Risk and Investment Strategy Optimisation for Mutual Funds Portfolio

Alen Šahinpašić and Mykyta Panchenko

March 2024

1 Abstract

This paper presents a comprehensive analysis of mutual funds based on a dataset obtained from Yahoo Finance. The analysis covers various aspects of mutual funds, including fund size, investment types, asset allocations, sector distributions, and portfolio optimization strategies. Through exploratory analysis, we examine the composition and characteristics of mutual funds, revealing insights into the distribution of fund sizes, prevalence of investment types, and diversification strategies. Asset allocation analysis provides further insights into the distribution of investments across different asset classes, while sector distribution analysis unveils the allocation of investments across various industries. Additionally, we explore portfolio optimization techniques, such as mean-variance optimization and the efficient frontier, aiming to maximize risk-adjusted returns while adhering to specified constraints. Leveraging linear programming and LP solver techniques, we formulate optimization problems to construct optimal portfolios tailored to specific risk-return objectives and sector allocation constraints.

2 What are Mutual Funds?

Definition and How it works:

A mutual fund is a managed portfolio of investments that investors can purchase shares of. Mutual funds combine money from many investors to buy a variety of investments. Professional managers decide which investments to buy and sell for the fund. A professional fund manager handles this mix of investments, and its assets and goals are detailed in the fund's prospectus.

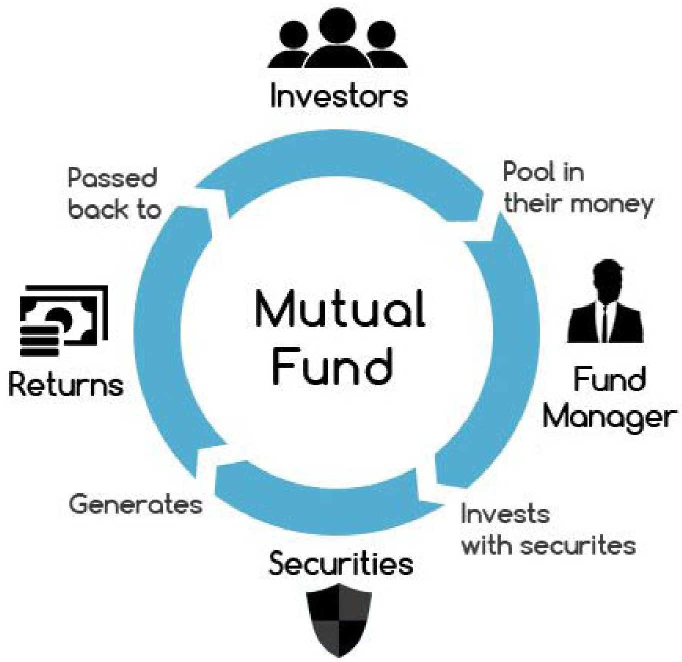

Figure 1: Visual representation of the process

Understanding Mutual Funds:

- A mutual fund owns a portfolio of investments funded by all the investors who have purchased shares in the fund. So, when an individual buys shares in a mutual fund, they gain part-ownership of all the underlying assets the fund owns. The fund's performance depends on how its collective assets are doing. When these assets increase in value, so does the value of the fund's shares. Conversely, when the assets decrease in value, so does the value of the shares.

- The mutual fund manager oversees the portfolio, deciding how to divide money across sectors, industries, companies, etc., based on the strategy of the fund. About half of the mutual funds held by American households are in index equity funds, which have portfolios that comprise and weigh the assets of indexes to mirror the SP 500 or the Dow Jones Industrial Average (DJIA). The largest mutual funds are managed by Vanguard and Fidelity. They are also index funds.

3 Data Overview

The file contains 24,821 Mutual Funds and with general aspects (as Total Net Assets, management company and size), portfolio indicators (as cash, stocks, bonds, and sectors), returns (as year-to-date, 2020-11) and financial ratios (as price/earning, Treynor and Sharpe ratios, alpha, and beta). This data was scraped from Yahoo Finance by Kaggle user Stefano Leone.

Link to dataset: https://www.kaggle.com/stefanoleone992/mutual-funds-and-etfs

3.1 Data Description

The datasets were meticulously curated to enable comprehensive comparisons in portfolio decision-making among Mutual Funds and. Neither dataset received financial backing from corporate entities or private organizations, and both were compiled from publicly available information. Each dataset is categorized by fund type, distinguishing between Mutual Funds (in "Mutual Funds.csv") and other datasets. Within these datasets, columns cover a spectrum of fund attributes, including financial ratios, sector allocations, risk indicators, and returns, spanning the timeframe from 2011 to 2020. While the original datasets were devoid of errors, they did contain null values in the form of empty columns, which were systematically removed during the data cleaning phase.

