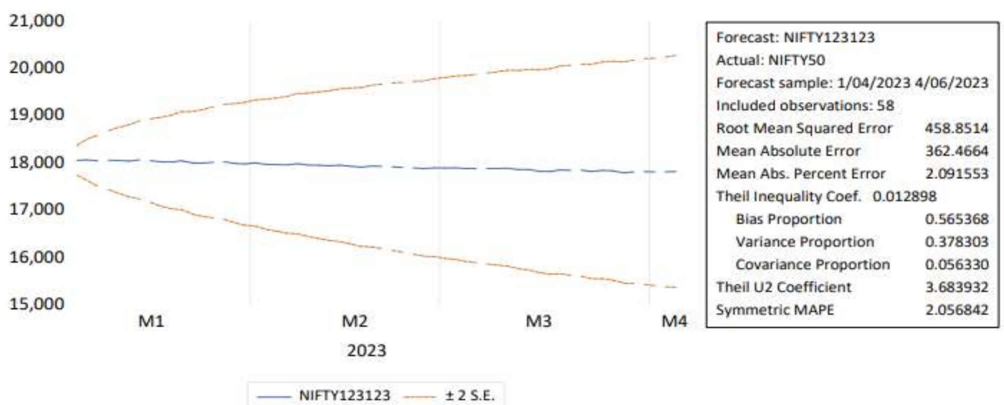
This is done via the forecast menu of Eviews and selecting the time period and specifying the parameters.

The output so received is compared to the actual real values.

Following were the results: -

1. Our forecast traces, follows the Nifty closes and in short run is in equilibrium.
2. Thus, the forecast clearly traces Nifty
3. Then the forecasted values were compared to the actual of the forecast, the results are included below.
4. It can be observed the error observed are very small and the mean error is -.19, thus showing a highly accurate model which can be utilized everywhere

Figure 4.38 Forecasted value with 2 Standard Deviation



Source Self Analysis

Figure 4.39 ARMA Chow Forecast Test

Chow Forecast Test
Equation: ARMANEW
Test predictions for observations from 1/04/2022 to 4/06/2022
Specification: NIFTY50 NEKKEI_DLOG(-2) SP500_DLOG(-3) GOLD_DLOG(-1) OIL_DLOG(-3) AR(2) MA(1) C

	Value	df	Probability
F-statistic	3.384395	(58, 825)	0.0000
Likelihood ratio	197.3741	58	0.0000

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	4094295.	58	70591.29
Restricted SSR	21302040	883	24124.62
Unrestricted SSR	17207745	825	20857.87

LR test summary:

	Value
Restricted LogL	-5759.389
Unrestricted LogL	-5660.702

Unrestricted log likelihood adjusts test equation results to account for observations in forecast sample

Unrestricted Test Equation:
Dependent Variable: NIFTY50
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 04/17/23 Time: 02:07
Sample: 4/13/2018 12/30/2021
Included observations: 833
Failure to improve objective (non-zero gradients) after 34 iterations
Coefficient covariance computed using outer product of gradients

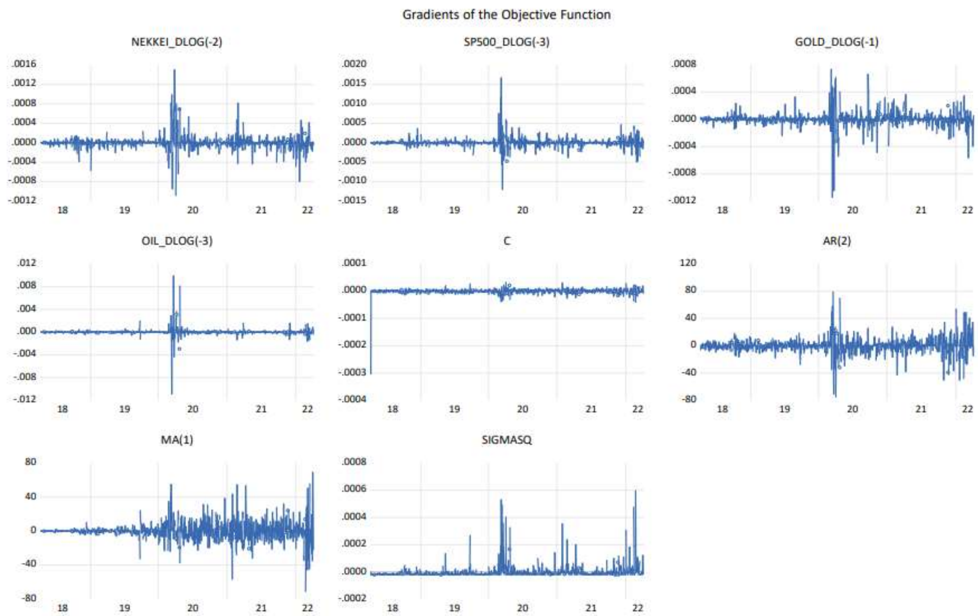
Variable	Coefficient	Std. Error	t-Statistic	Prob.
NEKKEI_DLOG(-2)	923.9905	213.2677	4.332539	0.0000
SP500_DLOG(-3)	824.3135	260.3936	3.165644	0.0016
GOLD_DLOG(-1)	603.6894	271.7443	2.221534	0.0266
OIL_DLOG(-3)	73.95952	44.72868	1.653514	0.0986
C	13343.51	2258.013	5.909405	0.0000
AR(2)	0.997291	0.003468	287.5808	0.0000
MA(1)	1.000000	0.745756	1.340921	0.1803
SIGMASQ	20657.56	657.1464	31.43524	0.0000

