# Integrated Machine Learning Framework for Portfolio Optimization and Risk Assessment: A Case Study of Nifty 50 Companies

**Nagavarshini R N**

*MA Financial Economics Madras School of Economics*

# Abstract

This study aims to construct optimized portfolios with the objective of achieving better returns while effectively managing overall portfolio risk. To achieve this goal, sophisticated machine learning models, such as Extreme Gradient Boosting (XGBoost), Agglomerative Hierarchical Clustering, and Generalized Autoregressive Conditional Heteroskedasticity (GARCH), have been used. Additionally, widely recognized optimization techniques such as Mean-Variance Optimization and quantitative methods like Hierarchical Risk Parity Asset Allocation have been utilized to formulate five distinct portfolios.

The portfolios are constructed using yahoo finance data of Nifty 50 companies from the period spanning 2014 to 2023. The primary objective of the study is to analyse the risk and returns of these portfolios to identify the most effective portfolio strategy. The research aims to provide valuable insights and enhance the understanding of the application of machine learning and optimization techniques for investors and portfolio managers.

# Key words

Nifty 50, Portfolio optimization, XGBoost, Mean variance optimization, Agglomerative hierarchical clustering, Sharpe ratio, Volatility ,GARCH,HRP

# Integrated Machine Learning Framework for Portfolio Optimization and Risk Assessment: A Case Study of Nifty 50 Companies

**Introduction**

The finance industry has experienced tremendous growth in recent decades, driven by advanced quantitative, mathematical, and analytical techniques. However, the process of investing and portfolio building has often been flawed due to the irrational thinking and behaviour of investors. Portfolio optimization, a critical domain in finance, has the potential to yield high profits when approached using sophisticated machine learning and optimization techniques. In this study, five distinct portfolios are constructed using such techniques, focusing on the Nifty 50 companies. For a comprehensive analysis, 10 years (2014-2024) of historical daily stock price data from Yahoo Finance is utilized.

The primary objective for any investor or portfolio manager is to minimize portfolio risk while maximizing returns, and one efficient method to achieve this is through diversification. Portfolios are exposed to systematic and market risks, and diversification can effectively manage market risk. One such approach involves selecting stocks from highly uncorrelated sectors. Thus, the Nifty 50 stocks are split into 11 sectors based on the Global Industry Classification Standard. The first portfolio is built by selecting one stock from each sector and optimizing them using the Mean-Variance Optimization technique.(\*

The second portfolio utilizes the Extreme Gradient Boosting (XGBoost) model to predict stock performance. The top 25 percentile stocks with the highest Sharpe ratio (indicating excess return for the volatility of holding the risky stock) are selected from the Nifty 50 stocks. Mean-Variance Optimization is then applied to optimize this portfolio.

The third portfolio, also constructed using XGBoost, analyses Nifty 50 stocks for their Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) values. The top 25 percentile stocks with the lowest MAPE values are chosen for their accuracy, and Mean- Variance Optimization is implemented.

The fourth portfolio considers all Nifty 50 stocks, irrespective of their performance metrics, and applies Mean-Variance Optimization.

The fifth model is built using the Agglomerative Hierarchical Clustering model, where 50 stocks are clustered based on pairwise similarity, and the Hierarchical Risk Parity Asset Allocation technique is used to weight the stocks in the portfolio.

Finally, GARCH models are developed to analyse stock volatility for the year 2023. Various risk and return metrics such as annualized returns, Sortino ratio, Calmar ratio, Maximum Drawdown, 95% Conditional Value at Risk (cVar), and 99% cVar are studied to determine the most efficient portfolio construction strategies.

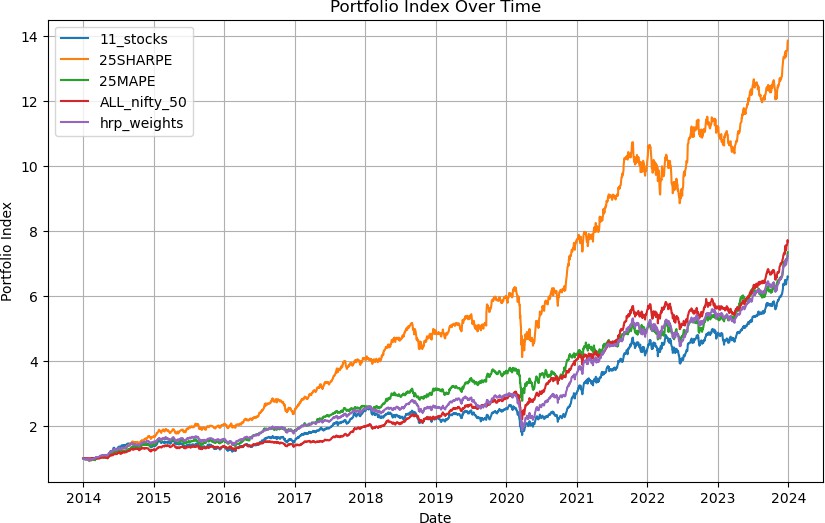


Fig.1. The portfolio indices of the five portfolios for the years 2014 – 2023

# Methodology

The stock market serves as a platform where investors engage in buying and selling ownership shares, offering the potential for returns and providing publicly traded companies with a means to raise capital through the issuance of stocks. The Nifty 50 index represents the performance of 50 large-cap companies across various sectors in the Indian economy. These companies, well-established and listed on the National Stock Exchange of India, span diverse industries including finance, information technology, pharmaceuticals, and consumer goods.

While the stock market can be profitable, it is not without risks, primarily due to market fluctuations influenced by various factors. These factors encompass investor sentiments, dynamic economic indicators, inflation, consumer spending, interest rates, geopolitical events, monetary policies, government regulations, and other market sentiments.

To mitigate the risks associated with market fluctuations, diversification is crucial. In order to do so five different portfolios have been created using machine learning techniques such as XGBoost and hierarchical clustering to select the stocks and mean variance optimization and Hierarchical Risk Parity(HRP) asset allocation to allocate weights to stocks and GARCH model to forecast volatility.

## Extreme Gradient Boosting(XGBoost)

XGBoost is a collection of weak classifier decision trees and it primarily focuses to

train the new decision tree to learn from the errors committed by the previous trees using a gradient descent-like procedure *(May.K,2022).* In this study, XGBoost classifier mosel has been used to select the top 25% of stocks with the highest Sharpe ratio, lowest Mean Absolute Percentage Error (MAPE), and lowest Root Mean Squared Error (RMSE) with accuracy of 73.42,.The model is trained using historical stock data and subsequently evaluated for its accuracy in prediction.

## Mean Variance Optimisation

Harry Markowitz introduced mean-variance optimization in 1952, forming the basis of modern portfolio theory. This theory guides investors in constructing portfolios that maximize expected returns for a given level of volatility or risk.It may be formulated as a convex optimization problem that can be solved analytically. *(Lop´ez de Prado, 2016, 2019).*

The goal here is to find the optimal portfolio that maximizes utility. The process begins by assuming a set of weights that produce the portfolio and calculating the expected return and risk of this portfolio.

Total utility can be calculated by multiplying the probability of discrete returns by their respective utility values.

The optimal portfolio is the one which maximizes utility and at the same time maximize its expected return minus λ times the variance .Markowitz analysed for various levels of the risk aversion coefficient to trace out the efficient portfolio Frontier. *(Uysal.E.,et al,2001)*

## Monte Carlo Method

The Monte Carlo method generates a large number of random results to compare them and obtain the most appropriate solution*(Wang.M, 2022)*. Here, this simulation is used to generate random portfolios, which are then plotted on the mean-variance efficient frontier. Here this stimulation is used to generate random portfolios which are then plotted on the mean variance efficient frontier.

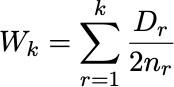
## Agglomerative Hierarchical clustering

Hierarchical clustering is a form of grouping algorithm, where a set of data points is divided into subsets or clusters. Assets within one cluster should be as similar as possible and different from assets within all other clusters. Agglomerative clustering who is the more common one, known as AGNES starts with all assets as single element clusters and then merge clusters until all assets are in the same cluster. The hierarchical clustering algorithm AGNES finds recursively nested clusters, resulting in a tree based representation called dendrogram.*(Eidenvall, A,. 2021)*

Suppose we have data with k clusters C1, C2, ..., Ck. Cr denoting the indices of observations in cluster r and the sum of pairwise distances Dr for cluster Cr, given by,



Where di,j is the Euclidean distance between data points i and j. Based on Dr, the cluster inertia Wk is then calculated as the within cluster , sum of squares around the cluster means



where nr being the number of data points in cluster r. Once the cluster inertia is calculated, the Gap statistic is given by the following equation, where E∗ n denotes expectation from the reference distribution based on n numbers of samples. The optimal numbers of clusters k will be the value that maximizes Gapn(k) *(Eidenvall, A,. 2021) .*

*A black text on a white background  Description automatically generated*

where E∗ n denotes expectation from the reference distribution based on n numbers of samples. The optimal numbers of clusters k will be the value that maximizes Gapn(k).

## Hierarchical Risk Parity Asset Allocation

The HRP approach was first introduced and published in 2016 with the aim of addressing two fundamental problems with modern portfolio theory: Instability and underperformance *(Nanakorn, N.et.al.,2021).*HRP approach only consider the ensuing order of the assets, obtained from the hierarchical clustering dendrogram. An allocation technique is then suggested based on recursive bisection. focuses on the obtained order at the bottom level of the dendrogram and ignores the obtained clusters at different levels of the dendrogram*(Eidenvall, A,. 2021) .*

## Generalized AutoRegressive Conditional Heteroskedasticity (GARCH)

It is a statistical model used to estimate volatility of financial returns by assuming that the variance of the current error term depends on the error term lagged squared and also the past variances. Metrics such as average volatility and standard deviation deviation of volatility of both the portfolio and the sensex and the ratio of portfolio volatility to sensex volatility are computed used this model. These metrices are crucial in understanding the risk associated with the portfolios. GARCH(p,q) used here involves two main components:

The autoregressive component (p) captures the persistence of volatility shocks over time.The moving average component (q) represents the impact of past squared error terms on the current conditional variance.

By applying these principles and diversifying across sectors, investors aim to navigate the inherent risks of the stock market while seeking higher returns.

## 11 STOCKS 11 SECTORS

Investment strategies could be active or passive. One passive way of investing involves investing in different sectors with low correlation, ensuring that the performance and fluctuations in one sector do not significantly impact another. According to the Global Industry Classification Standard (GICS), the stock universe can be categorized into 11

sectors: Information Technology, Healthcare, Consumer Discretionary, Consumer Staples, Industrials, Real Estate, Energy, Utilities, Materials, and Telecommunication Services *(Wang.M, 2022)*. The Nifty 50 companies are therefore categorized into these 11 sectors, with one stock selected from each sector. These 11 stocks are subsequently optimized using Markowitz mean-variance analysis.

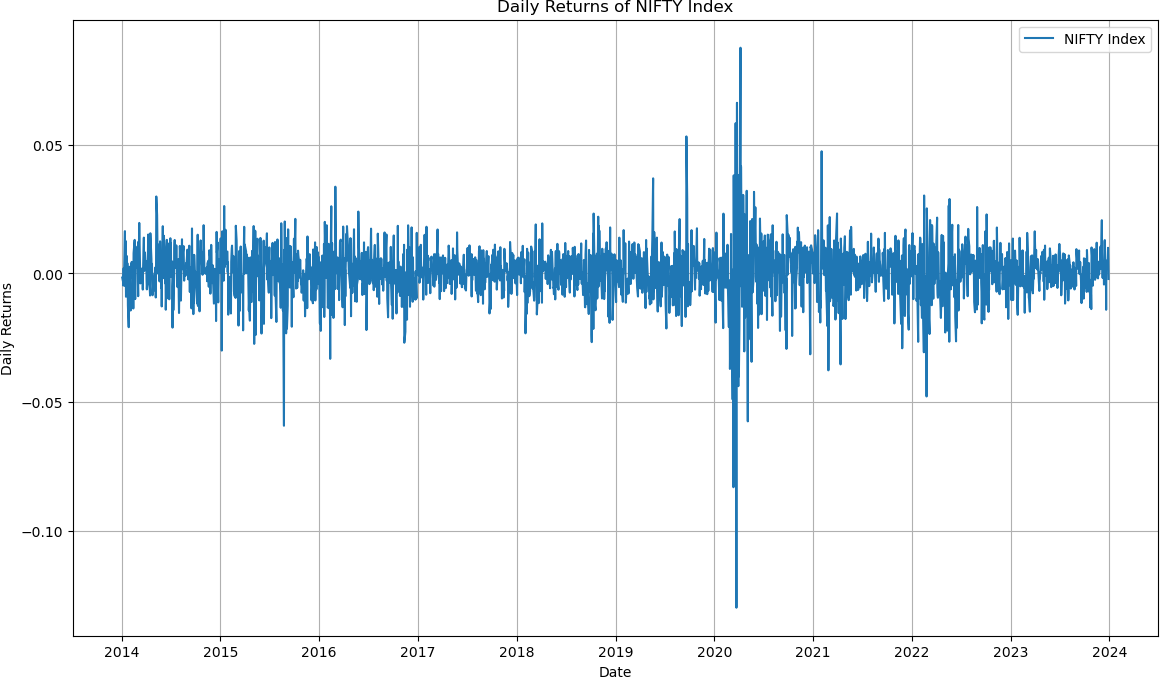


Fig.2. Daily returns of Nifty 50 index (2014-2023)

The stocks of Adani ports ,Bharti Airtel ,DLF ,HCL technologies ,Hindustan Unilever limited, ONGC Power grid ,Sun Pharma ,Tata Steel and Titan are chosen. Historic returns and their historical stock prices are given in figure 3 & 4.