3.2 Data Limitations

A significant data limitation we encountered during our analysis was the prevalence of 'NaN' or empty values in the dataset columns. These empty entries posed a challenge as they rendered many of our analytical functions ineffective, as they require numerical inputs. To address this issue, we opted to remove these 'NaN' values. However, this approach resulted in the removal of entire rows containing 'NaN' values, leading to the loss of potentially valuable data points. Despite the importance of the data in these rows, we had no choice but to discard them due to the presence of 'NaN' values. This was a frequent occurrence due to the high frequency of 'NaN' values in the dataset, resulting in a significant loss of data. Managing this issue was a major limitation we faced during our analysis.

4 Exploratory Analysis of Mutual Funds Dataset

The exploratory analysis of the Mutual Funds Dataset delves into the comprehensive examination of mutual funds data. Through this analysis, we investigate various aspects such as fund size, investment types, asset distributions, and sector allocations.

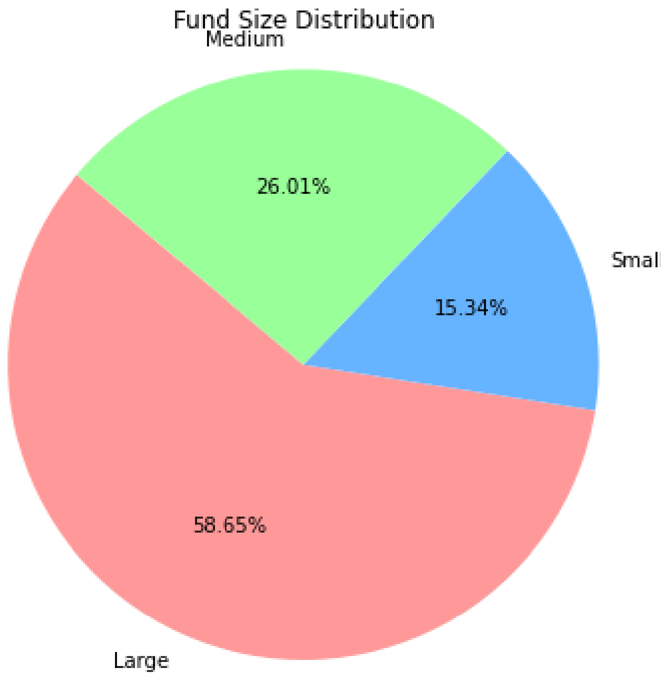

Figure 2: From this pie chart we can see that there is more large mutual funds (58.65%)than small funds (15.34%) and medium funds (26.01%) combined together. Fund size refers to the total amount of capital, typically in monetary terms, that a fund has available for investment. It serves as a crucial indicator of the fund's capacity to execute its investment strategies and meet its objectives. A fund's size can significantly impact its ability to diversify across assets, sectors, and geographies, influencing its risk profile and potential returns. Larger funds may enjoy economies of scale, potentially lowering operational costs and enhancing liquidity, while smaller funds might have more flexibility and agility in navigating niche markets. Ultimately, understanding a fund's size is essential for investors seeking to assess its suitability within their investment portfolio and align their expectations with its investment approach and potential outcomes.

