TASK 4 S&P 500 ANALYSIS

SUNRIT SARKAR

24-05-2024
Quantitative Analyst Intern

Introduction

S&P (Standard & Poor's) is a rating agency that provides credit ratings for countries, companies, and securities, assessing their creditworthiness. It is also a stock market index that tracks 500 of the largest publicly traded companies in the U.S., 502 to be exact at this point of time. For this task, we analyse the stocks in the index of S&P 500, traded at various US Stock Exchanges, we use some performance metrics to rate them and make a portfolio with the top 10 most profitable stocks among all the stocks present in the index. The portfolio is designed to minimize risks while maximising returns. All the analysis is made using 10 months of data.

Design of the Performance Metrics

The performance of a stock is generally determined by its profitability over a period of time. This is why, the change in the Adjusted Closing Price is used as the primary data source and its changes are modelled. The proposed hypothesis assumes the change in the price of a stock behaves like the Geometric Brownian motion. Geometric Brownian motion is a stochastic process, in which the change in the price is determined by 2 main factors:

- The drift, i.e., the effect of the market
- The random Brownian stochastic process, i.e., the randomness of the investors' decisions

As observed, this model captures the majority of the market forces acting on the stock price. In addition to this, the drift and random factors are linearly dependent on the present value of the stock price. This is also a realistic assumption as the market sentiments are often a result of the present market value of each stock of that particular company. To put it mathematically, let X_t be the adjusted closing price of each stock, and B_t be the random Brownian stochastic process, then,

$$dX_t = \mu X_t dt + \sigma X_t dB_t$$

So far, the rationality of the idea has been argued, a formal solution of the Stochastic Differential Equation has been presented in the next section of the report.

Solution of the Geometric Brownian Motion SDE

The written proof is provided, but before that some theories are presented:

Let B_t be the stochastic process describing the Brownian motion

- $B_0 = 0$
- For $0 \leq t_1 \leq t_2 \cdots \leq t_n$

  $(B_{t_1} - B_0) \perp\!\!\!\perp (B_{t_2} - B_{t_1}) \cdots \perp\!\!\!\perp (B_{t_n} - B_{t_{n-1}})$

  $\perp\!\!\!\perp$ denotes independence

- $(B_{t+\Delta t} - B_t) \sim N(0, \Delta t)$

  $\mathbb{E}[(B_{t+\Delta t} - B_t)] = 0$

  $V[(B_{t+\Delta t} - B_t)] = \Delta t = \mathbb{E}[(B_{t+\Delta t} - B_t)^2]$

Taylor's expansion

$$f(b) - f(a) = f'(a)(b-a) + \frac{1}{2} f''(a)(b-a)^2 + \cdots$$

$$\therefore \quad \Delta f(x) = f'(x)(\Delta x) + \frac{1}{2} f''(x).(\Delta x)^2 + \cdots$$

Now, we start with the formal solution of the SDE:

$$dX_t = \mu X_t dt + \sigma X_t dB_t$$

This cannot be integrated directly because of the presence of X_t on the RHS. To find a way to remove that, let us assume X_t has no random component for now. Thus,

$$dX_t = \mu X_t dt \Rightarrow \frac{dX_t}{X_t} = \mu dt \Rightarrow d(\ln X_t) = \mu dt$$

Thus, we find that taking $f(X_t) = \ln X_t$ solves this problem.

By Taylor's expansion;

$$d(f(X_t)) = f'(X_t).dX_t + \frac{1}{2} f''(X_t)(dX_t)^2 + \cdots$$

For convenience, we will take the first 2 terms. $f'(X_t)$, $f''(X_t)$ and (dX_t) are known. $(dX_t)^2$ can be found out using (dX_t).

$$(dX_t)^2 = (\mu X_t dt + \sigma X_t dB_t)^2 = X_t^2\left(\mu^2 (dt)^2 + 2\mu\sigma (dt)(dB_t) + \sigma^2 (dB_t)^2\right)$$

As (dB_t) is finite and (dt) is infinitely small, $(dt)^2$ and $(dt)(dB_t)$ terms become very small as compared to $(dB_t)^2$ and may be ignored.