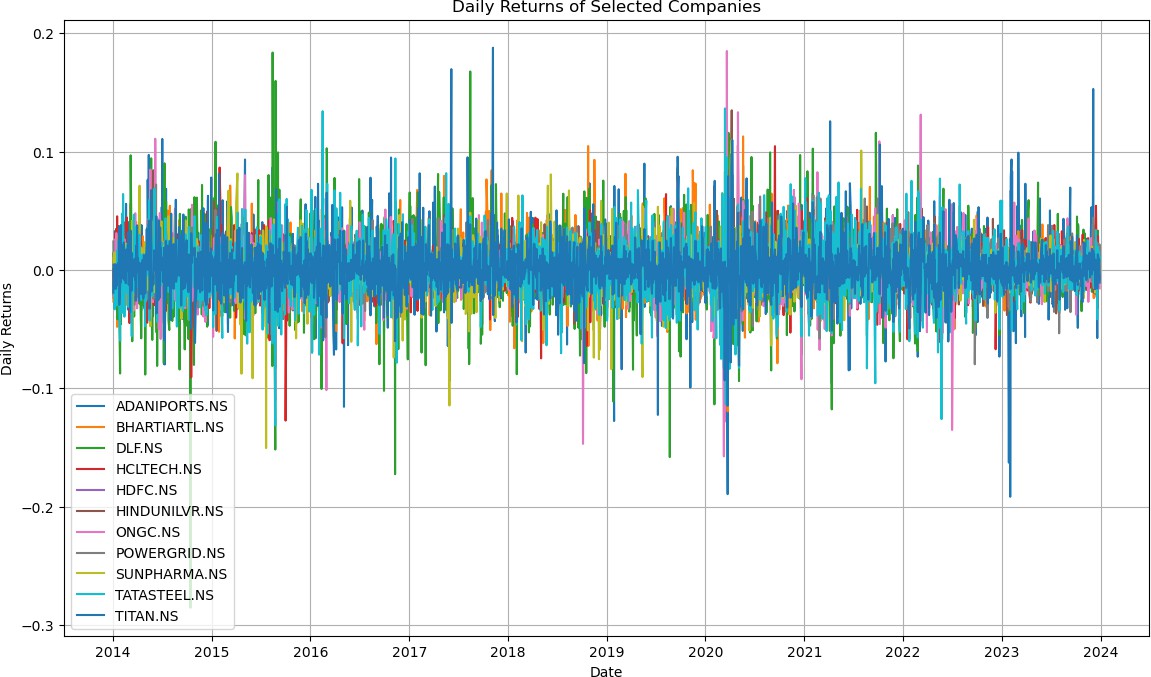


Fig. 3 .Returns of selected companies from 11 sectors(2014-2023)

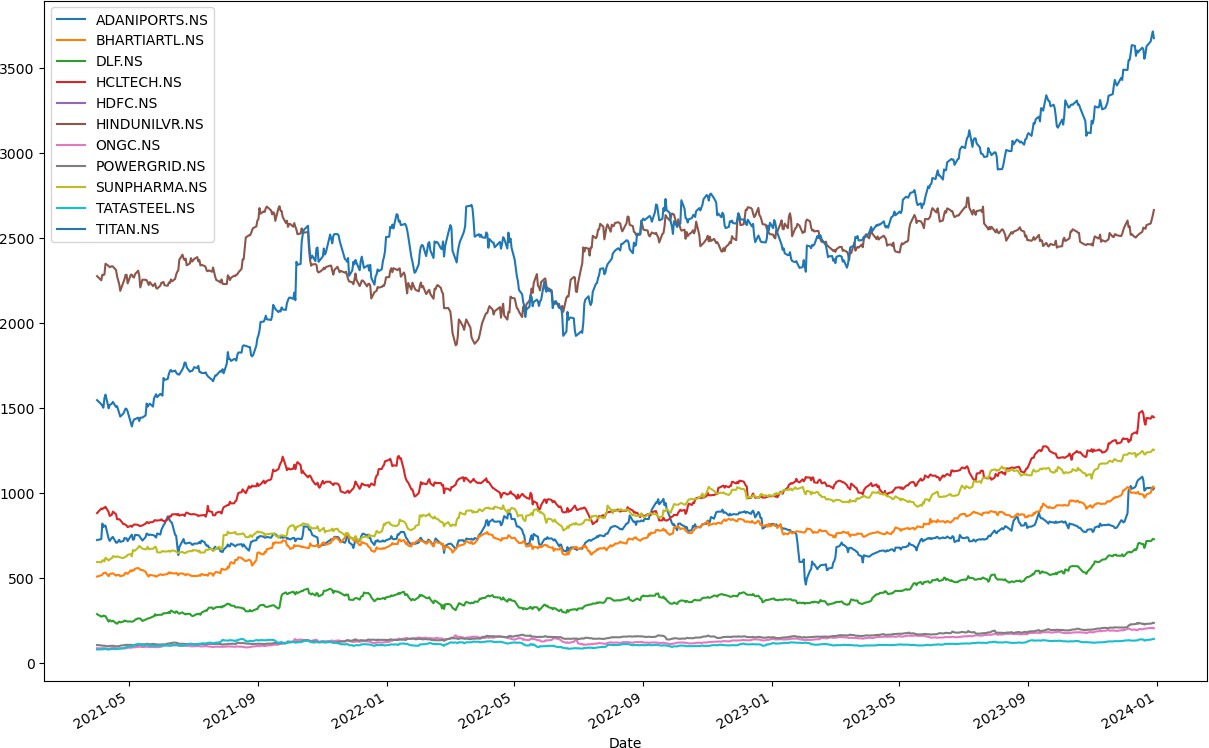


Fig.4. Historical stock prices of stocks in portfolio (2014-2023)

Mean variance optimization is implemented in steps. Firstly Ledoit-Wolf shrinkage is applied to estimate the most stable covariance matrix and then the estimated returns are calculate using the Capital Asset Pricing Model(CAPM) and the optimal portfolio allocation with maximized sharpe ratio is indicated, Monte Carlo stimulation is done by using a Dirichlet distribution and efficient Frontier with maximized sharp ratio is obtained.

The return ,volatility and Sharpe ratio of the portfolio are studied at the minimum volatility and maximum sharpe ratio point

|  |  |  |
| --- | --- | --- |
|  | At minimum volatility point | At maximum Sharpe point |
| Volatility | 0.1508 | 0.1747 |
| Returns | 16.55% | 20.87% |
| Sharpe ratio | 1.0976 | 1.1946 |

Though there is a considerable difference in the volatility between the two portfolios there is a significantly higher returns and higher sharper ratio associated with the portfolio with maximum Sharpe point .Therefore this portfolio is been selected the cummulative returns of this portfolio for the year 2023 is obtained which is 30.79%

Heat map is used to represent the covariance matrix which in turn represents the covariance between assets in a portfolio .High covariances are represented by strong colours while low covariances are present at by weaker colour .It measures the degree to which returns of two assets move in relation to one another and helps in assessing the risk associated with different asset pairings

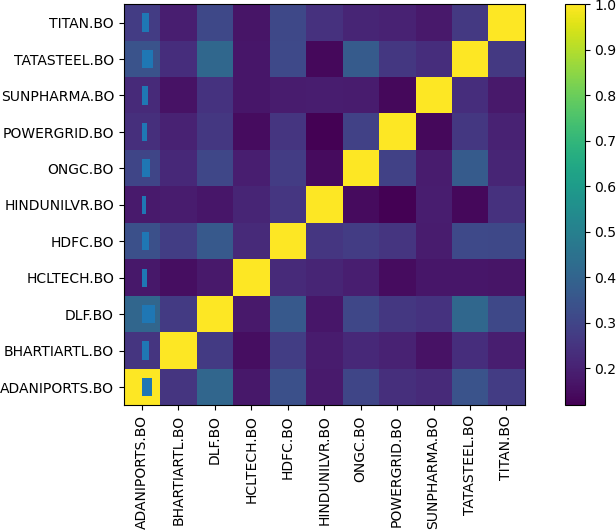


Fig.5. Heat map of the stocks in the portfolio

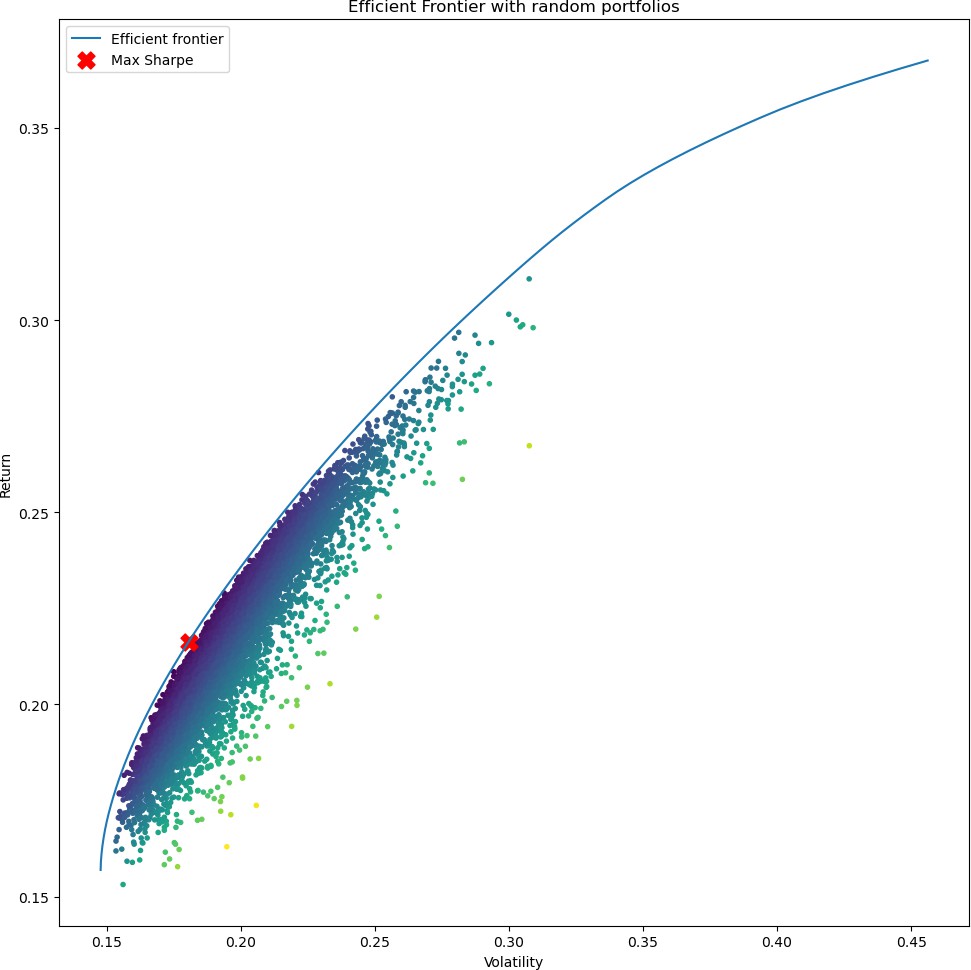


Fig.6.Efficient frontier of the portfolio

|  |  |
| --- | --- |
| Stocks | Weights |
| ADANI PORTS | 0.0843 |
| BHARTI AIRTEL | 0.0691 |
| DLF | 0.0745 |
| HCL TECH | 0.0767 |
| HDFC | 0.0675 |
| HINDUNILVR | 0.1554 |
| ONGC | 0.0913 |
| POWERGRID | 0.1251 |
| SUN PHARMA | 0.0679 |
| TATA STEEL | 0.1039 |
| TITAN | 0.0838 |

Table 2 Weights at maximum Sharpe ratio(1.194)

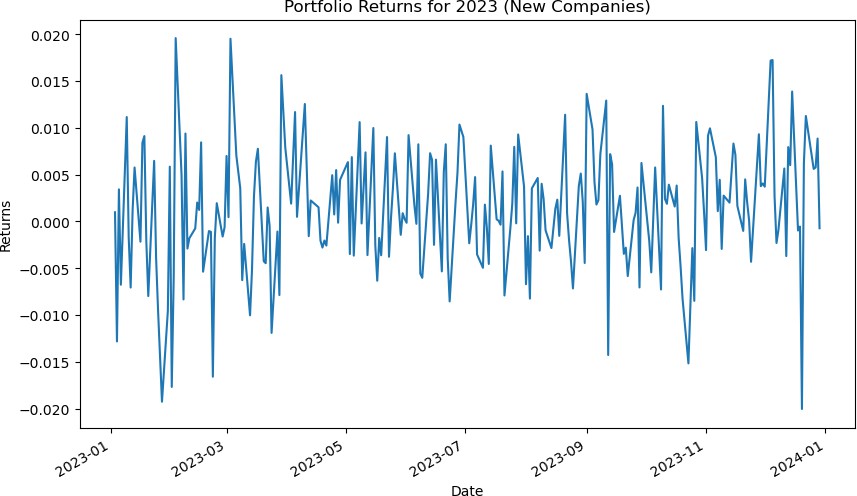


Fig.7. Portfolio (11 stocks from 11 sectors) returns for 2023

GARCH model is being used to forecast the volatility of the new portfolio and volatility of sensex for the year 2023.

