GROUP WORK PROJECT # 1 MScFE 652: PORTFOLIO MANAGEMENT
Group Number: 8244

FULL LEGAL NAME	LOCATION (COUNTRY)	EMAIL ADDRESS	MARK X FOR ANY NON-CONTRIBUTING MEMBER
Shivansh Kumar	India	shivansh.business23@gmail.com	
Danish Rizwan	Pakistan	danishrizwan.dr@gmail.com	
Bharat Swami	India	bharatswami1299@gmail.com	

Statement of integrity: By typing the names of all group members in the text boxes below, you confirm that the assignment submitted is original work produced by the group (excluding any non-contributing members identified with an "X" above).	
Team member 1	**Shivansh Kumar**
Team member 2	**Danish Rizwan**
Team member 3	**Bharat Swami**

Use the box below to explain any attempts to reach out to a non-contributing member. Type (N/A) if all members contributed. **Note:** You may be required to provide proof of your outreach to non-contributing members upon request.
(N/A)

Part 1

This report presents an analysis of the optimal mean-variance portfolio for a portfolio manager considering ten securities: TSLA, WMT, BAC, GS, LLY, MRK, GOOG, META, AAPL, and XOM. The period for estimating mean returns and the variance-covariance matrix is January 1, 2022, to December 31, 2024, using either **daily returns** data.

Daily returns are calculated using a log of daily price differences. Annual expected returns are calculated using the mean of daily return and multiplying it by 252 business day. The average expected return of all assets is in below Fig 1, which shows that LLY has the highest expected return while TSLA has the lowest expected return.

Figure 1: Expected Returns of Assets

Also from Fig 2, which is the heatmap of Covariance matrix of the assets, we see that TSLA has the highest variance followed by META, while all other assets have low (less than 0.1) and mostly positive variance/covariance.

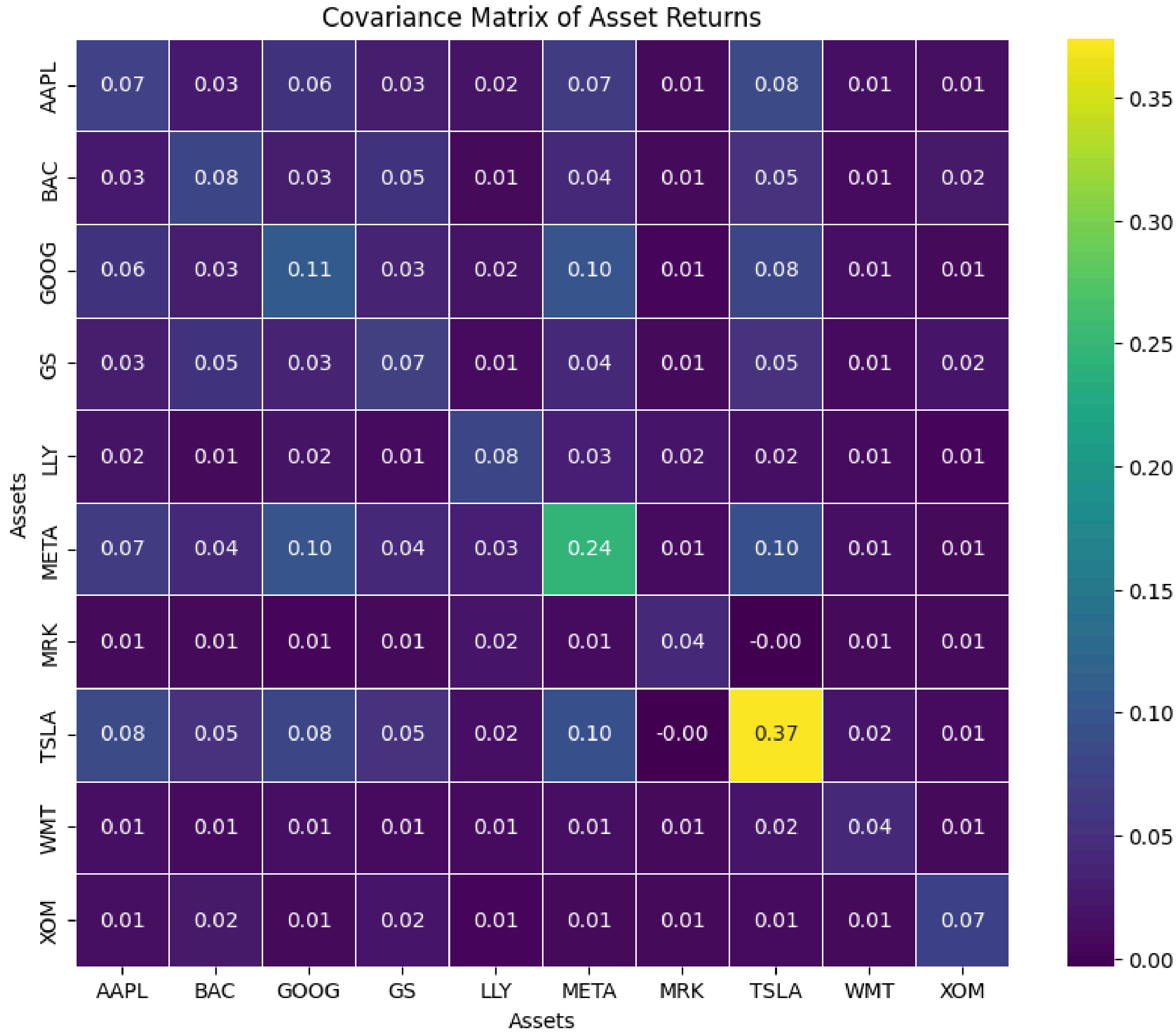

Figure 2 - Covariance Matrix of Asset Returns

Step 1: Determining the Optimal Mean-Variance Portfolio

The optimal mean-variance portfolio is computed under the following constraints:

- The sum of portfolio weights must equal 1.
- No short-selling is allowed (weights must be non-negative).