$$\therefore (dX_t)^2 \simeq \sigma^2 X_t^2 (dB_t)^2$$

Using the properties of Brownian motion,

$$\mathbb{E}[(dB_t)^2] = \mathbb{E}[(B_{t+dt} - B_t)^2] = \mathbb{E}[dt] = dt$$

Thus $(dX_t)^2 \simeq \sigma^2 X_t^2 dt$

Now we have,

$$d(\ln X_t) = \frac{1}{\cancel{X_t}}(\mu \cancel{X_t} dt + \sigma \cancel{X_t} dB_t) + \left(\frac{-1}{2\cancel{X_t^2}}\right)(\sigma^2 \cancel{X_t^2} dt)$$

$$= \left(\mu - \frac{\sigma^2}{2}\right)dt + \sigma dB_t$$

Integrating,

$$\int_0^t d(\ln X_t) = \left(\mu - \frac{\sigma^2}{2}\right)\int_0^t dt + \sigma \int_0^t dB_t$$

$$\ln(X_t/X_0) = \left(\mu - \frac{\sigma^2}{2}\right)t + \sigma(B_t - B_0)$$

$$\Rightarrow \ln X_t = \ln X_0 + \left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t \quad [\text{As } B_0 = 0]$$

Now that we have a formal solution of the SDE, it is analysed for better understanding of the distribution:

As $\ln X_0 + \left(\mu - \frac{\sigma^2}{2}\right)t$ is deterministic given t, and B_t is a Normal distribution given t.

$\therefore$ $\ln X_t$ is of the form $a + bZ$ where $Z \sim N(0, t)$

Thus $\ln(X_t)$ is also a Normal distribution

$$\therefore\ \mathbb{E}[\ln X_t] = \mathbb{E}[\ln X_0 + (\mu - \frac{\sigma^2}{2})t + \sigma B_t]$$

$$= \ln X_0 + (\mu - \frac{\sigma^2}{2})t + \sigma \mathbb{E}[B_t]$$

$$= \ln X_0 + (\mu - \frac{\sigma^2}{2})t$$

$$V(\ln X_t) = V(\ln X_0 + (\mu - \frac{\sigma^2}{2})t + \sigma B_t)$$

As $\ln X_0 + (\mu - \frac{\sigma^2}{2})t$ is deterministic, their variance is 0, thus

$$= V(\sigma B_t) = \sigma^2 V(B_t)$$

$$= \sigma^2 t$$

$$\therefore\ \ln X_t \sim N(\ln X_0 + (\mu - \frac{\sigma^2}{2})t,\ \sigma^2 t)$$

Similarly $\ln (X_t/X_0) \sim N((\mu - \frac{\sigma^2}{2})t,\ \sigma^2 t)$

If the integral was made from $t-1$ to t,

$$\ln (X_t/X_{t-1}) \sim N((\mu - \frac{\sigma^2}{2}),\ \sigma^2)$$

One problem left to tackle is the process of calculation of the mean and the variance of the distribution, which turns out to be fairly simple.

Thus, the mean and variance may obtained as follows:

$$(\mu - \frac{\sigma^2}{2}) = \frac{\sum_{t=1}^{n} \ln (\frac{X_t}{X_{t-1}})}{n} = m \text{ (let)}$$