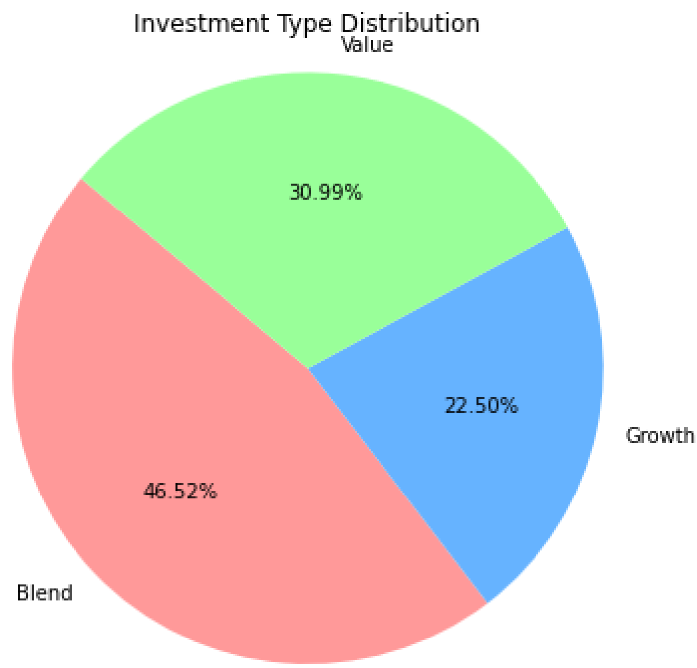

Figure 3: There are three main investment types for a fund: Growth, Value, and Blend. From here we can see that the dominant Investment type is Blend at 46.52%, followed by Value with 30.99% and then Growth with 22.5%. Growth funds primarily invest in stocks expected to increase in capital value rather than yield high income. These funds often target companies with strong growth potential, such as those in emerging industries or with innovative products/services. Value funds, on the other hand, follow a strategy that focuses on investing in stocks perceived to be undervalued based on fundamental characteristics. Blend funds offer a combination of both growth and value stocks within their portfolios. Importantly, blend funds exclusively invest in stocks and do not include fixed-income securities.

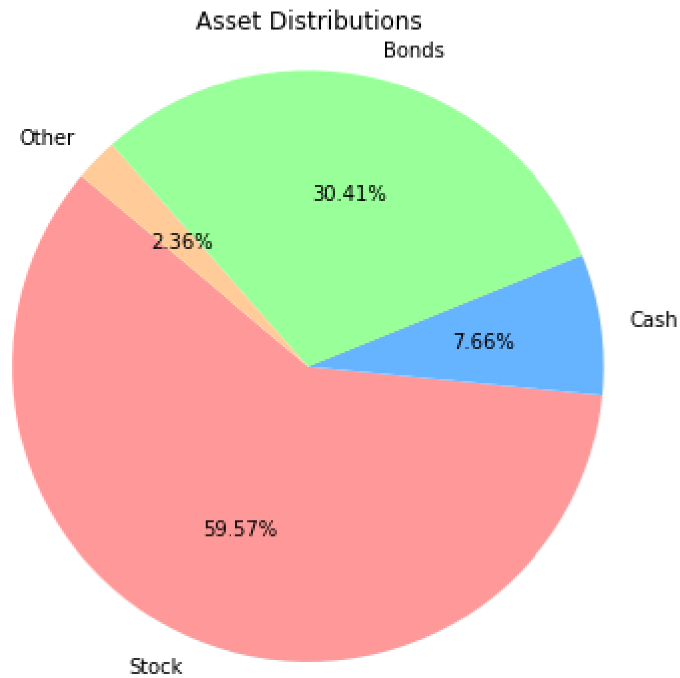

Figure 4: The graphs display the composition of the fund's assets, indicating the percentages allocated to each asset type. Each segment represents a specific asset category, including stocks, bonds, cash, and other investments. These distributions illustrate how the fund's investments are divided among different asset classes, providing insights into its overall portfolio structure. Understanding these distributions is essential for investors to assess the fund's diversification, risk exposure, and potential returns. In this case, funds have about half their assets in stocks and about a third in bonds, then they have the remaining tenth in cash and other.