Using historical return data, we compute the expected return vector and the variance-covariance matrix. The optimization is solved using CVXPY library.

The Optimal portfolio weights are as below

1. TSLA: -0.0
2. WMT: 0.0
3. BAC: -0.0
4. GS: 0.0
5. LLY: 1.0
6. MRK: -0.0
7. GOOG: -0.0
8. META: 0.0
9. AAPL: -0.0
10. XOM: 0.0

From above weights, we notice that the optimal portfolio consists entirely of **LLY** with a weight of **1.0**, while all other assets have weights of **0.0**. This suggests that, based on the mean-variance optimization, LLY provides the best trade-off between risk and return among the ten securities, given the constraints.

The expected return of the optimal portfolio is **35.99%**, since only LLY is included, this return is entirely dependent on LLY's estimated historical performance during the period

The portfolio's standard deviation is **28.52%**, reflecting the risk of investing solely in LLY. Compared to other assets in the portfolio, this is close to the standard deviation of all other assets except META and TSLA.

The **Sharpe Ratio of 1.26** shows that the portfolio provides a good return per risk, however, **it lacks diversification**, making it highly dependent on LLY's future performance.

Step 2: Incorporating the Maximum 15% Allocation Constraint

In this step, we implemented additional constraint thaat no asset can have a weight greater than 15% (0.15). The optimization process is re-run while adjusting the portfolio composition.

The Optimal portfolio weights are as below:

1. TSLA: 0.1
2. WMT: -0.0
3. BAC: -0.0
4. GS: 0.15
5. LLY: 0.15
6. MRK: 0.15
7. GOOG: 0.15
8. META: -0.0
9. AAPL: 0.15
10. XOM: 0.15

The portfolio is now more evenly distributed among multiple assets, with six assets (GS, LLY, MRK, GOOG, AAPL, XOM) each receiving the maximum allowable 15% allocation. TSLA is included at 10%, while WMT, BAC, and META are excluded.

The **expected return is 19.85%,** which is lower than the 35.99% in Step 1. This is because the optimizer is now restricted from putting all capital into the highest-return asset (LLY) and must spread investments across other assets.

The **portfolio's standard deviation is 16.48%**, which is lower than the 28.52% in Step 1. This indicates that diversification effectively reduces risk, as expected in a mean-variance optimized portfolio.

The **Sharpe Ratio is 1.20,** slightly lower than the 1.26 from Step 1. The decrease in the Sharpe Ratio suggests that, while risk has been reduced, the return has also dropped enough that the trade-off is slightly less efficient compared to the Step 1 portfolio.

Table 1 - Difference between Step 1 and 2

	Step 1 (LLY-only)	**Step 2 (Diversified, Max 15%)**
Portfolio Composition	100% LLY	Diversified across 7 stocks
Expected Return	35.99%	19.85%
Standard Deviation	28.52%	16.48%
Sharpe Ratio	1.26	1.2

Step 3

To analyze the impact of the maximum allocation constraint, we plot the efficient frontiers for both scenarios:

- **Unconstrained (except for no short-selling):** This portfolio has greater flexibility in asset allocation, potentially leading to higher expected returns.
- **Constrained (max 15% allocation):** This portfolio has a more diversified composition but might exhibit slightly lower returns due to restricted allocations.

Main Observation: The constrained efficient frontier is slightly lower than the unconstrained frontier, indicating that limiting asset weights reduces potential returns but may offer better diversification.