$$\sigma^2 = \frac{\sum_{t=1}^{n} (\ln (\frac{X_t}{X_{t-1}}) - m)^2}{(n-1)}$$

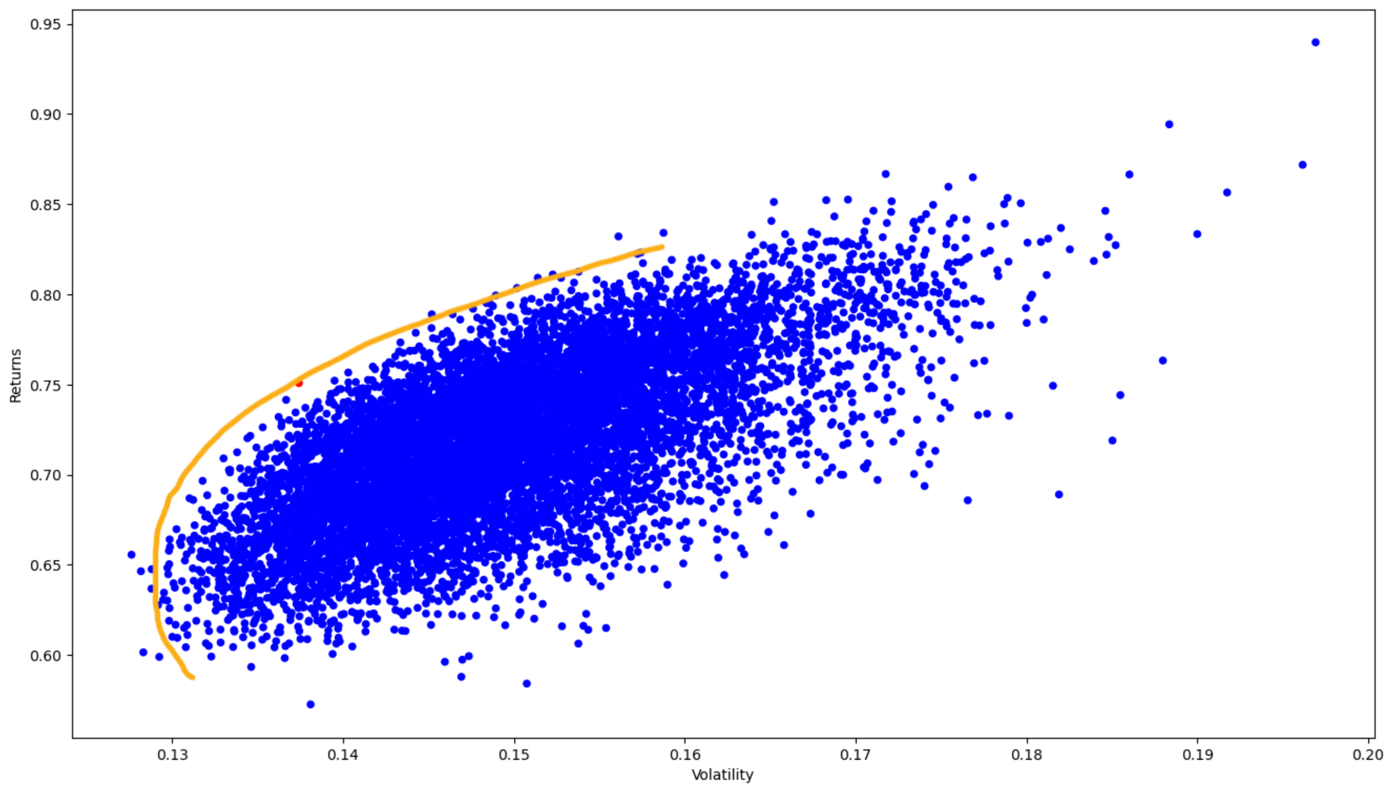
Now that the performance metrics has been decided, our focus is turned to optimizing the portfolio. The next section discusses the process of optimization in details.

Design of the Portfolio Optimization Model

For allotting percentages of the total investment to the different stocks, the Sharpe Ratio has been used. The Sharpe Ratio was developed by economist William F. Sharpe in 1966 to measure risk-adjusted investment returns. It is the ratio of the difference in the rate of returns of the stock and the risk-free rate, to the standard deviation of the portfolio's excess returns.

$$S_a = \frac{(rs - rf)}{\sigma_f}$$

The measure of risk has also to be taken care of. So, random weights are assigned to each stock and a plot between different returns and risk is obtained for different investment scenarios. The investment with the highest Sharpe ratio is chosen. This point lies on the efficient frontier, i.e., the surface that contains the maximum possible returns for each value of risk (volatility).



The efficient frontier is denoted by the yellow line. The weights that yield the maximum possible Sharpe ratio are used as investment percentages. The performance of the portfolio has also been analysed based on recent past performance.