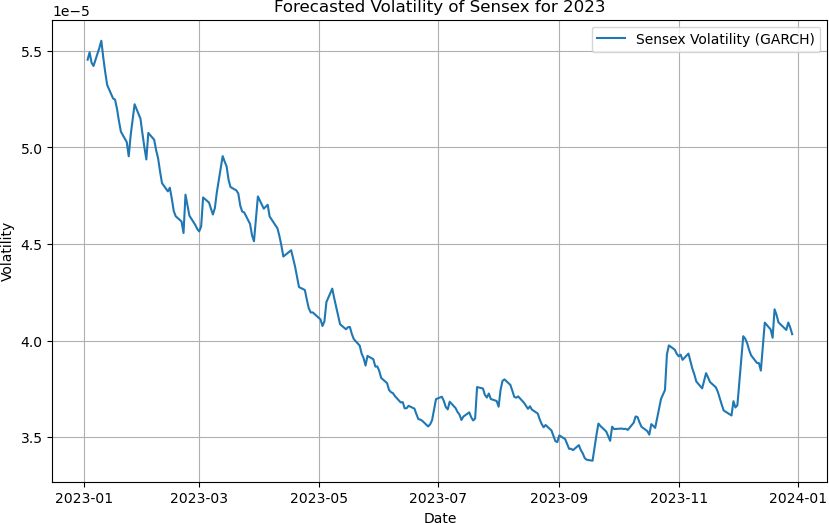
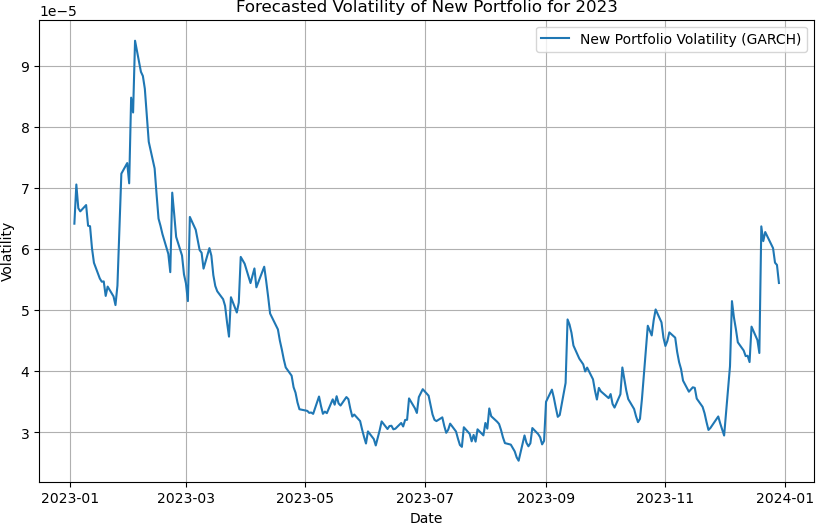


Fig.8.Forcasted volatility of portfolio & Sensex for 2023

|  |  |
| --- | --- |
| Average volatility of portfolio | 0.000043 |
| Average volatility of sensex | 0.000041 |
| Standard deviation of portfolio | 0.000014 |
| Standard deviation of sensex | 0.000006 |
| Ratio of portfolio volatility and sensex volatility | 1.063166 |

Table 3.Results from GARCH model

## Top 25 percentile stocks with highest Sharpe ratio

Sharpe ratio is a measure of risk adjusted i.e., the excess returns to those above and industry benchmark or the risk free rate of return .The higher the Sharpe ratio the better it is compared to the similar portfolios .Therefore this is taken as a metric and the top performing (25 percentile) stocks which fall under this category are being selected.

Sharpe ratio = [Fund Return – Risk-Free Return]/Standard deviation of the fund The XGBoost classifier model is trained on the historic data and it is used to classify the

performances and visualise the returns of the stocks. The model with an accuracy of 73.42% picked Bajaj finance,Britannia, Bajaj FinServ Titan ,Nestle India ,Asian paint, Powergrid,HDFC Bank ,Hindustan Unilever limited,JSW Steel and Eichermot as the top 25 percentile stocks based on Sharpe ratio.

Historic returns and their historical stock prices are given in figure 9 & 10.

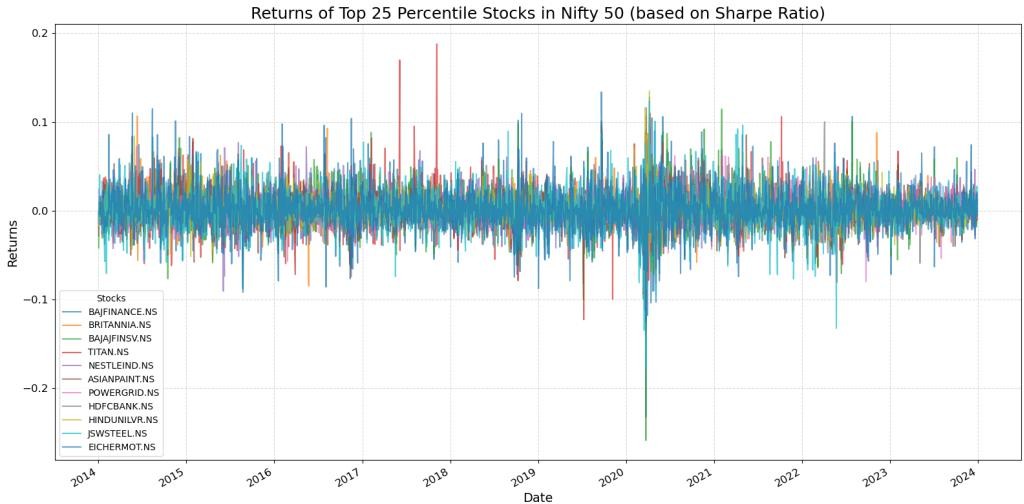


Fig.9.Returns of top 25 percentile stocks in Nifty 50 (2014-2023)

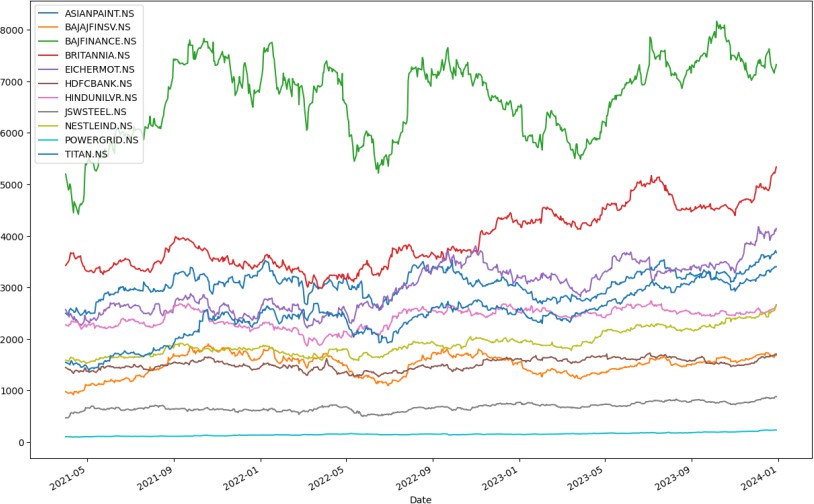


Fig.10Historic stock prices of the stocks (2014-2023)

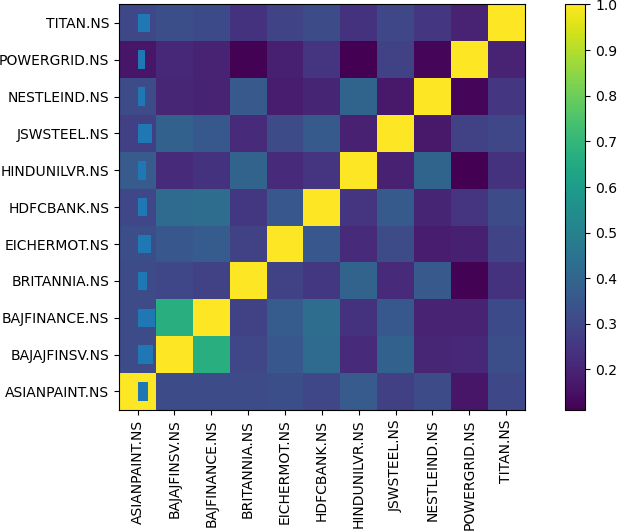


Fig.11.Heap map of the selected stocks

Similar to the previous portfolio mean variance optimization is being done to this portfolio.

The return ,volatility and Sharpe ratio of the portfolio are studied at the minimum volatility and maximum sharpe ratio point

|  |  |  |
| --- | --- | --- |
|  | At minimum volatility point | At maximum Sharpe point |
| Volatility | 0.1484 | 0.1628 |
| Returns | 24.58% | 29.29% |
| Sharpe ratio | 1.6558 | 1.7982 |

Though there is a considerable difference in the volatility between the two portfolios there is a significantly higher returns and slightly higher sharper ratio associated with the portfolio

with maximum Sharpe point .Therefore this portfolio is been selected and the cummulative returns of this portfolio for the year 2023 is obtained which is 22.87%

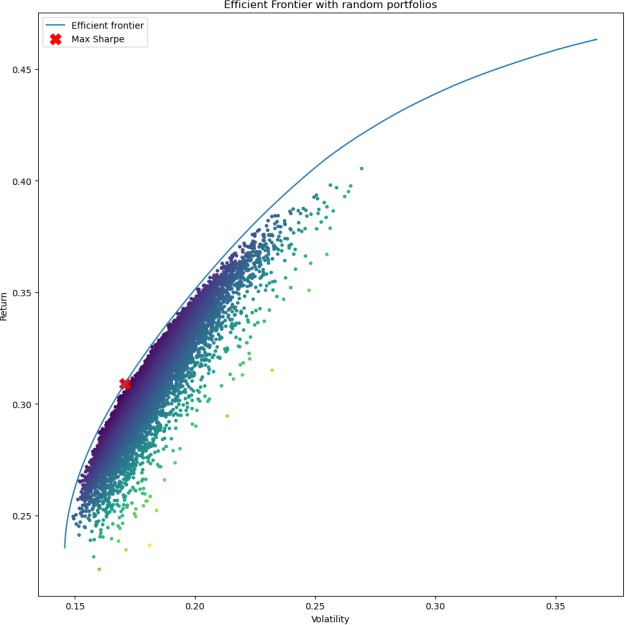


Fig.12.Efficient frontier of the portfolio

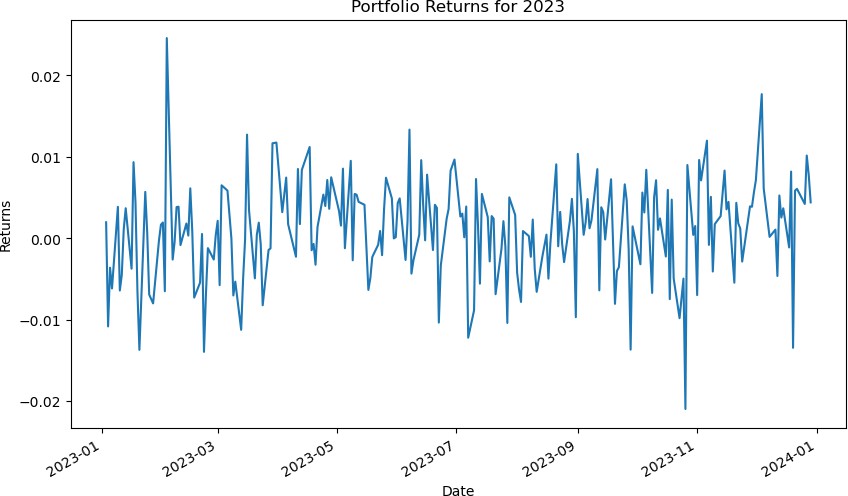


Fig.13.Cummulative returns of the portfolio for the year 2023

|  |  |
| --- | --- |
| STOCKS | WEIGHTS |
| ASIANPAINT.NS | 0.0865488739626426 |
| BAJAJFINSV.NS | 0.03612711716679851 |
| BAJFINANCE.NS | 0.08328986359394318 |
| BRITANNIA.NS | 0.05790175687712292 |
| EICHERMOT.NS | 0.10003793890806101 |
| HDFCBANK.NS | 0.12257941277206764 |
| HINDUNILVR.NS | .09769210313091992 |
| JSWSTEEL.NS | .07365154181402357 |
| NESTLEIND.NS | .15186365472020794 |
| POWERGRID.NS | 0.09705170070278274 |
| TITAN.NS | .09325603635142993 |

Table. 4 Wights of stocks in portfolio at highest Sharpe point

GARCH model is being used to forecast the volatility of the new portfolio and volatility of Sensex for the year 2023.