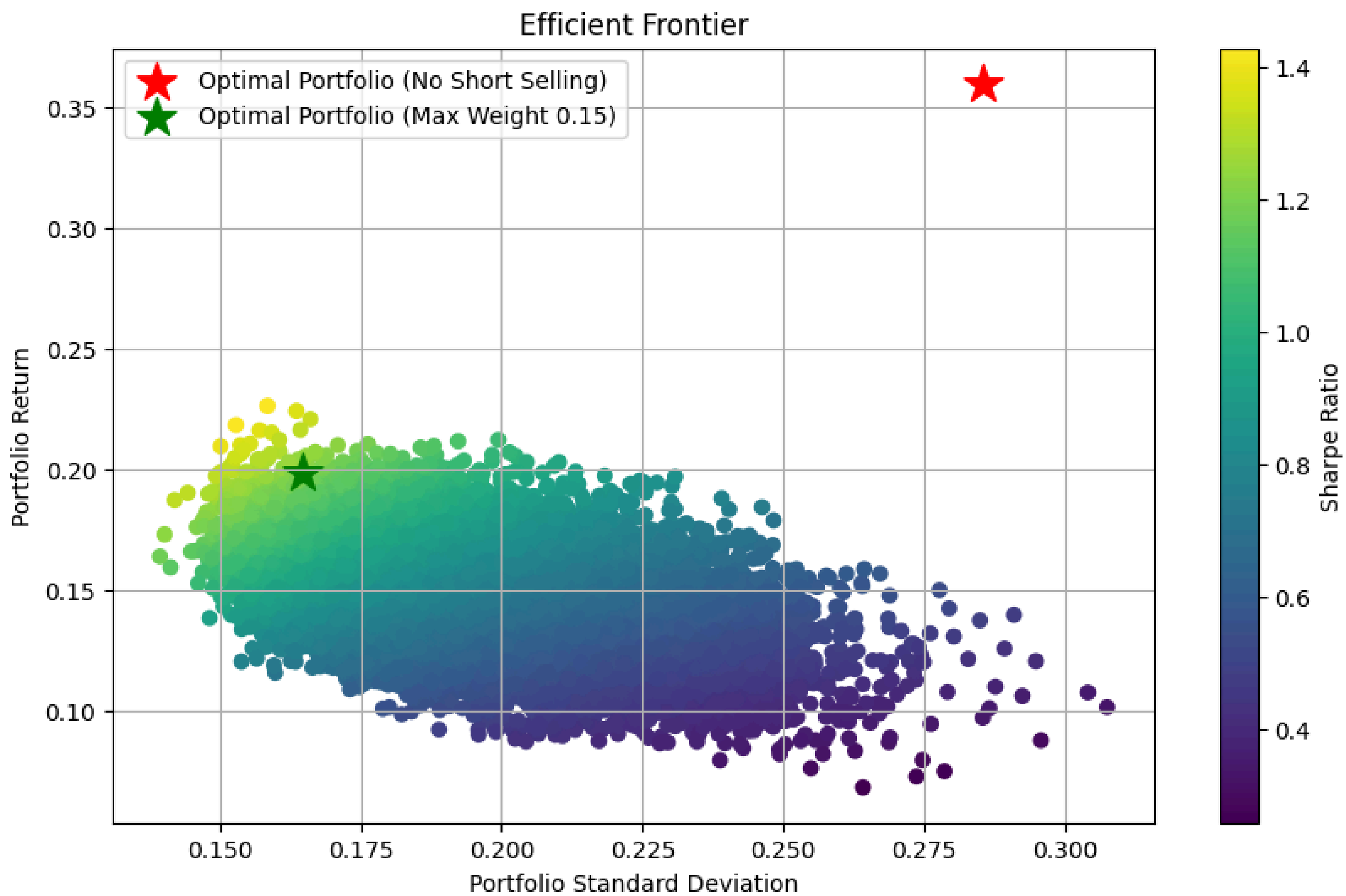

Figure 3 - Efficient Frontier and Optimal Portfolios

Step 4 . Out-of-Sample Performance (January 2025)

To evaluate real-world effectiveness, we assess the performance of both portfolios using returns from January 1-31, 2025 and optimal weights calculated in step 1 and 2.

Portfolio 1 (Step 1: LLY-Only Portfolio) Out-of-Sample Results

- Higher Return of 55.11% : The LLY-heavy portfolio had a strong performance in January 2025, significantly exceeding the in-sample expected return of 35.99%
- Higher Volatility of 39.33%: The standard deviation is much higher than in-sample (28.52%), due to price changes.
- Sharpe Ratio of 1.40: The Sharpe Ratio is only slightly better than in-sample (1.26), meaning the higher return came at the cost of higher risk.

Portfolio 2 (Step 2: Diversified Portfolio) Out-of-Sample Results

- Extremely High Return of 68.86% : The diversified portfolio outperformed the concentrated LLY portfolio, despite having a more balanced allocation.
- Much Lower Volatility of 13.53%: The diversified portfolio saw a dramatic drop in standard deviation compared to in-sample (16.48%) and far lower than Portfolio 1 (39.33%). This suggests that spreading risk across multiple assets effectively reduced volatility.
- Sharpe Ratio of 5.09: The Sharpe Ratio skyrocketed, indicating that Portfolio 2 achieved significantly higher returns with much lower risk. A Sharpe Ratio above 3.0 is considered exceptional.

From above, we can say that the constrained, diversified portfolio delivered superior risk-adjusted returns, making it the better long-term investment choice.

Part 2

For this part we will analyze the portfolio found in Part 1, Step 2 given constraints, we will try to explain the exposure of this portfolio to each of Fama-French 5 factors: Market Risk Premium, SMB, HML, CMA, RMW. Essentially, we will estimate the following linear regression model:

$$E[R_p] - R_f = \alpha + \beta_1(R_m - R_f) + \beta_2 SMB + \beta_3 HML + \beta_4 RMV + \beta_5 CMA + \epsilon$$

Here,

- $E[R_P] - R_f$: Excess return of the portfolio over the risk-free rate
- α: Intercept (constant term)
- β's: Coefficients for the factors
- $R_m - R_f$: Market risk premium (Mkt-RF in Fama-French data)
- SMB: Small Minus Big
- HML: High Minus Low
- RMW: Robust Minus Weak
- CMA: Conservative Minus Aggressive
- ϵ: Residuals

Step 5:

Now we will understand the role of each of the 5 factors in the above model, but let’s first understand what the is Fama-french 5 factor is, Researchers have always been intrigued by return and risk which has led to financial models quantifying risk and estimating the expected returns. (Chen) The Fama French five-factor model is nothing but an improved version of the Fama French three-factor model, it is build on 5 factors that captures different aspects of the relationships between risk and returns, it includes Market Risk Premium, SMB, HML, CMA, RMW, let’s understand one by one what each one of these signifies (Chen):

- **Market Risk Premium:** it is the excess return of the market portfolio over the risk-free rate, it basically captures the exposure to systematic market risk. (For our analysis we are using U.S 10-Yr treasury note rate as risk free rate.)
- **SMB(Small minus Big):** It measures the performance of small-cap stocks relative to large-cap stocks. Positive exposure indicates a tilt toward smaller firms.
- **HML(High minus Low):** It calculates High book-to-market (value) stocks minus low book-to-market (growth) stocks. In this Positive beta implies value exposure.
- **CMA(Conservative mins Aggressive):** It calculates Firms with conservative (low) investment strategies minus aggressive (high) investment. In this Positive beta signals exposure to firms that invest conservatively.
- **RMW(Robust minus Weak):** It calculates Robust (high) profitability firms minus weak (low) profitability firms. In this Positive loading indicates exposure to profitable companies.