Performance analysis of the Portfolio

2 Greeks are used for this analysis:

- β: It is a measure of the volatility of a stock or portfolio w.r.t the market, in this case the market index S&P 500. This is used to determine the risk in a stock/portfolio in the market. If beta is less than 1, it means the stock or portfolio is less volatile than the market.

$$\beta = \frac{cov(portfolio\ return, market\ return)}{\sigma_{market}}$$

- α: It is the measure of returns of a stock or portfolio w.r.t. the market, which, again, in this case is the market index S&P 500, over the risk-free interest rate.

$$\alpha = r_s - r_f - (\beta * (r_m - r_f))$$

Usage of the Ideas

- The stocks are analysed using the Geometric Brownian motion model and their mean and variances are found out.
- The target is set at 10%+ increase in stock price within a year, so accordingly, the stocks with the highest probability of achieving this, as predicted by the Brownian model, are picked, after checking if they were in business for more than 2 years.
- The Sharpe ratio is used to construct the investment percentages between the stocks in the portfolio.
- The feasibility and performance of the portfolio is analysed by the Greeks β and α.
- The investment is visualized using pie charts.

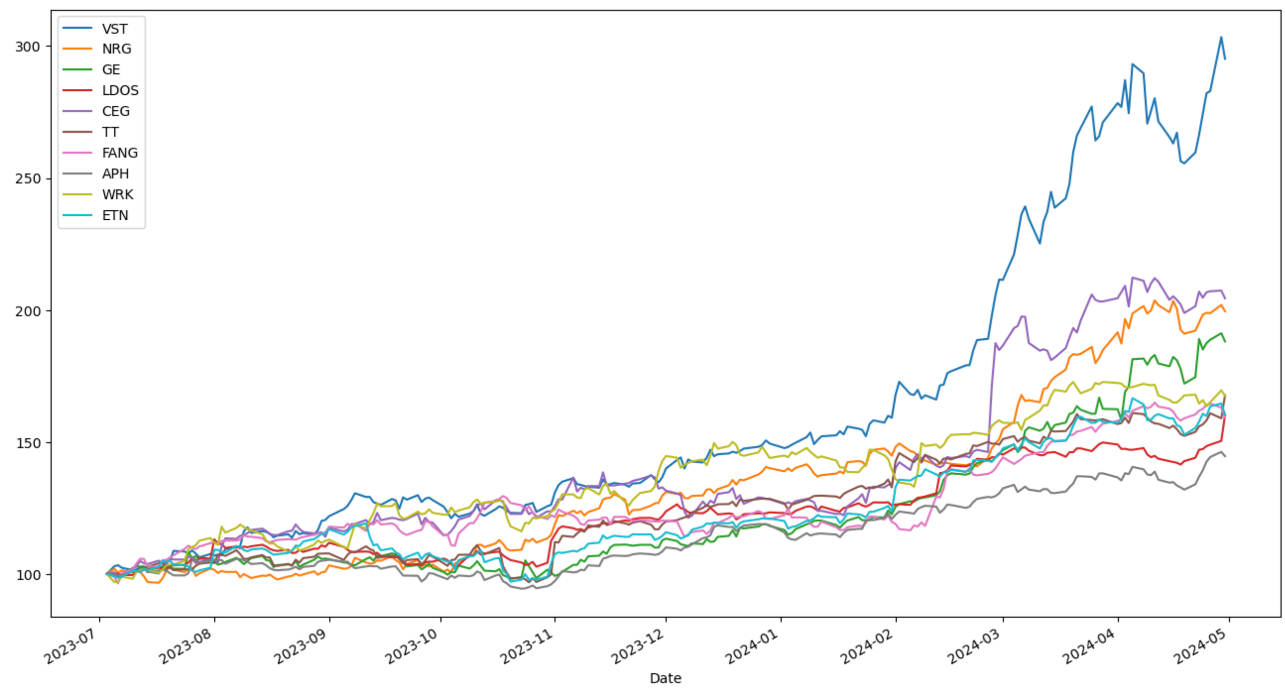
Results