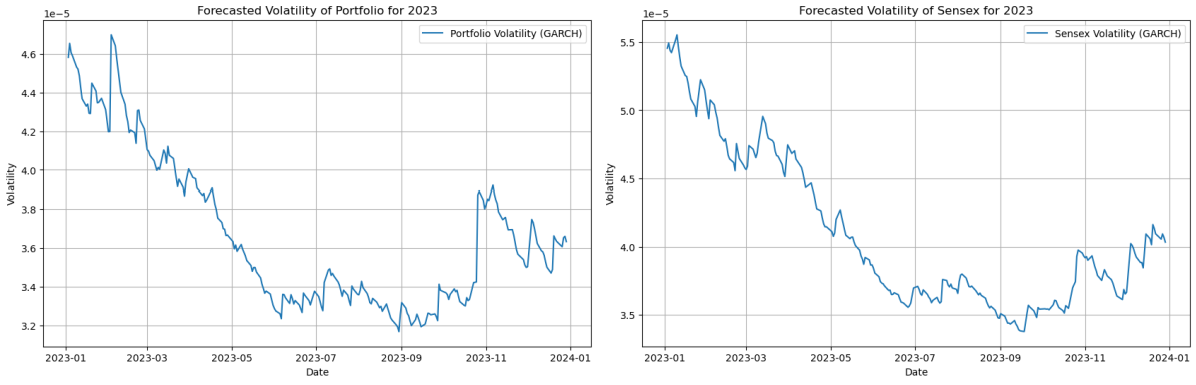


Fig.14Forcasted volatility of portfolio & Sensex for 2023

|  |  |
| --- | --- |
| Average volatility of portfolio | 0.000037 |
| Average volatility of sensex | 0.000041 |
| Standard deviation of portfolio | 0.000004 |
| Standard deviation of sensex | 0.000006 |
| Ratio of portfolio volatility and sensex volatility | 0.898571 |

Table 5.Results from GARCH model

## Top 25 percentile companies with lowest MAPE

The root means square error(RMSE) and mean absolute percentage error(MAPE) aretwo well known metrices in accessing error.Here both the values are being calculated and the one that is being smaller is chosen as the metric .The smaller the error the better is the model .On using XGBoost modelMAPE is found to be a better metric.

The top 25 percentile companies with least RMSE were IOC,JSWSTEEL and POWERGRID with RMSE values 2.016,10.502 and 2.101 respectively.

The top 25 percentile companies with least MAPE were HDFC BANK,NESTLEIND and POWERGRID with MAPE values 1.285,1.337 and 1.363 respectively. On comparing the values MAPE is found to be the least therefore its chosen.

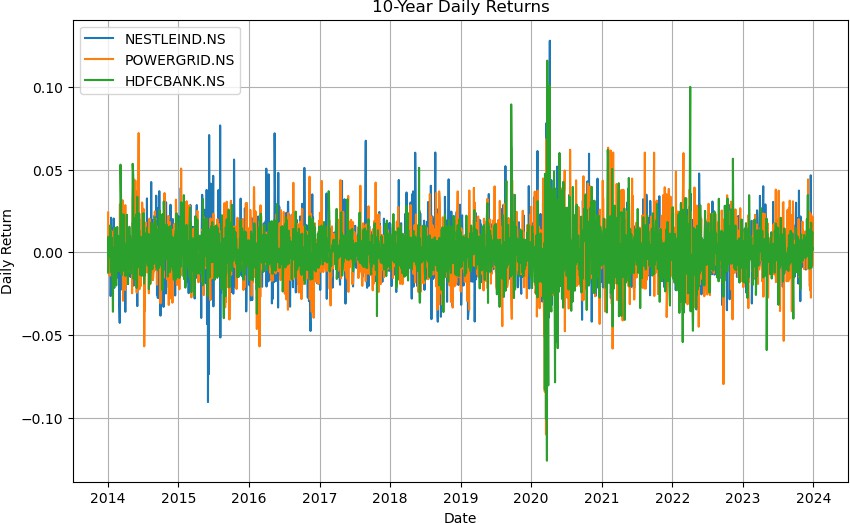


Fig.15.Daily returns of top 25 percentile companies with least MAPE

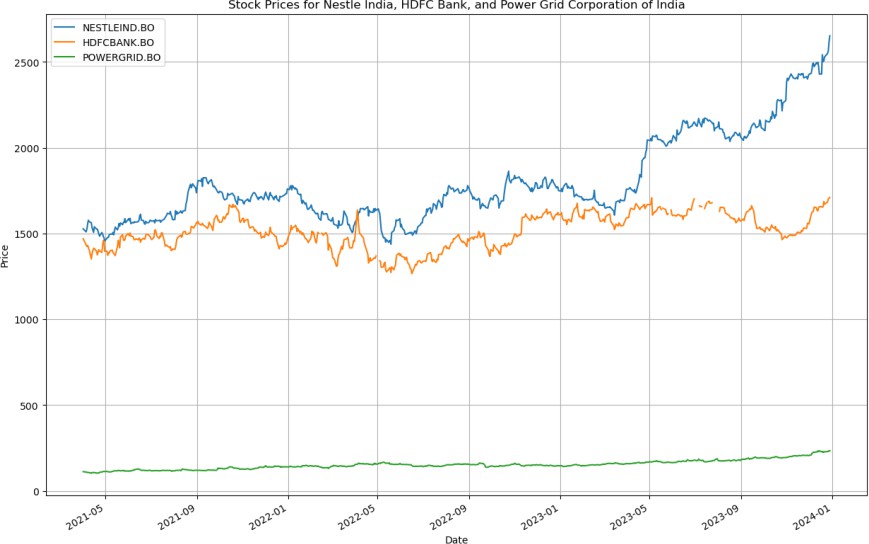


Fig.16Historic stock prices of the selected stocks(2014-2023)

Similar to the pervious portfolios the mean variance optimization is used to weight the stocks in the portfolio.The return ,volatility and Sharpe ratio of the portfolio are studied at the minimum volatility and maximum Sharpe ratio point

|  |  |  |
| --- | --- | --- |
|  | At minimum volatility point | At maximum Sharpe point |
| Volatility | 0.1583 | 0.1585 |
| Returns | 22.57% | 22.63% |
| Sharpe ratio | 1.4257 | 1.5885 |

The difference in the volatility between the two portfolios is zero but there is slightly higher returns and higher sharper ratio associated with the portfolio with maximum Sharpe point

.Therefore this portfolio is been selected and the cummulative returns of this portfolio for the year 2023 is obtained which is 33.00%

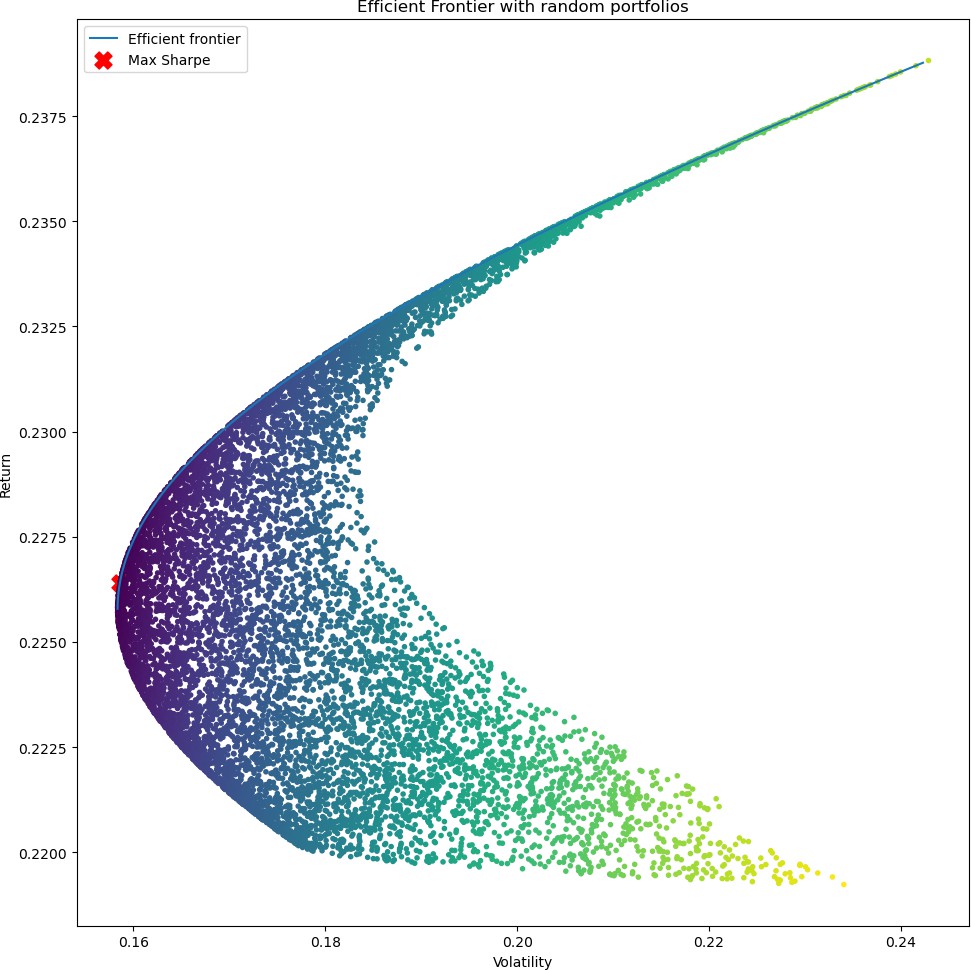


Fig.17.Efficient frontier of portfolio

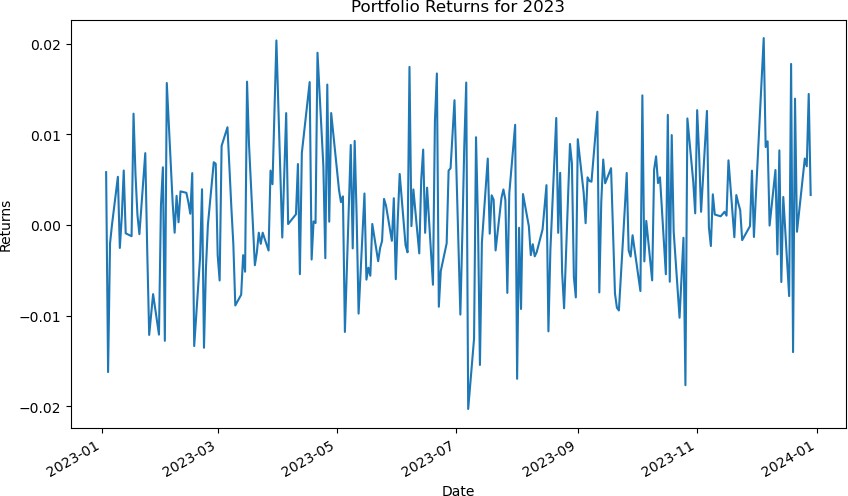


Fig.18.Cummulative returns of portfolio for the year 2023

GARCH model is being used to forecast the volatility of the new portfolio and volatility of Sensex for the year 2023.