In our analysis we are using pandas datareader to get "F-F_Research_Data_5_Factors_2x3_daily" data, which includes daily Fama French five factor data accordingly by given time duration as you can see in Figure 4.

	Mkt-RF	SMB	HML	RMW	CMA
Date					
2022-01-03	0.0073	0.0042	0.0080	-0.0117	0.0059
2022-01-04	-0.0029	-0.0008	0.0364	0.0063	0.0147
2022-01-05	-0.0228	-0.0094	0.0259	0.0139	0.0125
2022-01-06	0.0000	0.0038	0.0175	-0.0056	-0.0001
2022-01-07	-0.0048	-0.0096	0.0202	-0.0012	0.0085

Figure 4: Fama French Five-Factor Data

After that we plotted the Correlation heatmap matric of the Fama French five factor model data. You can see the same in Figure 5.

	Mkt-RF	SMB	HML	RMW	CMA
Mkt-RF	1.000000	0.218657	-0.391141	-0.372612	-0.471425
SMB	0.218657	1.000000	0.134794	-0.430618	-0.086428
HML	-0.391141	0.134794	1.000000	0.356232	0.646585
RMW	-0.372612	-0.430618	0.356232	1.000000	0.295510
CMA	-0.471425	-0.086428	0.646585	0.295510	1.000000
RF	0.067500	-0.010683	-0.059453	0.002866	-0.093562

Figure 5: Correlation matrix Fama-French five-factor model

The correlation matrix for the FF model signifies many insightful relationships, the Mkt-RF (market risk premium) has strong negative correlations with HML (-0.39), RMW (-0.37), and CMA (-0.47), indicating that value, profitability, and conservative investment factors tend to underperform during high market returns. HML and CMA are strongly positively correlated (0.65), reflecting the overlap between value and conservative investment strategies. RMW (profitability) shows moderate positive correlations with HML (0.36) and CMA (0.30), but negative correlations with Mkt-RF and SMB (-0.43), suggesting profitability factors struggle in bull markets or when small-cap stocks thrive.

Step 6 :

In this section, we will discuss our Regression methods as well as as data split method, but before that let's talk about Data preparation and MVO for the portfolio.

We have used Yahoo Finance to download daily Adj.Close Data for TSLA, WMT, BAC, GS, LLY, MRK, GOOG, META, AAPL, XOM tickers from Jan 1, 2022 to Dec 31, 2024. You can see it in the given plot below in Figure 6.

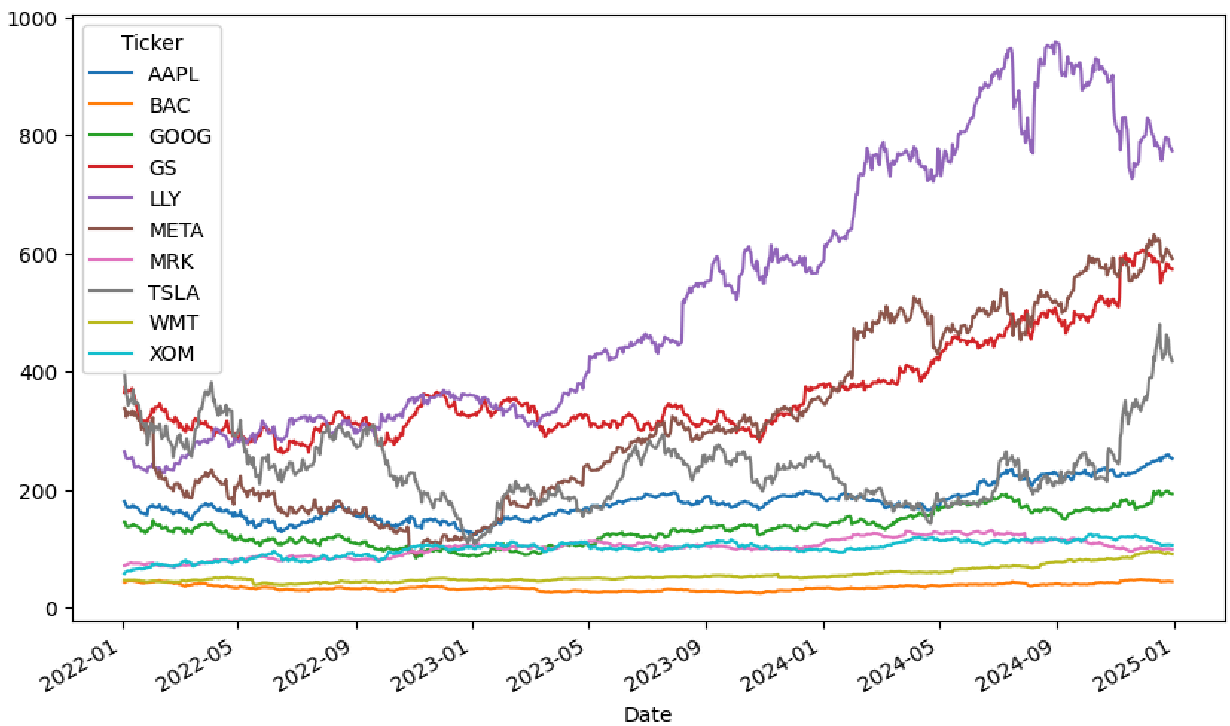

Figure 6: Ticker Time series Plot

After that we have used the PyPortfolioOpt Python package ("PyPortfolioOpt") for our MVO of our portfolio to find and efficient frontier using given constraints in Part 1 Step 2.
Using the package we found the efficient frontier of our portfolio, we have extracted the clean weight of how much should be allocated to per assets, you can see this in Figure 7 below.