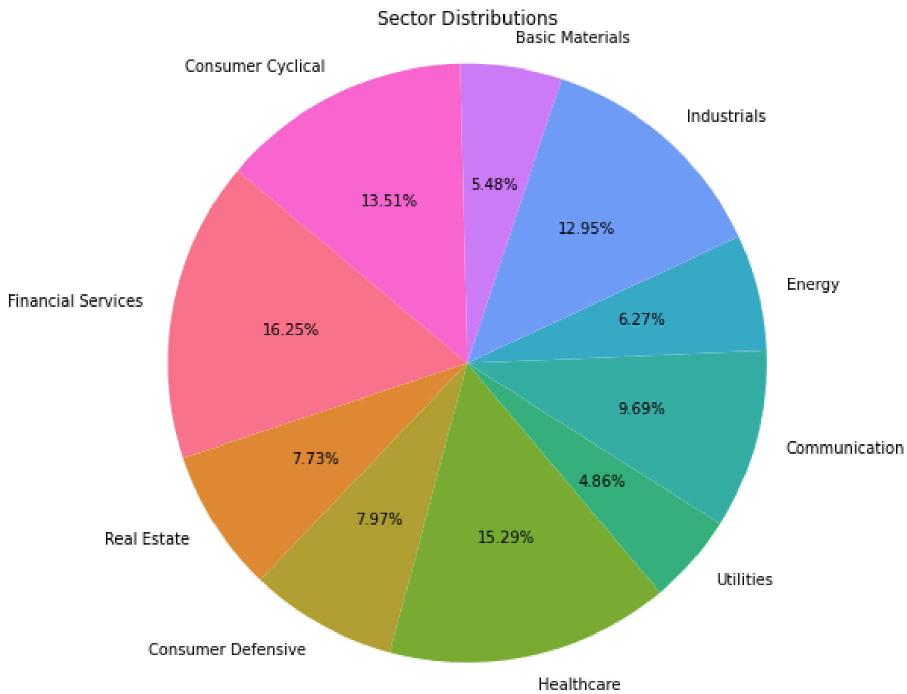

Figure 5: The funds sector distribution provides insight into how it allocates investments across various industries. With a significant emphasis on Financial Services and Healthcare sectors at 16.25% and 15.29% respectively, the funds appear positioned to capitalize on opportunities within these segments. Additionally, allocations to sectors such as Consumer Cyclical and Industrials at 13.51% and 12.95% suggest a focus on areas sensitive to economic cycles. Meanwhile, sectors like Utilities and Basic Materials receive smaller allocations, indicating a potentially more conservative stance. Overall, this diversified sector allocation aims to balance growth opportunities with risk management within the funds portfolio.

5 LP for Portfolio Optimization (Example with MF Returns Maximisation)

1. **Set up LP problem**:

   - We define a linear programming (LP) problem named "Portfolio_Optimization" with the objective of maximizing returns. Mathematically, this can be represented as:

   $$\text{Maximize } Z = \sum_i r_i \times x_i$$

   - Here, Z represents the objective function (total returns), r_i represents the return of fund i, and x_i represents the allocation of fund i.

2. **Total allocation constraint**:

   - We impose a constraint that the total allocation across all funds must equal 1, indicating that all available investment capital is fully allocated. Mathematically, this constraint can be represented as:

   $$\sum_i x_i = 1$$

   - Here, x_i represents the allocation of fund i, and the sum of all allocations equals 1.

3. **Sector allocation constraints**:

   - Based on the provided data, we impose constraints on sector allocations. These constraints restrict the allocation of funds to specific sectors to adhere to investment strategies or risk management objectives.
   - For example, if we want to limit the allocation to the technology sector (S_{tech}) to 30percent, we would have the constraint:

   $$\sum_{i \in S_{tech}} x_i \leq 0.3$$

   - Similarly, if we want to ensure at least 10 percent allocation to the healthcare sector (S_{health}), we would have the constraint:

   $$\sum_{i \in S_{health}} x_i \geq 0.1$$

```python
# Set up LP problem
prob = LpProblem("Portfolio_Optimization", LpMaximize)
prob += objective

# Total allocation constraint
prob += lpSum(allocation[fund] for fund in funds) == 1

# Sector allocation constraints based on the provided data
prob += lpSum(allocation[fund] for fund in funds if data.loc[data['fund_symbol'] == fund]['fund_sector_technology'].values[0]) <= 0.3
# Limit allocation to technology sector to 30%
prob += lpSum(allocation[fund] for fund in funds if data.loc[data['fund_symbol'] == fund]['fund_sector_healthcare'].values[0]) >= 0.1
# Ensure at least 10% allocation to healthcare sector
prob += lpSum(allocation[fund] for fund in funds if data.loc[data['fund_symbol'] == fund]['fund_sector_consumer_cyclical'].values[0]) <= 0.2
# Limit allocation to consumer cyclical sector to 20%

# Solve the LP problem
prob.solve()

# Check the status of the solution
print("Status:", LpStatus[prob.status])

# Print optimized allocation
print("Optimized Allocation:")
for fund, var in allocation.items():
    if var.varValue > 0:
        print(fund, ": ", var.varValue)
```

Figure 6: Code Solution: We decided to utilise LP solver code for solving this optimisation problem by coding our LP problem and the above-mentioned constraints. Bellow you can find detailed explanation on how it works.

4. **Solve the LP problem**:

    - We utilize an LP solver to find the optimal solution that maximizes the objective function while satisfying all constraints. The solver explores possible allocations of funds to maximize returns within the defined constraints.