- The top stocks as predicted by the model

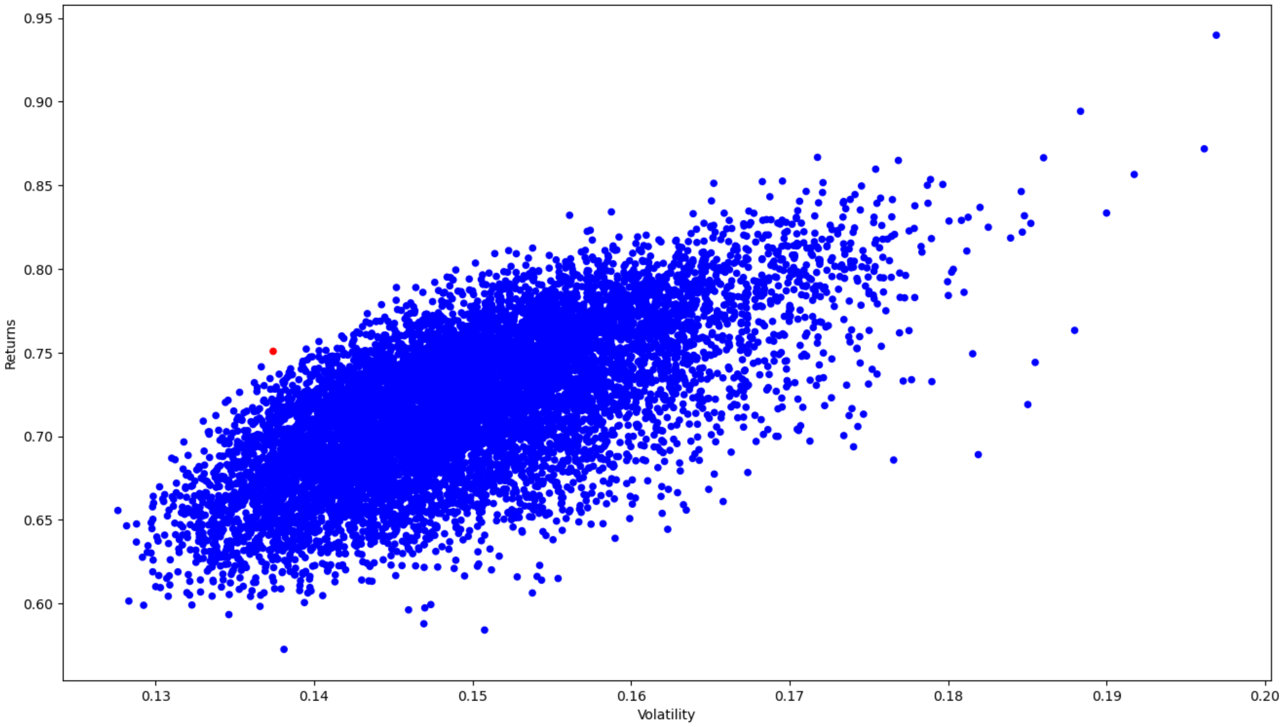
Date	VST	NRG	GE	LDOS	CEG	TT	FANG	APH	WRK	ETN
2023-07-03	25.694530	36.428875	86.037918	87.520782	90.965714	189.299576	125.053558	83.519714	28.461912	197.827286
2023-07-05	26.433165	37.135761	86.331909	87.748444	91.501976	187.883820	123.774132	83.301178	27.578245	196.121445
2023-07-06	26.541494	36.603180	85.314842	87.540581	90.369873	187.923401	120.523392	83.142235	27.626799	194.987503
2023-07-07	26.295284	36.913044	86.029976	87.144653	90.697594	187.309586	125.480042	83.062775	28.160885	195.687592
2023-07-10	26.058922	36.632233	87.907120	87.134758	92.951866	191.675766	127.081711	84.314377	27.937540	200.223312
...	...	...	...	...	...	...	...	...	...	...
2024-04-24	70.459999	72.209396	159.190002	128.899994	186.160004	296.489990	203.226135	116.309998	47.201408	317.839508
2024-04-25	72.480003	72.458015	161.259995	129.880005	188.009995	299.369995	205.067886	119.010002	46.535343	316.154572
2024-04-26	72.699997	72.408287	162.350006	130.360001	188.369995	304.529999	205.721405	120.489998	46.972759	323.333069
2024-04-29	77.959999	73.542000	164.490005	131.619995	188.610001	300.859985	203.840057	122.129997	48.255192	325.536499
2024-04-30	75.839996	72.669998	161.820007	140.220001	185.940002	317.339996	199.156464	120.769997	47.678593	317.311096