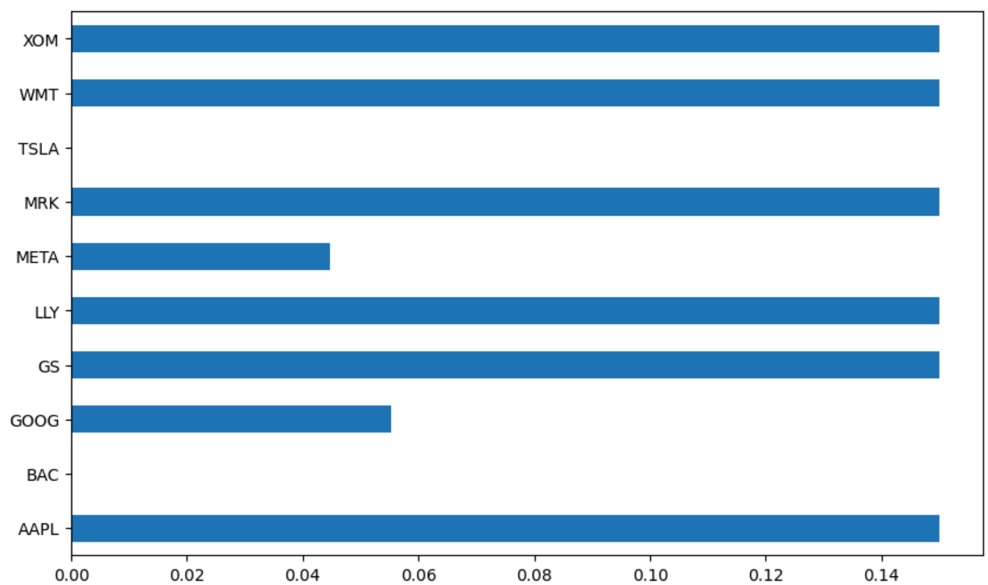

Figure 7: Portfolio weights

We can also see the plot of our Efficient frontier portfolio in Figure 8 below.

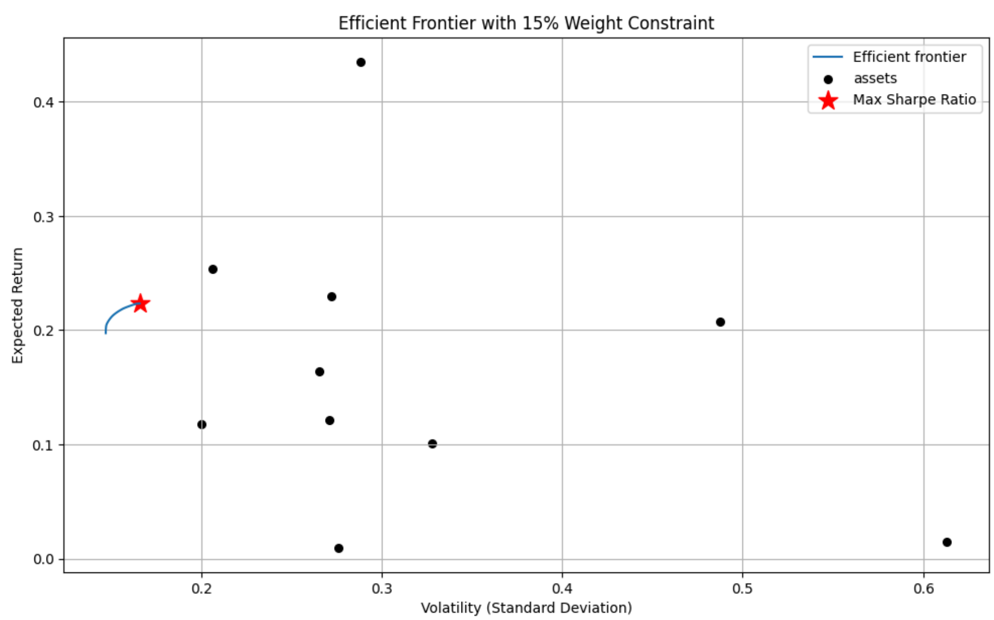

Figure 8: Efficient frontier Portfolio

Now we have to find the Excess market returns, which is $R_p - R_f$, to find the risk-free rate we have used U.S 10 year treasury note annual rate i.e 4.53% at the time of analysis we then calculated the Portfolio access return and merged this data to the Fama french data to build a final data frame for our regression analysis. You can see the final Data frame in Figure 9 below.

Date	Portfolio_Excess	Mkt-RF	SMB	HML	RMW	CMA
2022-01-04	0.002466	-0.0029	-0.0008	0.0364	0.0063	0.0147
2022-01-05	-0.007582	-0.0228	-0.0094	0.0259	0.0139	0.0125
2022-01-06	0.000039	0.0000	0.0038	0.0175	-0.0056	-0.0001
2022-01-07	0.005439	-0.0048	-0.0096	0.0202	-0.0012	0.0085
2022-01-10	0.003637	-0.0015	-0.0030	-0.0028	-0.0005	-0.0037

Figure 9 : Final DataFrame

Further, we have used two types of regression: OLS and one Robust model, i.e., Huber regression. However, before discussing the model and the results more, let's discuss how we have divided the train-test data.

Our goal with the data split was to maintain the temporal order to avoid look-ahead bias in our data, for this we have simply split the data into 2 parts Training Data: Jan 2022–Sept 2023 and Testing Data: Oct 2023–Dec 2024.

Step 7:

So Finally we are using 2 kinds of models OLS and Huber, OLS estimates the relationship between the dependent variable and one or more independent variables by minimizing the sum of squared residuals, while the Huber regression is a robust regression technique designed to reduce the influence of outliers. It combines the squared loss (used in OLS) for small residuals and the absolute loss for large residuals.