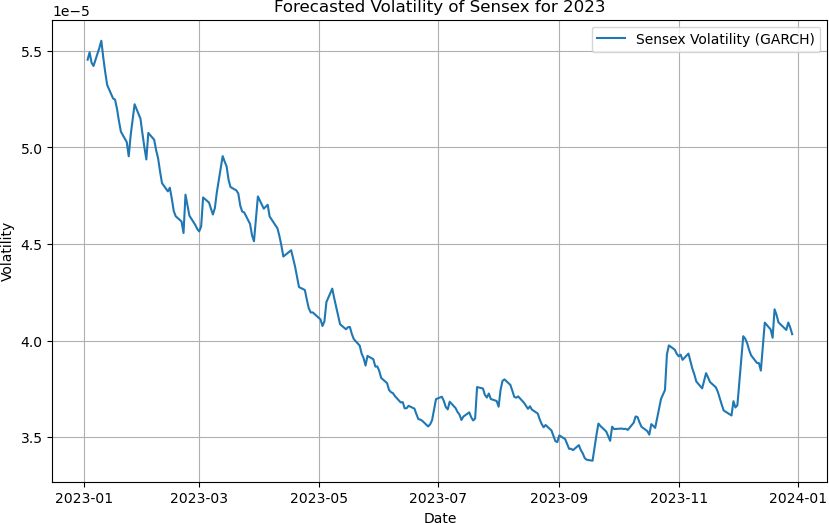
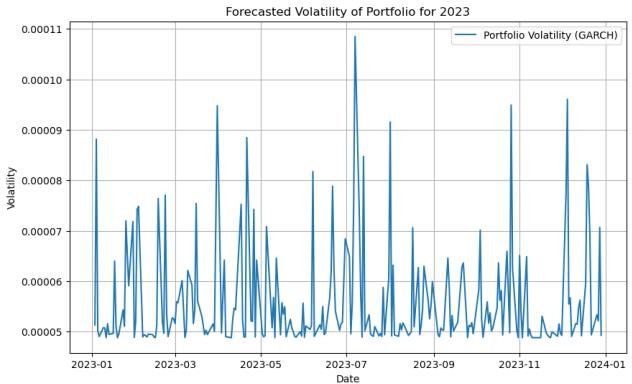


Fig.19 Forcasted volatility of portfolio & Sensex for 2023

|  |  |
| --- | --- |
| Average volatility of portfolio | 0.000056 |
| Average volatility of sensex | 0.000041 |
| Standard deviation of portfolio | 0.00001 |
| Standard deviation of sensex | 0.000006 |
| Ratio of portfolio volatility and sensex volatility | 1.369688 |

Table 5.Results from GARCH model

## All Nifty 50 companies

In order to construct a portfolio of the Nifty 50 companies all the stocks are being selected irrespective of their performances in the market and the mean variance optimization is applied.

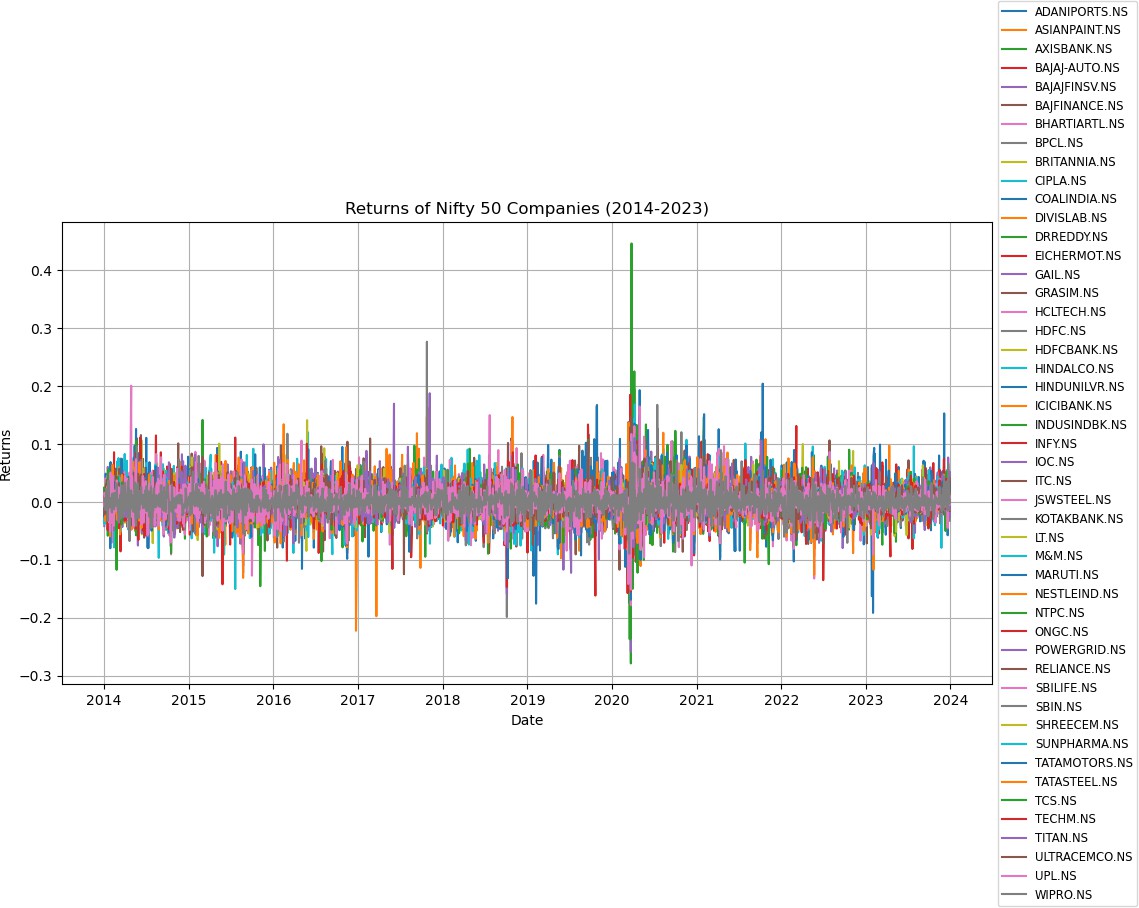


Fig.20.Returns of the Nifty 50 companies (2014-2023)

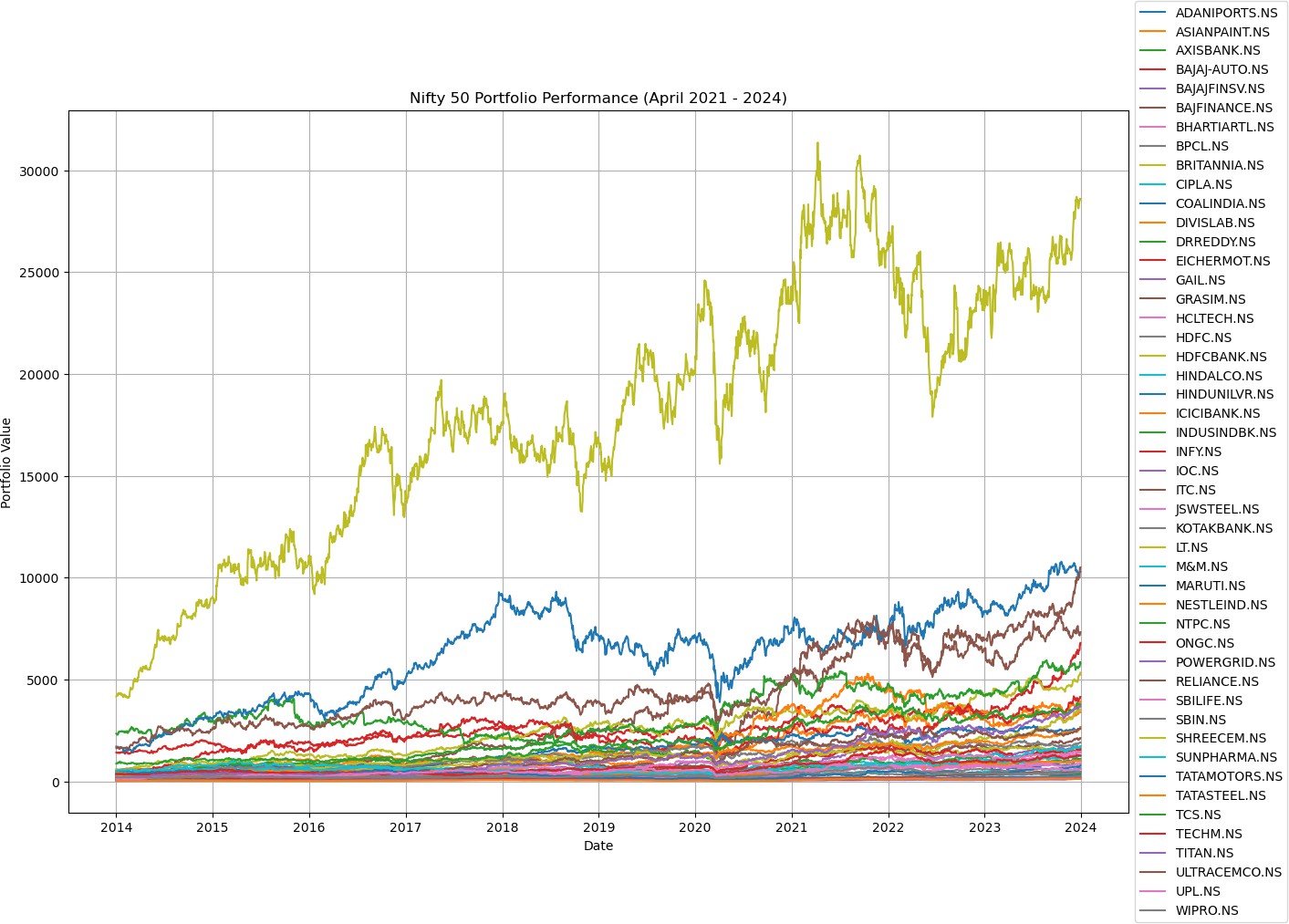


Fig.21.Historical stock returns of the Nifty 50 companies

On optimizing using the mean variance optimization method the cumulative returns of the portfolio was found to be 30.22% for the year 2023.

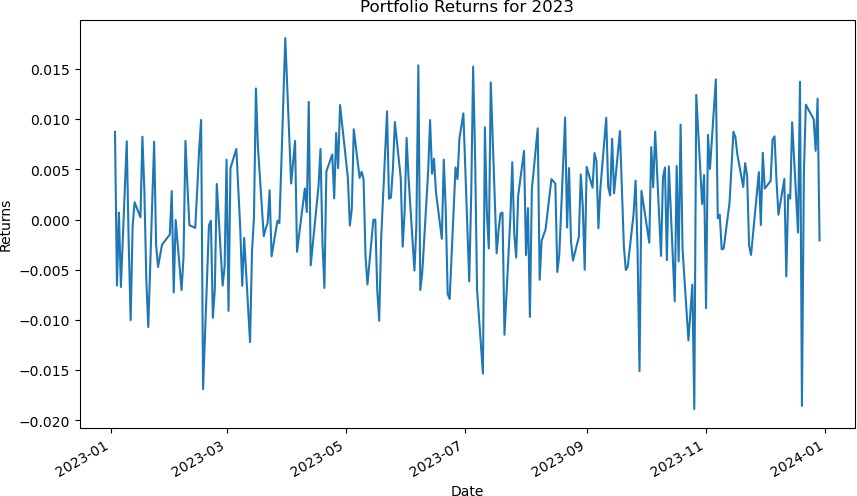


Fig.22.Portfolio returns for 2023

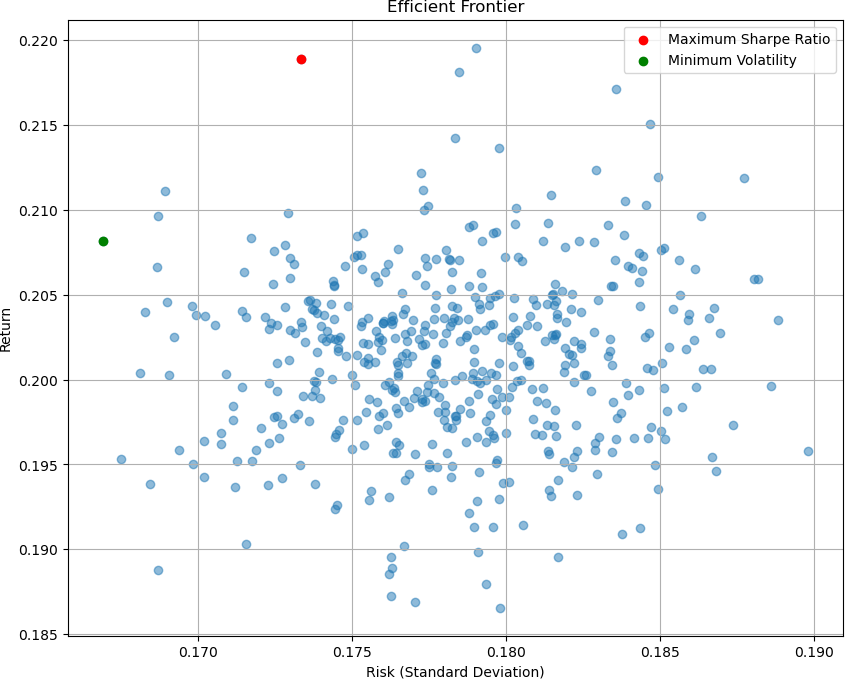


Fig.23.Efficient frontier of mean variance optimization

It was also noted that the mean variance optimization method did not significantly diversify the portfolio across all the stocks in the universe .