5. **Check the status of the solution**:

    - After solving the LP problem, we check the status to ensure that the solution is feasible and whether any issues were encountered during optimization. This step provides insight into the success or failure of the optimization process.

6. **Print optimized allocation**:

    - Finally, we print the optimized allocation, indicating how much of the total investment capital should be allocated to each fund. This allocation represents the optimal portfolio composition that maximizes returns while satisfying all constraints.

Results:

- The optimization process successfully found an optimal solution with a status of "Optimal".

- The optimized allocation suggests investing 20 percent of the portfolio in the LSHUX fund and 80 percent in the RSNYX fund, based on the constraints and objective function provided.

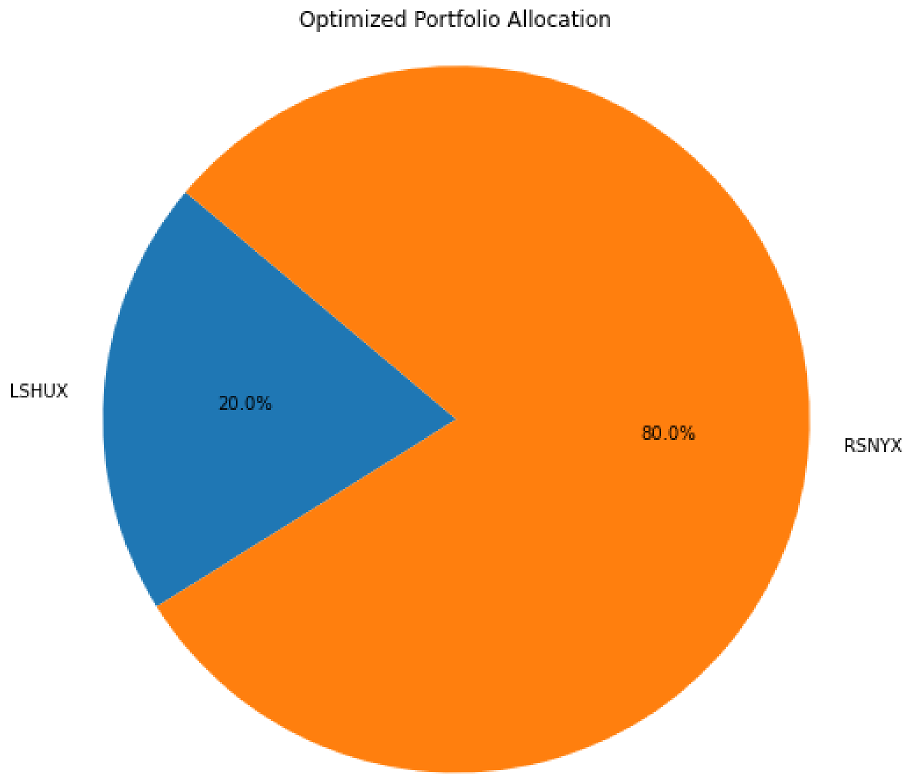

Figure 7: Pie Chart: Here we have a pie chart of the results from the simple optimised portfolio using LP solver code. We can see that here the smaller blue part of the pie chart represents the LSHUX fund with 20 % suggested asset allocation and the bigger orange part of the pie chart represents RSNYX fund with 80 % suggested asset allocation

By following these steps, we effectively utilized linear programming techniques to optimize portfolio allocation while considering constraints on sector allocations, ultimately achieving a balanced and efficient investment portfolio. Adjustments to the constraints and objectives can be made to tailor the optimization process to specific investment goals and preferences.

6 Optimising Trading Portfolio for Mutual Funds

Portfolio Optimization

- A portfolio is constructed using the Riskfolio library.
- Historical data is leveraged to compute asset statistics.
- Employing the Mean-Variance model for portfolio optimization.
- The aim is to maximize the Sharpe ratio, ensuring a balanced approach to risk and return.

The Mean-Variance model computes the portfolio's expected return by aggregating the weighted expected returns of each individual asset:

$$E(R_p) = w_1 \cdot E(R_1) + w_2 \cdot E(R_2) + \ldots + w_n \cdot E(R_n)$$

The Sharpe ratio, a key measure for portfolio optimization, is calculated as:

$$\text{Sharpe Ratio (SR)} = \frac{E(R_p - R_f)}{\sigma_p}$$

This formula quantifies the excess return of the portfolio (over the risk-free rate) relative to its risk (volatility). A higher Sharpe ratio indicates a better risk-adjusted performance.