R-squared	0.996422	Mean dependent var	12485.93
Adjusted R-squared	0.996392	S.D. dependent var	2404.330
S.E. of regression	144.4226	Akaike info criterion	12.80187
Sum squared resid	17207745	Schwarz criterion	12.84725
Log likelihood	-5323.980	Hannan-Quinn criter.	12.81927
F-statistic	32823.64	Durbin-Watson stat	1.938550
Prob(F-statistic)	0.000000		

Inverted AR Roots	1.00	-1.00
Inverted MA Roots	-1.00	

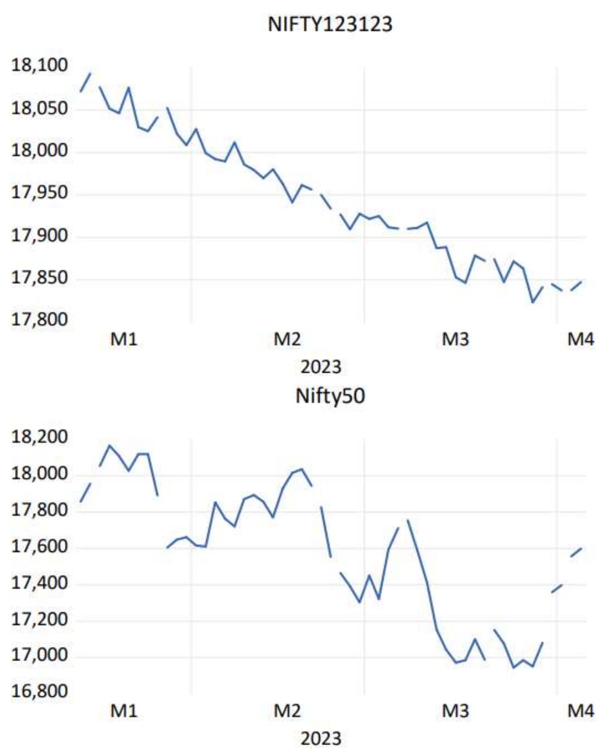
Source Self Analysis

Figure 4.40 Gradient of Objective Function



Source Self Analysis

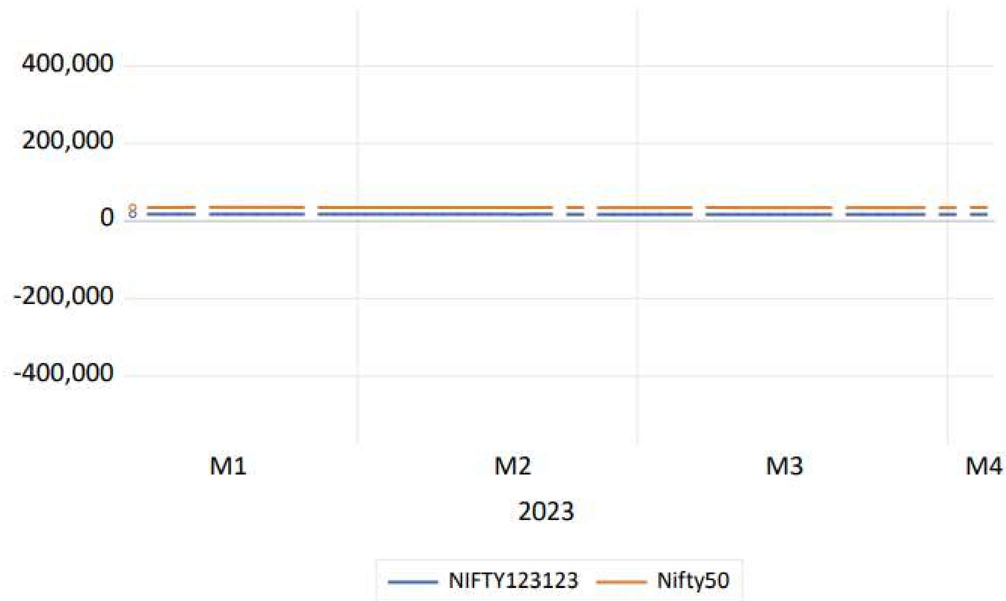
Figure 4.41 Nifty 50 and Forecasted value



Source Self Analysis

As seen in Figure 40 the Nifty Forecasted by the name of Nifty 123123 is following similar trend as the actual Nifty 50 thought the trough formed in Nifty are bigger than forecasted but the crests are also larger which leads to the net average which is similar to the model forecasted

Figure 4.42 Nifty 50 and Forecasted Value 2



Source Self Analysis

Figure 4.43 Forecasted and Actual Value with Average Error Percentage

	A	B	C	D
1	Date	Actual	Forecast	Residual Error Percentage
2	04-01-2023	18042.95	18083.33	-0.002237993
3	05-01-2023	17992.15	18092.93	-0.005601332
4	06-01-2023	17859.45	18073.88	-0.012006529
5	10-01-2023	17914.15	18087.29	-0.009664986
6	11-01-2023	17895.7	18080.97	-0.010352766
7	12-01-2023	17858.2	18071.94	-0.011968731