|  |  |
| --- | --- |
| STOCKS | WEIGHTS |
| BHARTIARTL.NS | 0.01751 |
| DIVISLAB.NS | 0.1456 |
| HCLTECH.NS | 0.02367 |
| INFY.NS | 0.09284 |
| NESTLEIND.NS | 0.30518 |
| NTPC.NS | 0.02308 |
| POWERGRID.NS | 0.089 |
| RELIANCE.NS | 0.01045 |
| TCS.NS | 0.09785 |
| TITAN.NS | 0.19482 |

Table 6 Weights of the stocks in the portfolio

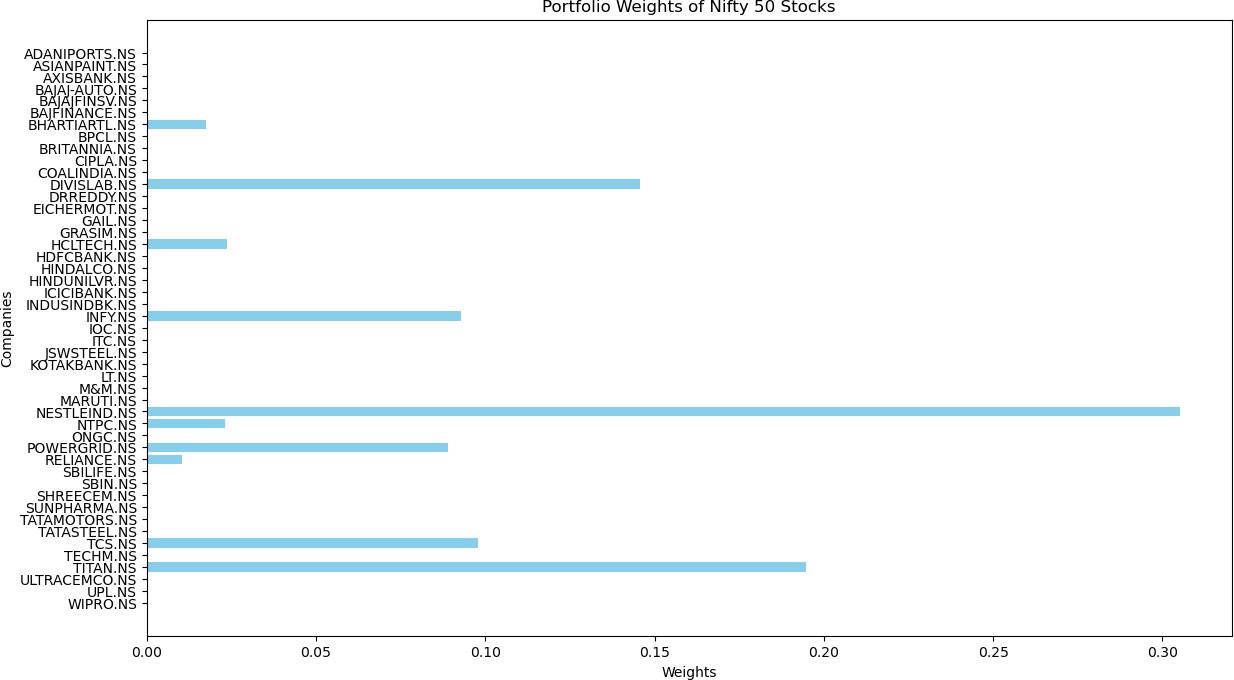


Fig.24.Bar plot of weights of all the stocks in the portfolio

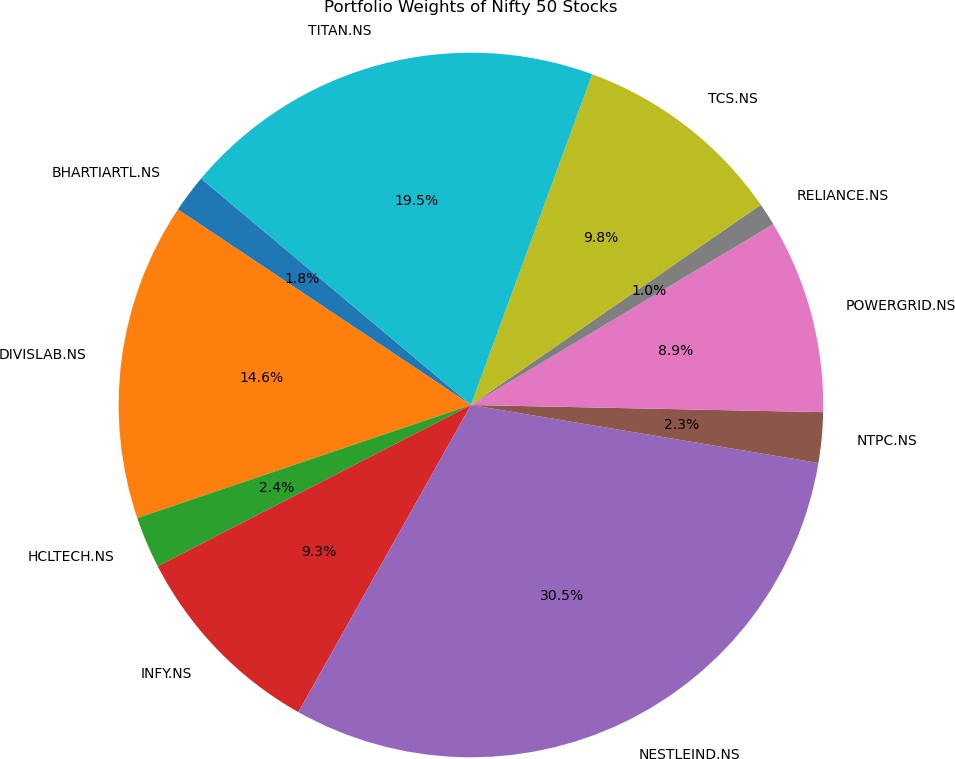


Fig.25.Pie chart of all the stocks in the portfolio

GARCH model is being used to forecast the volatility of the new portfolio and volatility of Sensex for the year 2023.

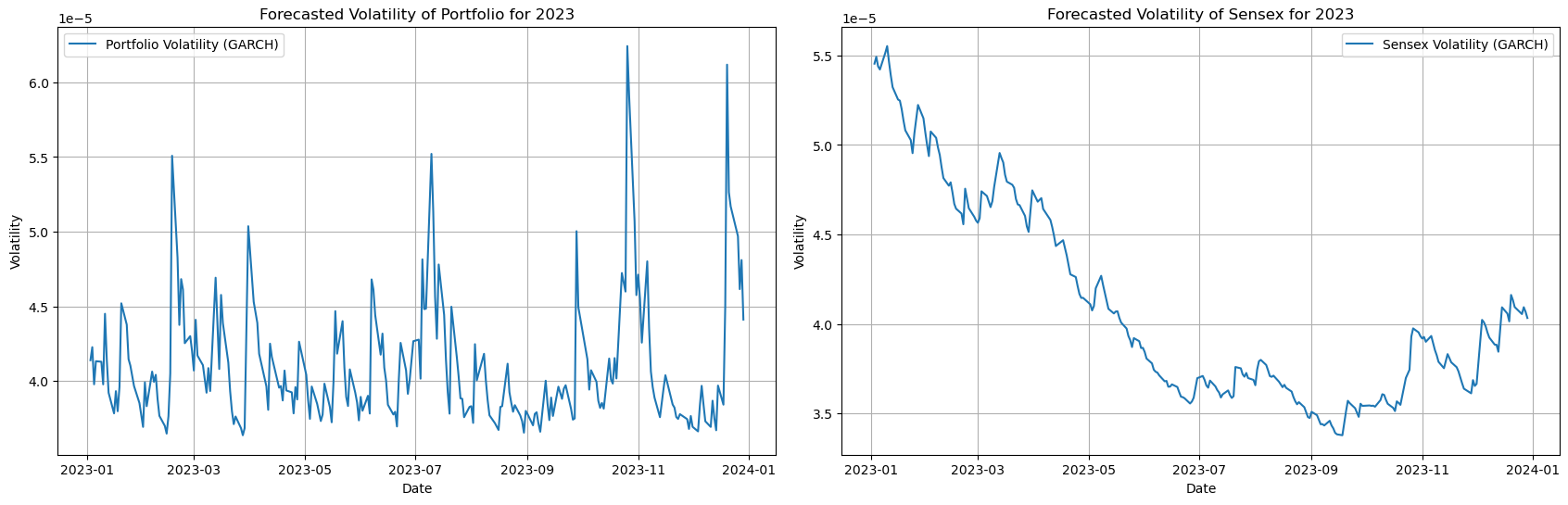


Fig.26 Forcasted volatility of portfolio & Sensex for 2023

|  |  |
| --- | --- |
| Average volatility of portfolio | 0.000041 |
| Average volatility of sensex | 0.000041 |
| Standard deviation of portfolio | 0.000004 |
| Standard deviation of sensex | 0.000006 |
| Ratio of portfolio volatility and sensex volatility | 1.002458 |

Table 7 Results from GARCH model

**Agglomerative Hierarchical Clustering & Hierarchical Risk Parity Asset Allocation**

The stocks of the Nifty 50 companies are clustered using Agglomerative Hierarchical clustering model . To do so the covariance matrix is calculated based on the change in stock prices and hierarchical clustering is performed using Ward’s method on the correlation Matrix and the dendogram and Heat map are used to visualize the relation between all the Nifty 50 companies on the basis of their price movements.