209 rows × 10 columns

- The change in their Adjusted Closing Price over the last 10 months

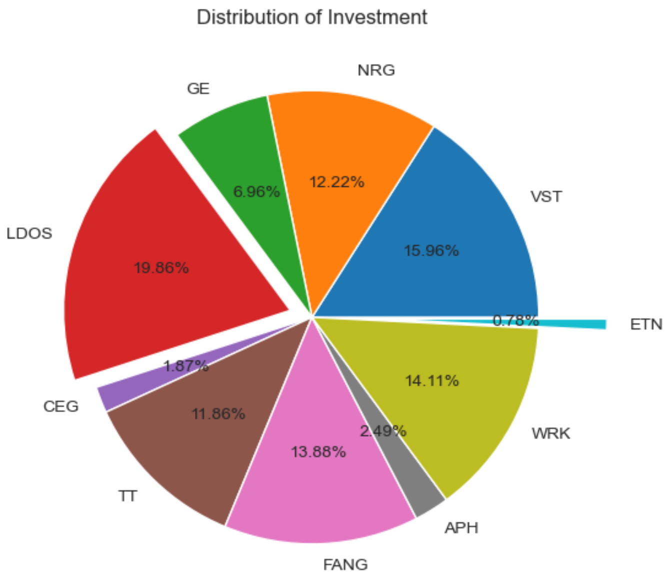


- The visualization of different investment scenarios (the dot coloured red represents the scenario with the highest Sharpe ratio)



- The investment percentages are found and visualised

```
Stock : VST -> 15.96%
Stock : NRG -> 12.22%
Stock : GE -> 6.96%
Stock : LDOS -> 19.86%
Stock : CEG -> 1.87%
Stock : TT -> 11.86%
Stock : FANG -> 13.88%
Stock : APH -> 2.49%
Stock : WRK -> 14.11%
Stock : ETN -> 0.78%
Volatility :  0.13741461394071328
Returns :  0.7507336154589691
```



- The Greeks β and α are found out

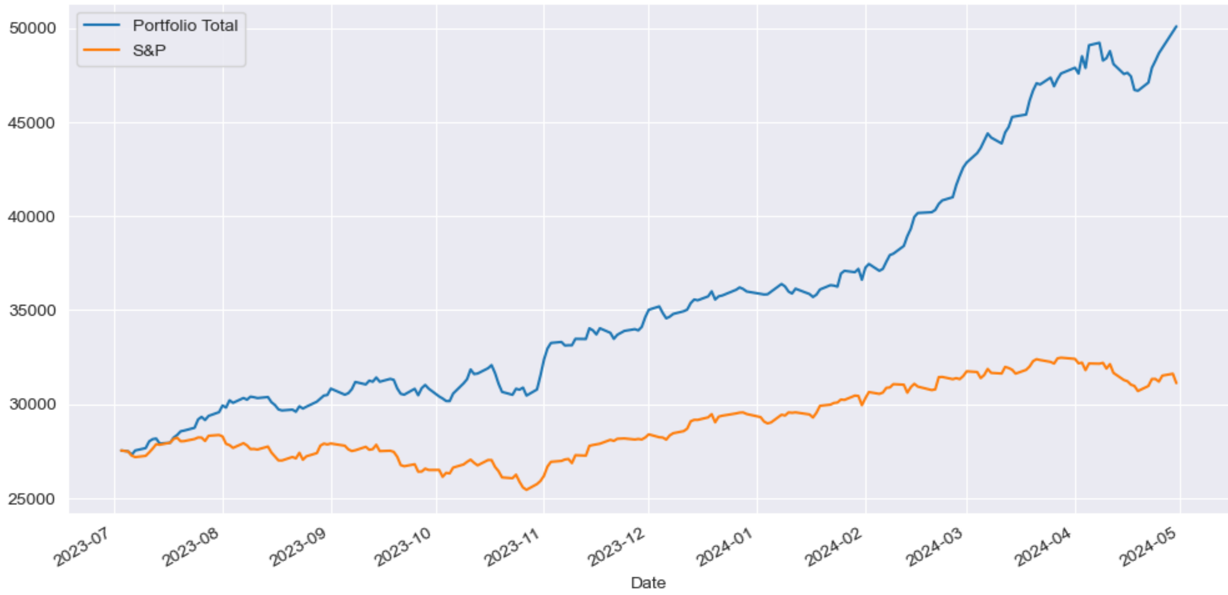
In [32]:
```
#the beta of the portfolio with respect to the market
beta_portfolio = find_portfolio_beta(total_portfolio_df, dt.datetime(2023,7,1), dt.datetime(2024,5,1))
beta_portfolio
```