8	13-01-2023	17956.6	18092.92	-0.007591638
9	17-01-2023	18053.3	18076.96	-0.001310564
10	18-01-2023	18165.35	18051.63	0.00626027
11	19-01-2023	18107.85	18046.33	0.003397422
12	20-01-2023	18027.65	18076.29	-0.002698078
13	23-01-2023	18118.55	18029.73	0.004902158
14	24-01-2023	18118.3	18025.2	0.005138451
15	25-01-2023	17891.95	18041.52	-0.008359625
16	27-01-2023	17604.35	18052.62	-0.025463593
17	30-01-2023	17648.95	18022.29	-0.021153666
18	31-01-2023	17662.15	18008.8	-0.019626716
19	01-02-2023	17616.3	18027.55	-0.023344857
20	02-02-2023	17610.4	17999.29	-0.022082974
21	03-02-2023	17854.05	17992.15	-0.00773494
22	06-02-2023	17764.6	17989.44	-0.012656632
23	07-02-2023	17721.5	18011.92	-0.016388003
24	08-02-2023	17871.7	17985.92	-0.00639111
25	09-02-2023	17893.45	17979.41	-0.004803993
26	10-02-2023	17856.5	17969.8	-0.006345028
27	13-02-2023	17770.9	17980.1	-0.011772054
28	14-02-2023	17929.85	17963.31	-0.001866162
29	15-02-2023	18015.85	17941.24	0.004141353
30	16-02-2023	18035.85	17961.52	0.004121236
31	17-02-2023	17944.2	17956.33	-0.000675984
32	21-02-2023	17826.7	17950.14	-0.006924445
33	22-02-2023	17554.3	17933.83	-0.021620344
34	24-02-2023	17465.8	17926.89	-0.026399592
35	27-02-2023	17392.7	17909.63	-0.02972109
36	28-02-2023	17303.95	17927.71	-0.036047261
37	01-03-2023	17450.9	17921.69	-0.026977978
38	02-03-2023	17321.9	17925.12	-0.034824124