Now we will understand the result of these regressions and understand which factors and how they help explain the exposure of the portfolio returns on our portfolio. We will see the result of our models applied to training data.

Table 2: OLS & Huber model Beta's comparison.

Factors	OLS Regression	OLS P > | value |	Huber Regressiom
Alpha	0.0006	0.005	0.0003
Mkt-RF	0.8519	0.000	0.8571
SMB	-0.2112	0.000	-0.1936
HML	0.0333	0.371	0.0439
RMW	0.0263	0.505	0.0328
CMA	0.3593	0.000	0.3457

The comparison of OLS and Huber regression models provides insight that portfolio excess returns are primarily driven by market exposure (Beta 0.85, $p < 0.001$) and a significant tilt towards large-cap stocks (negative SMB, $p < 0.001$). The CMA factor is also significant ($p < 0.001$), indicating a preference for firms with conservative investment policies, while the HML (value, $p = 0.371$) and RMW (profitability, $p = 0.505$) factors remain weak indicators. Huber regression, by down-weighting outliers, provides more robust estimates with slightly lower alpha (0.0003 vs. 0.0006 in OLS) and minor adjustments in factor loadings (e.g., market beta at 0.8571 vs. 0.8519). The reduction in SMB magnitude (-0.1936 vs. -0.2112) and slight variations in other factor coefficients suggest a more resilient model against extreme observations, reinforcing systematic risk factor influences.

Now we will use these models to predict on testing data and calculate the performance metrics like MAE and R-squared and understand which of our models have performed well.

Table 3: Performance Metrics Comparison OLS vs Huber

Performance Metrics	OLS Regression	Huber Regression
MAE	0.003746	0.003718
R-Squared	0.5265	0.5334

The performance metrics further support our evidence of the robustness of the Huber model, which achieves a lower mean absolute error (MAE: 0.003718 vs. 0.003746 in OLS) and a slightly higher R-squared value (0.5334 vs. 0.5265). This indicates that Huber regression provides better predictive accuracy and a stronger overall fit to the data by mitigating the influence of outliers. Consequently, it offers a more reliable assessment of systematic risk exposure, making it a preferable choice for modeling financial returns in the presence of extreme observations.

Part 3

EDA

In this part we are going to select 5 assets from our 10 assets list and see if we can produce the same result as direct optimization by running simulations. We are going to run a number of simulations to see how much is sufficient to come closer to an optimized result.

First we have to decide which assets to pick from the list. To do such we are going to generate the expected returns and covariance matrix for our selected time period on daily logarithmic returns of these assets. As we can see all assets are not that correlated , and no one is negatively related.

Some more insights from expected returns and covariance matrix-

- TSLA and META have very high variance as compared to others.
- WMT and MRK have very low variance.
- LLY has the highest expected returns among the list.
- TSLA and BAC have lowest expected returns.

So make our research as diverse as possible we are going to select below assets from the list-

- TSLA
- WMT
- LLY
- BAC
- META

Step 8

In this step we are going to use data for the above assets. We are also going to work under some constraints which are-

- Weights have to be sum equal to 1.
- There is no short selling which implies we can't use negative weights in our portfolio.

a) Optimization approach-

In this we are going to use the python module **cvxpy** which uses optimization formulas for the Quadratic Programming we are using.

After running the code in shared notebook we get the following weights for assets-

Table 4: Asset Weights for optimization solution under step 8 constraints

Asset Ticker	Weights
TSLA	0.0
WMT	0.0
LLY	1.0
BAC	0.0
META	0.0

And the corresponding Expected Returns and Volatility are-

- Expected return optimal portfolio: **36.12%**
- Expected std. deviation optimal portfolio: **28.54%**

Reason for all assets weights being zero and all concentrated to LLY-

As we can see the expected returns from the LLY are the highest among all the assets and the volatility is in lowest range with WMT and BAC which makes this asset perfect for buying. Since we are using the mean-variance optimization problem, these are expected solutions.

Now, we are going to run the monte carlo simulation to see whether we can replicate the results from optimization.

b) Monte-Carlo simulations

As discussed above we are going to run monte-carlo simulation with different number of simulation to how much simulation we need to run to produce similar results

Table 5: Asset Weights for Monte-Carlo simulations solution under step 8 constraints

Number of simulations	Results			
	Weights	Expected Returns	Expected Volatility	Time taken (sec)
1000	TSLA: 0.0734 WMT: 0.2877 LLY: 0.5607 BAC: 0.0108 META: 0.0674	28.14%	20.47%	0.0
10000	TSLA: 0.0288 WMT: 0.2959 LLY: 0.6591 BAC: 0.0106 META: 0.0055	30.65%	21.18%	
100000	TSLA: 0.0029 WMT: 0.2940 LLY: 0.6840 BAC: 0.0144 META: 0.0047	31.45%	21.56%	11
500000	TSLA: 0.0095 WMT: 0.2258 LLY: 0.7344 BAC: 0.0250 META: 0.0054	31.77%	22.49%	57
1000000	TSLA: 0.0098 WMT: 0.1142 LLY: 0.8054 BAC: 0.0033 META: 0.0673	32.96%	24.61%	109
3000000	TSLA: 0.0041 WMT: 0.2133 LLY: 0.7733 BAC: 0.0083 META: 0.0010	32.78%	23.31%	320
5000000	TSLA: 0.0142 WMT: 0.0314 LLY: 0.9354 BAC: 0.0037 META: 0.0153	34.81%	27.11%	528