Out[32]: 0.6852652383177071

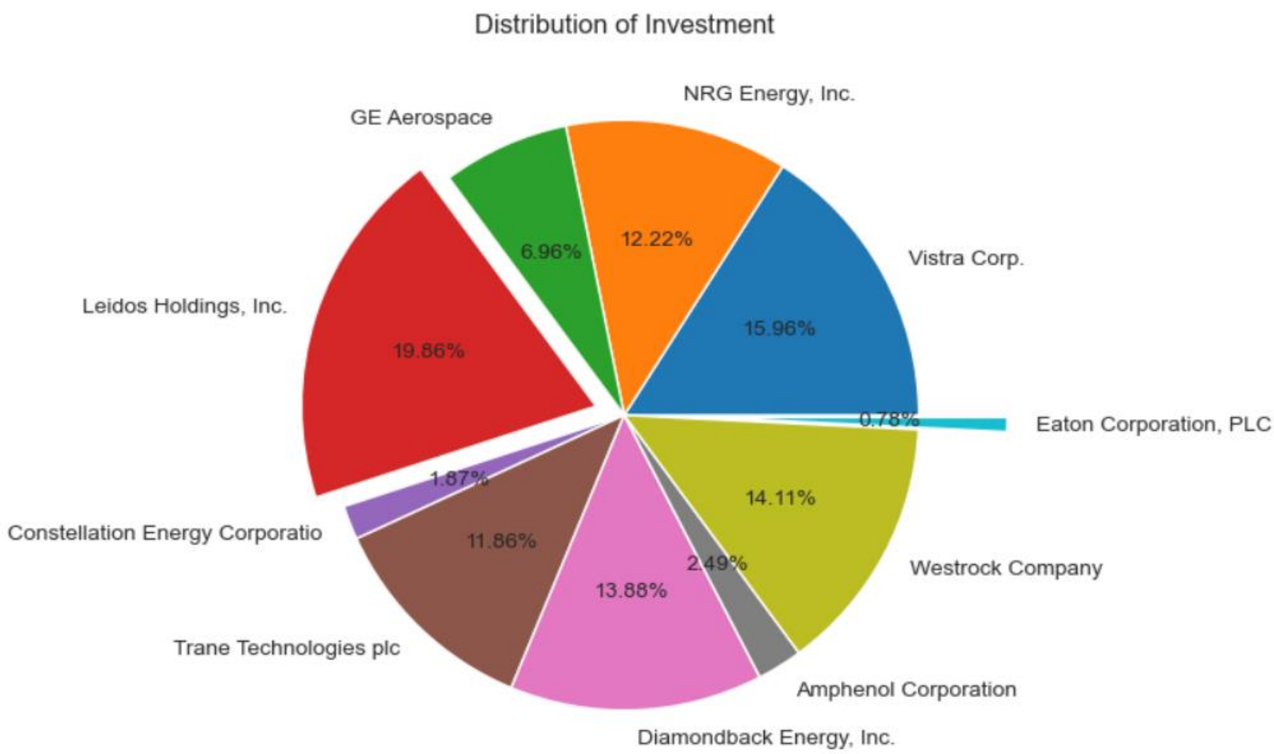
In [33]:
```
#the alpha of the portfolio with respect to the market
risk_free_rate = (0.0442)
port_alpha = portfolio_roi - risk_free_rate - (beta_portfolio * (snp_roi - risk_free_rate))
port_alpha
```