39	03-03-2023	17594.35	17911.84	-0.018044997
40	06-03-2023	17711.45	17910.18	-0.011220425
41	08-03-2023	17754.4	17910.1	-0.008769657
42	09-03-2023	17589.6	17911.08	-0.018276709
43	10-03-2023	17412.9	17917.29	-0.028966456
44	13-03-2023	17154.3	17887.39	-0.042735058
45	14-03-2023	17043.3	17888.35	-0.049582534
46	15-03-2023	16972.15	17852.92	-0.051895016
47	16-03-2023	16985.6	17846.42	-0.050679399
48	17-03-2023	17100.05	17878.64	-0.045531446
49	20-03-2023	16988.4	17872.18	-0.052022557
50	22-03-2023	17151.9	17874.21	-0.042112536
51	23-03-2023	17076.9	17847.37	-0.045117674
52	24-03-2023	16945.05	17871.59	-0.054679095
53	27-03-2023	16985.7	17863.27	-0.051665224
54	28-03-2023	16951.7	17823.44	-0.051424931
55	29-03-2023	17080.7	17841.14	-0.044520424
56	31-03-2023	17359.75	17844.8	-0.027941071
57	03-04-2023	17398.05	17837.53	-0.025260302
58	05-04-2023	17557.05	17837.56	-0.015977058
59	06-04-2023	17599.15	17847.35	-0.014102954
60			Average	-0.019951335

Source Self Analysis

This negative Average Error Percentage represent that the model is predicting value slightly higher than the actual value and hence the negative sign.

As shown by the test conducted the error of the ARMA model build is very less and within acceptable ranges The figure 43 represents the value that market can fluctuate to assuming that the fluctuations remain between 2 σ standard deviations.

4.8 Limitation/ Further Scope

- **Slight over fitting**: - Since the data the data used and the equation formed is focused on getting better R square value.
- **Correlation: -** Since the Indexes are highly corelated as the impact on one index can be observed on other indexes to some extend hence the correlation cannot be removed entirely.
- **Better Volatility Models: -** Can use better volatility models like arch and GARCH.
- **Machine Learning and Artificial Intelligence: -** ML AI can be used wherein we can utilise our ARMA equation so as get more sensitive and give better results. Moreover, these ML/AI would be much more receptive to change

- **Program Limitations: -** E Views is an iterative method without searching for close formed solution while it is the standard procedure even in python A close formed solution provide better results
- **Other Limitations: -** Stochastic model was not utilised which are much more flexible in the short run

Chapter-5 Conclusion

When there is a market volatility and uncertainty the investor moves towards investing in safer options which is tangible asset which in our case is gold. The negative coefficient of gold is the indicator of such shifts. Similarly, when the markets on a bull run then the investors are enticed by the higher return invest greater amount in the stock market which is represented in the index with its rise. But in our result, this is not the case as shown by the coefficients of our VAR model which for Nifty 50 had positive coefficients for gold and oil which was also seen in S&P 500 Equation as well but not in Nikkei 225 index which showed a negative coefficient for gold and positive coefficient for oil. This report tries to establish that there is relationship between indexes gold and oil prices and how the fluctuations in one result to fluctuations in another. There are limited literature proving this is the case. This paper uses daily prices of gold and oil along with daily values indexes in order to determine the relationship through VAR model.

This research report uses 2 models one VAR which is further improved by VECM model and ARIMA model to show their relationship. The first model uses Nifty 50 as the dependent variable and using Nikkei 225, S&P 500, gold and oil as independent variables using the first order. The selection of lag is done on the basis of test conducted which comes out at the lag of 6. The VAR estimates were calculated on the basis of the fact that for the coefficient to be accepted first it should be significant i.e., the probability should be less than 0.05 or 5% and if there are multiple terms then the terms with highest t value are chosen for incorporating the maximum effect of that variable on the basis of which we created the VAR equation the test indicated that the equation created had a R square value which was acceptable but when we had to test for assumptions the model failed in the cointegration test or Granger test. But this process was not futile as it did establish the effect of these variables. In order to correct the equations, the project employed VECM method which introduced another error term which led to the correction of model and the condition of no cointegration being satisfied. The second model used was the ARMA model that also utilised the same framework where all the variables were considered an endogenous variable and were of the first order. The R Square value was also acceptable in this case but the value was lower in this test as compared to the one obtained in the VECM model. The ARMA model was further tested by forecasting the data of the later months as the data for this test was till 3 January and the forecast was for 4 January 2023 to 6 April 2023. The average error in the returns were -0.019 or 1.9% the negative sign indicates that the model predicted value slightly lower than the actual value.