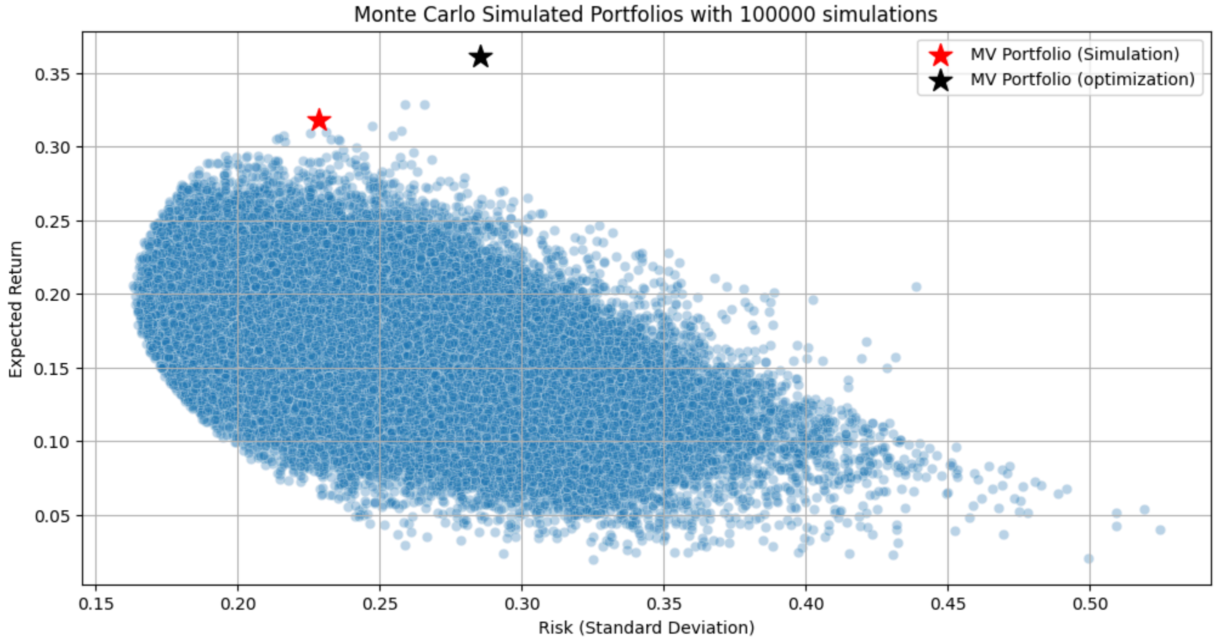

Figure 6: Portfolio Frontier with Monte-Carlo Simulation with maximum weight allowed to 100% of portfolio

As we can see from the results table we are getting closer to an optimized solution but as we increase the number of simulations we are using more time and resources which is not good. The main reason why it is very difficult to reach the optimized solution is because all the assets have weight equal to zero except LLY. And we know in monte-carlo solution we run a random number generator which generates 5 numbers (in our case) and then we normalize them with a sum equal to one. So chances for particular weights with all weights equal to 0 except LLY are very slim because we are normalizing. But except that we are getting closer to optimized results. And all other results still have LLY asset maximum weights and others near zero.

After running above simulations we can say **100000** is a sufficient number of simulations for producing a good approximation if we have time and power as constraints.

Step 9

In this step we are again going to run the optimizations and simulations on our problem with same assets but this time we have some more constraints added, which are-

- Weights have to be sum equal to 1.
- There is no short selling which implies we can't use negative weights in our portfolio.
- No asset can be more than 30% of total portfolio weight, which means all the assets weights are in range 0 to 0.3.

a) Optimization Solution

Below are results from running the optimization, same as before-

Table 6: Asset Weights for optimization solution under step 9 constraints

Asset Ticker	Weights
TSLA	0.0
WMT	0.3
LLY	0.3
BAC	0.1
META	0.3

And the Expected Return and Risk are-

- Expected return optimal portfolio: 23.36%
- Expected std. deviation optimal portfolio: 21.81%

Again we can see the assets have distributed the weights among themselves in a way to produce the maximum reuters with min risk. Since we are not allowing the negative returns, weights are distributed in above fashion. Now maximum weights are given assets with maximum returns.

b) Monte-Carlo simulations

Now we are again going to try producing the optimized result using the monte-carlo simulations. Below are the results are running monte-carlo simulations for different number of simulations.

Table 7: Asset Weights for Monte-Carlo simulations solution under step 9 constraints

Number of simulations	Results			
	Weights	Expected Returns	Expected Volatility	Time taken (sec)
1000	TSLA: 0.0176 WMT: 0.2966 LLY: 0.2914 BAC: 0.1425 META: 0.2518	22.13%	20.55%	0.0
10000	TSLA: 0.0071 WMT: 0.2898 LLY: 0.2939 BAC: 0.1228 META: 0.2864	22.68%	21.45%	2
100000	TSLA: 0.0053 WMT: 0.2936 LLY: 0.2995 BAC: 0.1044 META: 0.2972	23.16%	21.78%	24
500000	TSLA: 0.0130 WMT: 0.3000 LLY: 0.2994 BAC: 0.1157 META: 0.2718	22.84%	21.10%	111