Out[33]: 0.7159283083967999

- The past performance with the market is visualized. It may be observed that the portfolio is correlated to the market and their bearish and bullish periods align.



- The final visualization of the investment



Observations and Conclusions

- From the Greeks, we may reach a conclusion that, as β is between 0 and 1, the portfolio has a lower overall volatility than the market but is positively correlated to the market, which is obviously a positive news. The α being greater than 0 indicates that the volatility adjusted returns of the portfolio is greater than the market index, which again is positive. So, by these parameters, we have achieved a portfolio with less volatility and higher returns than the market, as per historical behaviours.

- The correlation between the different stocks among the top 10 stocks have been observed:

	VST	NRG	GE	LDOS	CEG	TT	FANG	APH	WRK	ETN
VST	1.000000	0.559243	0.258634	0.168479	0.471989	0.249623	0.122520	0.300681	0.044967	0.340161
NRG	0.559243	1.000000	0.327873	0.173011	0.449188	0.270069	0.240820	0.267228	0.139022	0.348039
GE	0.258634	0.327873	1.000000	0.104241	0.299322	0.303764	0.141066	0.478579	0.091855	0.433586
LDOS	0.168479	0.173011	0.104241	1.000000	0.089536	0.253980	0.154247	0.200899	0.072805	0.346460
CEG	0.471989	0.449188	0.299322	0.089536	1.000000	0.240790	0.152356	0.235415	0.133187	0.334170
TT	0.249623	0.270069	0.303764	0.253980	0.240790	1.000000	-0.092204	0.481867	0.054361	0.578825
FANG	0.122520	0.240820	0.141066	0.154247	0.152356	-0.092204	1.000000	0.155374	0.283811	0.088472
APH	0.300681	0.267228	0.478579	0.200899	0.235415	0.481867	0.155374	1.000000	0.229734	0.528567
WRK	0.044967	0.139022	0.091855	0.072805	0.133187	0.054361	0.283811	0.229734	1.000000	0.079187
ETN	0.340161	0.348039	0.433586	0.346460	0.334170	0.578825	0.088472	0.528567	0.079187	1.000000

  It may be observed that none of the stocks have correlation more than 0.6, mostly having values near 0.2. This indicates that the stocks are not very correlated with each other. This observation helps suggests that our portfolio is diversified and any crisis with one company would not have much effect on the whole portfolio.

- The percentages of the total investment were found out, but due to discretisation of shares, we have to buy a specific number of shares of each stock. The number of shares to be bought, with a total investment of US$ 50,000 are as follows:

```
VST -> 105
NRG -> 84
GE -> 22
LDOS -> 71
CEG -> 5
TT -> 19
FANG -> 35
APH -> 10
WRK -> 148
ETN -> 1
```

- With the current 10-year US bonds rate, i.e., the risk-free rate at 4.42%, 10% or more profit in a year is a realistic target in present market scenario. This is the basis of our performance metrics. The stocks which are having the highest probability of achieving this are chosen. It might be noted that major companies like GOOGL, MSFT or AAPL are not present in the list. This is because their growth is gradual and suitable for long term investments (at least in the present scenario), but for medium or short-term investments, other companies are preferred.

References

- The list of tickers of the stocks in S&P 500 index: Kaggle_Datasets
- Stock Data Fetching: YFinance
- Data Visualization: Matplotlib, Seaborn
- More on Geometric Brownian Motion: ScienceDirect_Article
- More on Sharpe Ratio and Modern Portfolio Theory: CFI_Article
- More on market α and β: Time_Magazine_Article