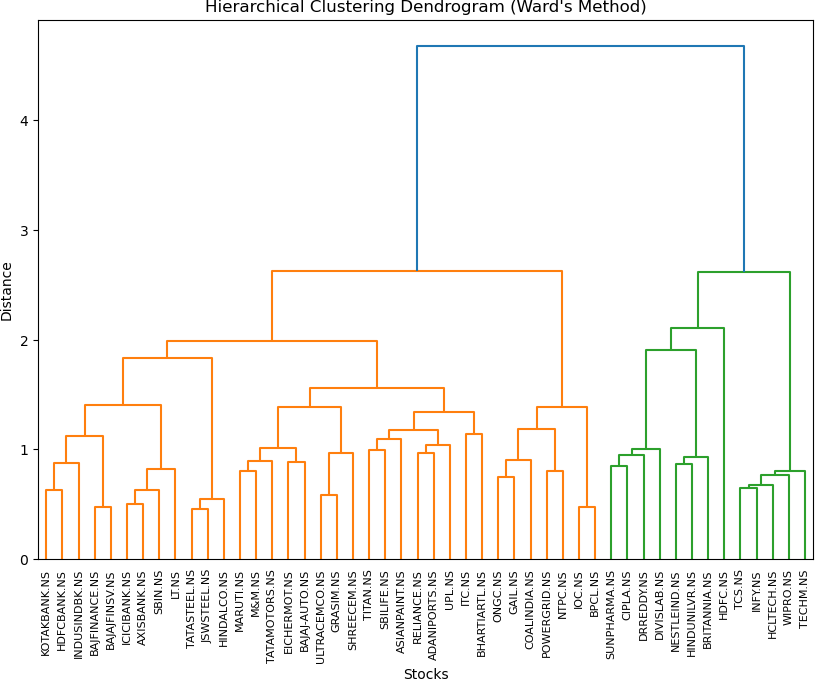


Fig.27 Hierarchical clustering dendogram of Nifty 50 companies

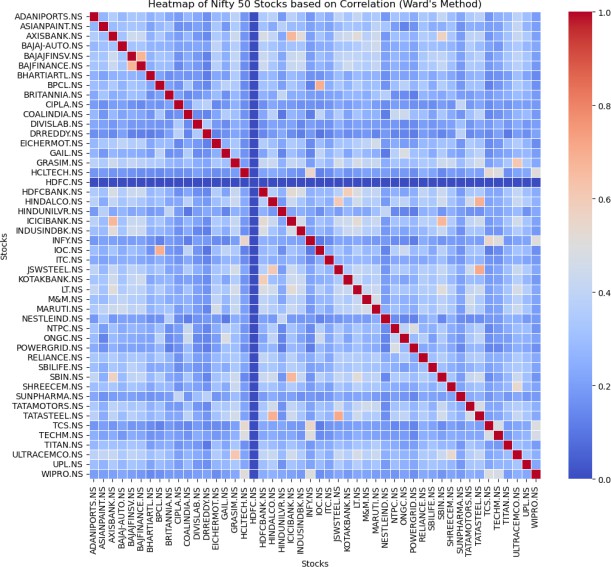


Fig.27.Heat map of the Nifty 50 companies

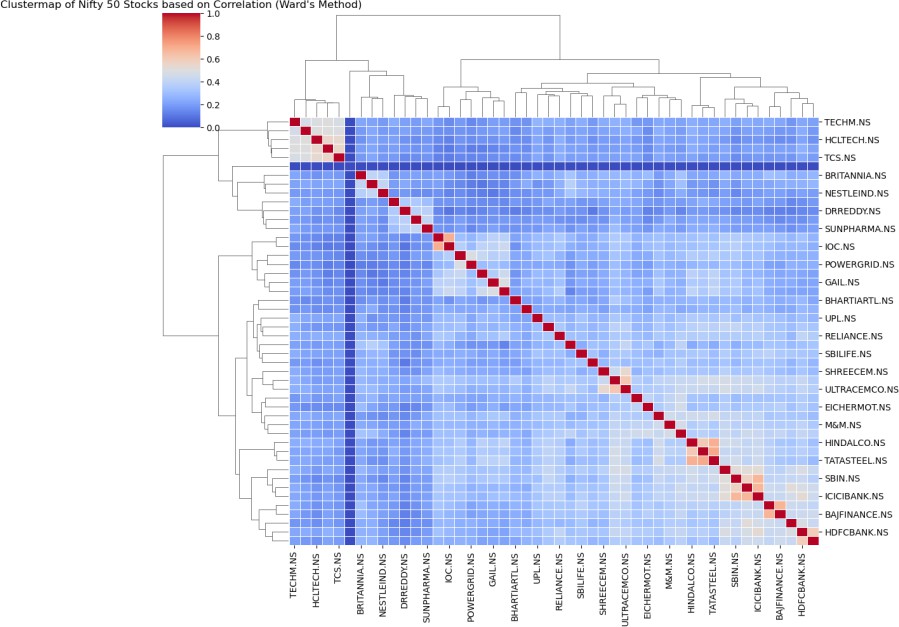


Fig.28 Cluster map of the Nifty 50 companies.

Gap statistics helps in determining optimal number of cluster by identifying the point where the gap value starts to decrease after reaching a maximum .The gap statistics for each number of clusters is calculated as the difference between the log of the average sum of square distance for the reference data set and the log of sum of square distance of the actual data set

.This helps identifying optimal number of clusters to be considered which is 10 here.

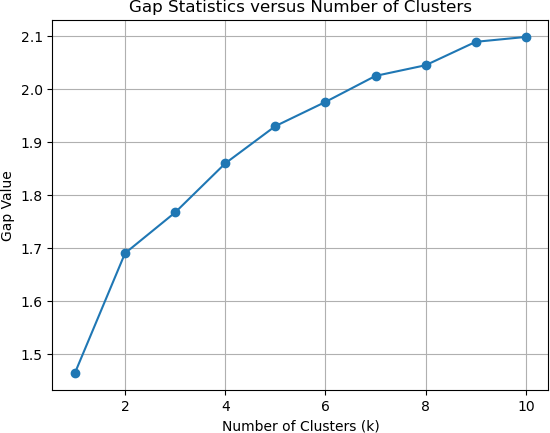


Fig.29.Gap statistics versus number of clusters

Cluster 1: BAJAJ-AUTO.NS, EICHERMOT.NS, GRASIM.NS, M&M.NS, MARUTI.NS, SHREECEM.NS, TATAMOTORS.NS, ULTRACEMCO.NS

Cluster 2: ADANIPORTS.NS, ASIANPAINT.NS, BHARTIARTL.NS, ITC.NS, RELIANCE.NS, SBILIFE.NS, TITAN.NS, UPL.NS

Cluster 3: BPCL.NS, COALINDIA.NS, GAIL.NS, IOC.NS, NTPC.NS, ONGC.NS, POWER GRID.NS

Cluster 4: CIPLA.NS, DIVISLAB.NS, 'DRREDDY.NS, SUNPHARMA.NS Cluster 5: AXISBANK.NS, ICICIBANK.NS, LT.NS, SBIN.NS

Cluster 6: BAJAJFINSV.NS, BAJFINANCE.NS, HDFCBANK.NS, INDUSINDBK.NS, KOTAKBANK.NS

Cluster 7: BRITANNIA.NS, HINDUNILVR.NS, NESTLEIND.NS Cluster 8: HINDALCO.NS, JSWSTEEL.NS, TATASTEEL.NS

Cluster 9: HCLTECH.NS,INFY.NS, TCS.NS, TECHM.NS, WIPRO.NS

Cluster 10: HDFC.NS

|  |  |  |
| --- | --- | --- |
| HRP | Weights:  Ticker | Weight |
| 0 | ADANIPORTS.NS | 0.013454 |
| 1 | ASIANPAINT.NS | 0.006020 |
| 2 | AXISBANK.NS | 0.049415 |
| 3 | BAJAJ-AUTO.NS | 0.005383 |
| 4 | BAJAJFINSV.NS | 0.009785 |
| 5 | BAJFINANCE.NS | 0.050904 |
| 6 | BHARTIARTL.NS | 0.011752 |
| 7 | BPCL.NS | 0.014854 |
| 8 | BRITANNIA.NS | 0.033917 |
| 9 | CIPLA.NS | 0.007529 |
| 10 | COALINDIA.NS | 0.009270 |
| 11 | DIVISLAB.NS | 0.042703 |
| 12 | DRREDDY.NS | 0.007547 |
| 13 | EICHERMOT.NS | 0.010826 |
| 14 | GAIL.NS | 0.043539 |
| 15 | GRASIM.NS | 0.010609 |
| 16 | HCLTECH.NS | 0.009203 |
| 17 | HDFC.NS | 0.025055 |
| 18 | HDFCBANK.NS | 0.060915 |
| 19 | HINDALCO.NS | 0.019181 |
| 20 | HINDUNILVR.NS | 0.012469 |
| 21 | ICICIBANK.NS | 0.019481 |
| 22 | INDUSINDBK.NS | 0.036662 |
| 23 | INFY.NS | 0.008592 |
| 24 | IOC.NS | 0.005840 |
| 25 | ITC.NS | 0.045753 |
| 26 | JSWSTEEL.NS | 0.007549 |
| 27 | KOTAKBANK.NS | 0.007730 |
| 28 | LT.NS | 0.038150 |
| 29 | M&M.NS | 0.008617 |
| 30 | MARUTI.NS | 0.005969 |
| 31 | NESTLEIND.NS | 0.031367 |
| 32 | NTPC.NS | 0.011776 |
| 33 | ONGC.NS | 0.006145 |
| 34 | POWERGRID.NS | 0.035651 |
| 35 | RELIANCE.NS | 0.013463 |
| 36 | SBILIFE.NS | 0.009891 |
| 37 | SBIN.NS | 0.042368 |
| 38 | SHREECEM.NS | 0.022046 |
| 39 | SUNPHARMA.NS | 0.017388 |
| 40 | TATAMOTORS.NS | 0.029460 |
| 41 | TATASTEEL.NS | 0.042144 |
| 42 | TCS.NS | 0.046472 |
| 43 | TECHM.NS | 0.008492 |
| 44 | TITAN.NS | 0.014186 |
| 45 | ULTRACEMCO.NS | 0.030478 |

Hierarchical Risk Parity allocation technique is suggested based on recursive bisection and inverse variance allocation. The method starts with all of the ordered assets as one set, with al l unit weights equal to one. *(Eidenvall, A,. 2021)* The sets are then bisected into two subsets a nd the subsets are then bisected until all subsets includes only one asset. At each step of th e bisection, assets weights in the two subsets are rescaled with a factor α and (1 − α) respect ively. We obtain factor α by calculating the inverse variance allocation of the two subsets. The subsets variance is also determined by using an inverse variance allocation within the sub set. *(Eidenvall, A,. 2021)*

Recursive bisection in HRP rather focuses on the obtained order at the bottom level of the de ndrogram and ignores the obtained clusters at different levels of the dendrogram.

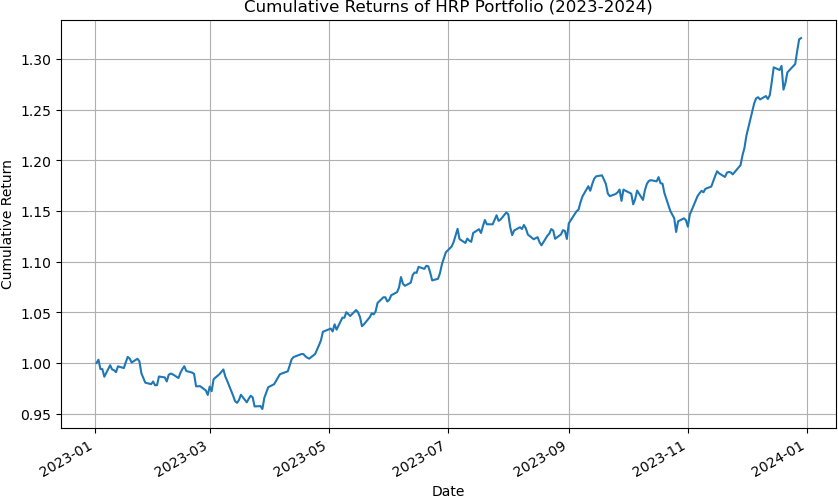


Fig.30.Cummulative returns of HRP portfolio

It can also be noticed that the HRP asset allocation technique has successfully diversified the portfolio unlike the mean variance optimization method which allocated funds to only 10 stocks in the Nifty 50 stocks universe.

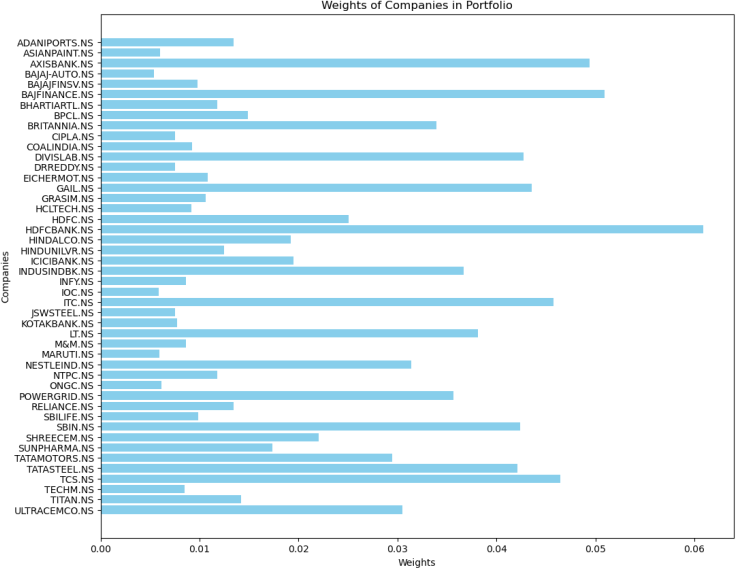


Fig.31.Bar plot of weights of stocks in the portfolio

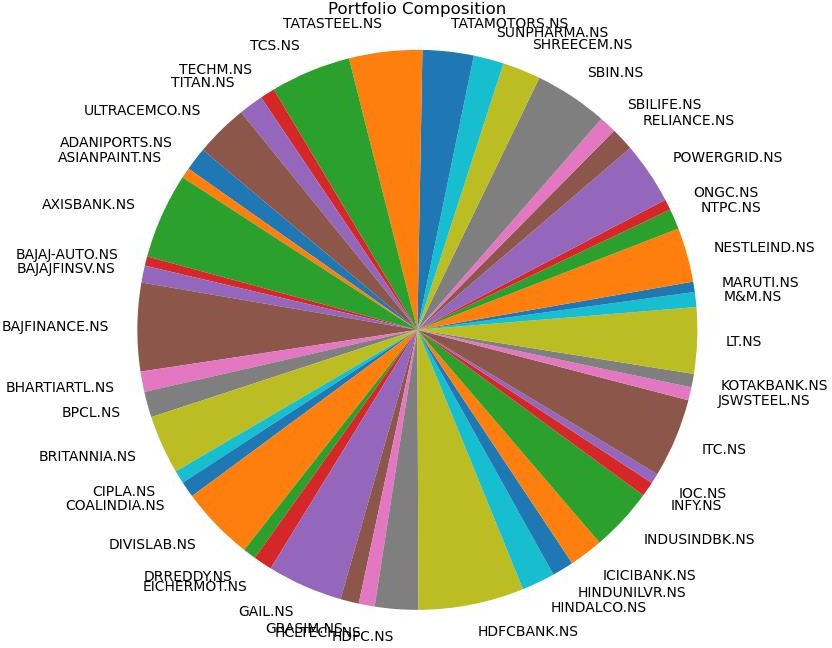


Fig.32.Pie chart of weights of stocks in the portfolio

GARCH model is being used to forecast the volatility of the new portfolio and volatility of Sensex for the year 2023.