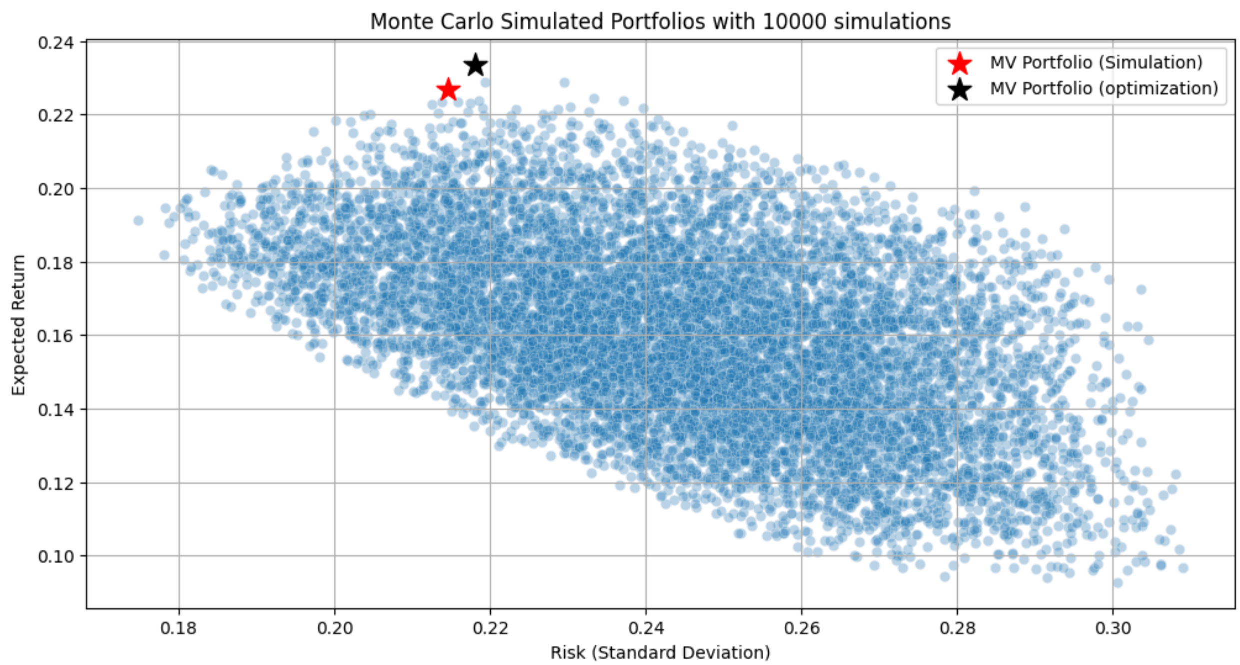

Figure 7: Portfolio Frontier with Monte-Carlo Simulation with maximum weight allowed to 30% of portfolio with 100000 simulations

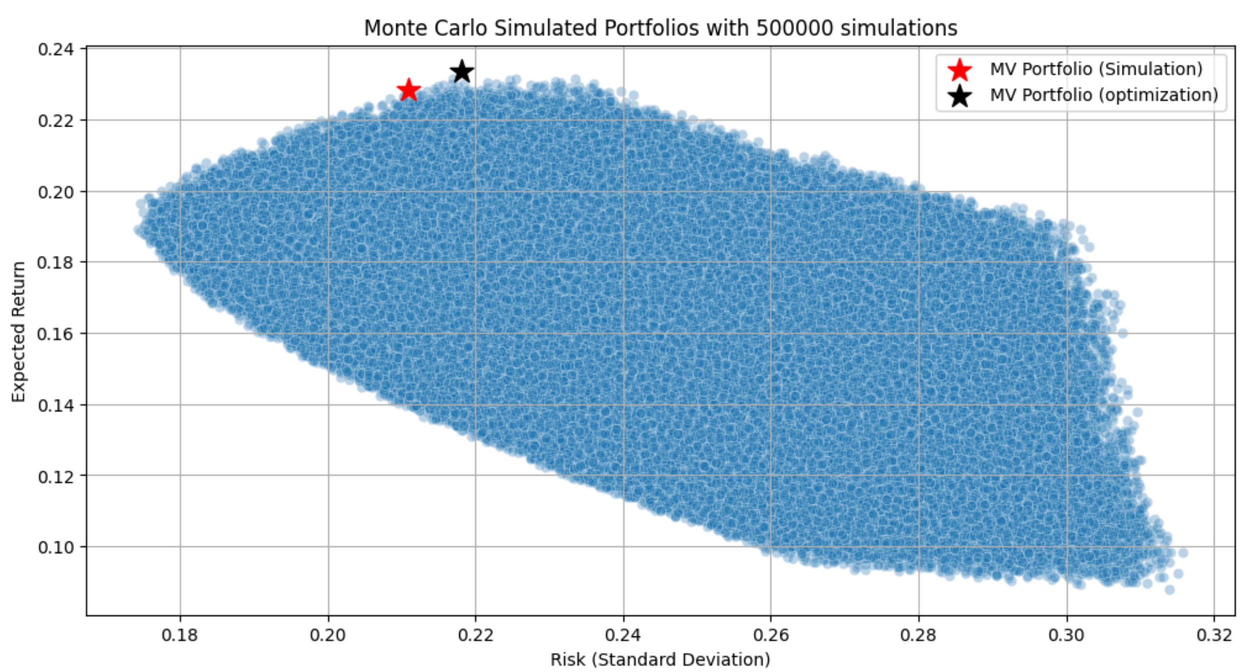

Figure 8: Portfolio Frontier with Monte-Carlo Simulation with maximum weight allowed to 30% of portfolio with 500000 simulations

As we can see from the results table we can approximately produce the optimized solution using the monte-carlo simulations. The number of simulations which can produce great results with time and power constraints is again **100000** which is good and runs under 1 minute of time.

Reason of not getting the optimized solution via simulation is the same as before if it is very unlikely to get the weight as zero while we are normalizing the random generated weights. It is not impossible but it is very unlikely to produce results in some simulations.

References

1. Chen, Daniel. “Fama French Five Factors.” *quantconnect*, August 2019, https://www.quantconnect.com/research/15262/fama-french-five-factors/p1.
2. “PyPortfolioOpt.” *Installation — PyPortfolioOpt 1.5.4 documentation*, https://pyportfolioopt.readthedocs.io/. Accessed 2 February 2025.