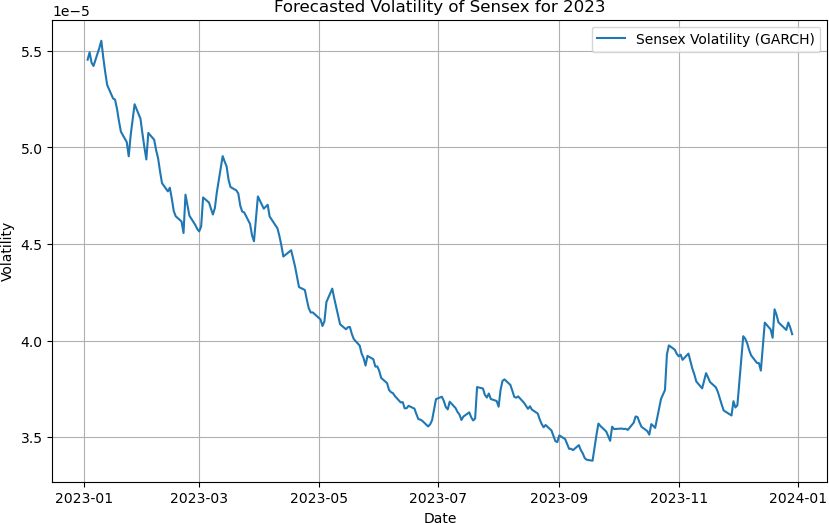
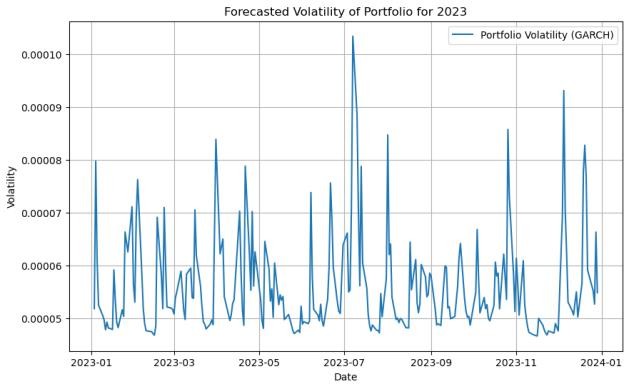


Fig.33 Forcasted volatility of portfolio & Sensex for 2023

|  |  |
| --- | --- |
| Average volatility of portfolio | 0.000056 |
| Average volatility of sensex | 0.000041 |
| Standard deviation of portfolio | 0.000009 |
| Standard deviation of sensex | 0.000006 |
| Ratio of portfolio volatility and sensex volatility | 1.370067 |

Table 8.Results from GARCH model

**Results**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 11 stocks from 11sectors(MPT allocation) | Top 25 percentile stocks with highest Sharpe ratio(MPT allocation) | Top 25 percentile stocks with lowest MAPE(MPT  allocation) | All Nifty 50 stocks (MPT allocation) | All Nifty 50 stocks (Hierarchical Risk Parity) |
| Annualised Returns | 0.213 | 0.207 | 0.216 | 0.237 | 0.218 |
| Volatility | 0.011 | 0.011 | 0.010 | 0.009 | 0.010 |
| Standard deviation | 0.027 | 0.026 | 0.045 | 0.042 | 0.013 |
| Sharpe Ratio | 1.893 | 1.885 | 1.018 | 1.380 | 3.266 |
| Sortino ratio | 2.365 | 2.665 | 4.898 | 2.667 | 3.638 |
| Calmar Ratio | 0.646 | 0.598 | 0.807 | 0.902 | 0.666 |
| Maximum Drawdown | -0.329 | -0.347 | -0.268 | -0.263 | -0.626 |
| 95% cVar | -0.025 | -0.026 | -0.021 | -0.021 | -0.001 |
| 99% cVar | -0.044 | -0.045 | -0.037 | -0.036 | -0.003 |
| Cummulative returns | 0.307 | 0.228 | 0.329 | 0.302 | 0.320 |
| Volatility ratio to Sensex | 1.060 | 0.898 | 1.369 | 1.002 | 1.370 |

Table 19.Summary of the performances of the five distinct portfolio

## Sortino ratio

Sortino’s Ratio is = [Fund Return – Risk-Free Return]/Downside Risk

It measures the performance of the investment relative to the downward deviation. Unlike Sharpe, it doesn't take into account the total volatility in the investment.

## Calmar ratio

Calmar ratio = Average Annual rate of return / Maximum drawdown

## Maximum Drawdown

It measures the maximum fall in the value of the investment which is a significant indication of stock’s market performance.

## 95% cVar& 99% cVar

Expected Shortfall is a risk measure that indicates the average value of a possible loss in an investment that exceeds a 95% and 99% confidence level respectively.

## Volatility ratio to sensex

It measures the relative changes in the portfolio’s price relative to market indicators changes.

## Cummulative returns

Aggregate levels of gain or loss over a period of time (10 years here)

On considering the annualised returns of the five portfolio the portfolio with all Nifty 50 companies optimised using mean variance optimisation seems to be the highest. Also the Clamar ratio of the of the portfolio seems to be the highest of all the values and the maximum drawdown values in the table also conclude that this portfolio has the lowest maximum drawdown .On comparing the sharpe ratio of the five portfolios , the portfolio which has been constructed using agglomerative hierarchcial clustering seems to have significantly higher sharpe ratio making it better pick in comparison with the other four portfolios as it yields higher risk adjusted returns.

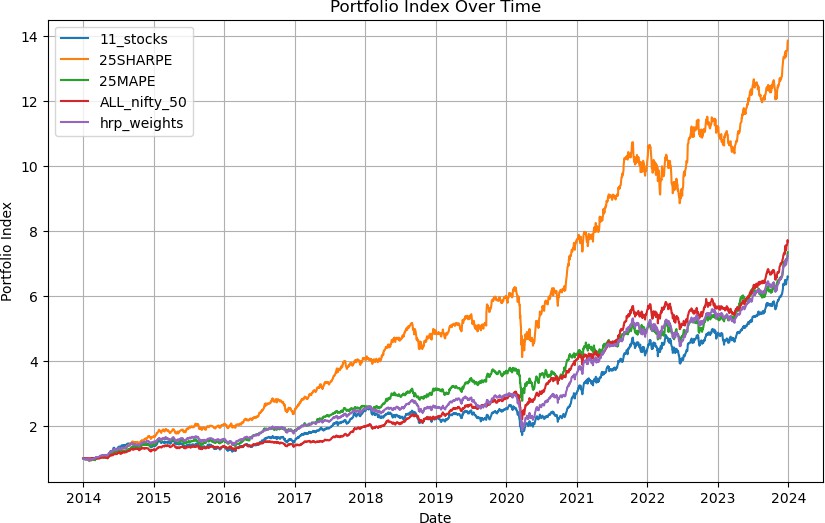


Fig.34. The portfolio indices of the five portfolios for the years 2014 – 2023

**Conclusions**

On considering the above criterions the portfolio obtained using agglomerative hierarchical clustering and weighted using hierarchical risk parity seems to be a considerably better pick as it has a significantly higher Sharpe ratio and extremely low 95% cVar and 99%cVar. Also the cumulative returns of this portfolio is 32% and the standard deviation, volatility of this portfolio is significantly lower .The down side of this portfolio could be a lesser annualised return and higher volatility ratio. However if the Sortino ratio is to be considered as a metric then the portfolio with top 25% stocks with lowest MAPE could be considered. The portfolio

with the top 25% stocks with highest Sharpe ratio and weighted using of mean variance optimisation though has a sharpe a ratio of 1.885 which is quite less than the portfolio weighted using HRP it seems to have the lowest volatility ratio to the sensex.

What is considered a good portfolio is always a balance between how much risk one is willing to take in comparison to the return where the clamar ratio is a good measured for that

.In that case the portfolio with all Nifty 50 stocks weighted using mean variance optimisation seems to be a good pick .However, it all depends on the investors to consider whichever metric is suitable with respect to their goals and comfort levels of risk.

**References**

Here are the references arranged in alphabetical order:

1. Behr, P., Güttler, A., & Miebs, F. (2008). Is minimum-variance investing really worth the while? An analysis with robust performance inference. EDHEC-Risk working paper.
2. Choueifaty, Y., & Coignard, Y. (2008). Toward Maximum Diversification. Journal of Portfolio Management - J PORTFOLIO MANAGE, 35(Oct).
3. Christoffersen, P., et al. (2012). Is the Potential for International Diversification Disappearing? A Dynamic Copula Approach. Review of Financial Studies, 25(May).
4. Eidenvall, A. (2021). Hierarchical Clustering To Improve Portfolio Tail Risk Characteristics. LUTFMS-3409-2021.
5. Hansen, K. B. (2020). The virtue of simplicity: On machine learning models in algorithmic trading. Big Data & Soc., 7(1), 2053951720926558.
6. Karina Marvin. (2015). Creating diversified portfolios using cluster analysis. Princeton University.
7. Kaufman, L., & Rousseeuw, P. J. (1990). Finding groups in data: An introduction to cluster analysis. Wiley, New York.
8. Kuchler, T., Piazzesi, M., & Stroebel, J. (2023). Housing market expectations. In Handbook of Economic Expectations. Academic Press, pp. 163-191.
9. May, K. (2022). Forecast Based Portfolio Optimisation Using XGBoost. Doctoral dissertation, University of the Witwatersrand, Johannesburg.
10. Molyboga, M. (2020). A Modified Hierarchical Risk Parity Framework for Portfolio Management. The Journal of Financial Data Science, 2(3).
11. Nanakorn, N., & Palmgren, E. (2021). Hierarchical Clustering in Risk-Based Portfolio Construction.
12. Obeidat, S., et al. (2018). Adaptive Portfolio Asset Allocation Optimization with Deep Learning. International Journal On Advances in Intelligent Systems, 11(1 and 2), 25–34.
13. Pardo. (2008). The Evaluation and Optimization of Trading Strategies, 2nd edition. John Wiley & Sons, Inc.
14. Ren, Z. (2005). Portfolio construction using clustering methods. Worcester Polytechnic Insitute.
15. Scott-Ram, R. (2000). Managing portfolio risk for periods of stress. World Gold Council.
16. Stein, D. M., & Garland, J. P. (2008). Investment Management for Taxable Investors. In

F.J. Fabozzi. Handbook of finance (Vol. II). John Wiley & Sons Inc.

1. Tatsat, H., Puri, S., & Lookabaugh, B. (2020). Machine Learning and Data Science Blueprints for Finance From Building Trading Strategies to Robo-Advisors Using Python. O’Reilly Media, Inc, California, United States.
2. Thuraisingham, B. M., & Ceruti, M. G. (2000). Understanding data mining and applying it to command, control, communications and intelligence environments. Proceedings 24th Annual International Computer Software and Applications Conference, Taipei, Taiwan, pp. 171–175.
3. Uysal, E., Trainer, F. H., & Reiss, J. (2001). Revisiting mean-variance optimization. Journal of Portfolio Management, 27(4), 71-82.
4. Wang, M. (2022, December). A Portfolio Strategy Based on XGBoost Regression and Monte Carlo Method. In 2022 2nd International Conference on Economic Development and Business Culture (ICEDBC 2022) (pp. 896-902). Atlantis Press.