6.1 Efficient Frontier

The efficient frontier is the set of optimal portfolios that offer the highest expected return for a defined level of risk or the lowest risk for a given level of expected return. It is represented by a curve in a graph where the x-axis represents the risk (standard deviation of the portfolio return) and the y-axis represents the return (expected portfolio return) riskfolio-lib.readthedocs.io.

	AMMIX	DCGVX	CLPAX	AFAZX	DVIPX	CSCZX	CMNZX	CHCRX	ABAEX
1	0.000000	0.003744	-0.008040	0.000000	0.006000	-0.006444	0.000000	0.000000	-0.013986
2	0.001000	0.006631	0.004053	0.012484	-0.001988	-0.004864	0.000000	-0.004959	-0.013171
3	0.000999	0.009057	0.000000	0.012330	0.001992	-0.002851	0.000000	0.001661	0.016427
4	0.006986	-0.011832	-0.001009	0.009744	-0.002982	0.004902	-0.001832	0.009950	-0.007071
5	-0.003964	-0.001652	0.007071	0.006031	-0.002991	0.002033	-0.001835	-0.011494	-0.020346

Figure 8: Data Representation Table with Returns over Time Periods

	AMMIX	DCGVX	CLPAX	AFAZX	DVIPX	CSCZX	CMNZX	CHCRX	ABAEX
weights	0.409303	3.378763e-12	0.15868	0.141518	0.191846	7.998876e-12	5.089144e-12	0.098653	4.046945e-10

Figure 9: Portfolio Optimisation Table with Weighted Values for Mutual Funds over time

Here's how to interpret the Figure 8:

- Each row represents a time period, with consecutive rows indicating successive time periods.
- Each column represents an asset or fund, with the asset symbol serving as the column header.
- The numerical values in the DataFrame represent the percentage returns of each asset for the corresponding time period.
- Positive values indicate positive returns, while negative values indicate negative

6.2 Classical Mean Variance Optimisation Problem Formulation

Maximum Sharpe Ratio Portfolio with Weights w and Covariance Matrix Σ

$$\text{Maximize} \quad \frac{R(w) - rf}{\sqrt{w^T \Sigma w}}$$

$$\text{Subject to} \quad A \cdot w \geq B$$

$$R(w) \geq \bar{\mu}$$

6.3 Comparative Analysis of 4 Risk Measures

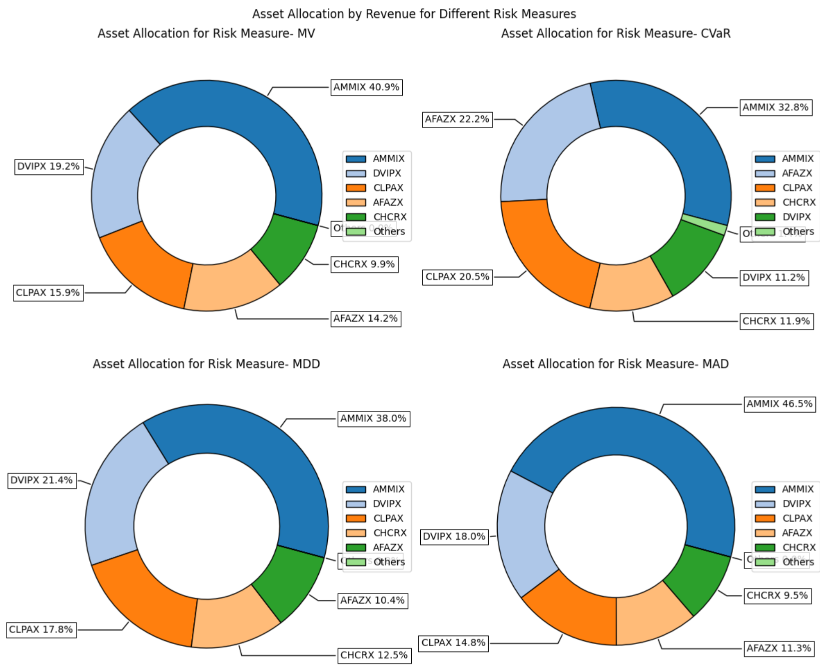

Figure 10: Here we have 4 different asset allocation suggestions based on the different risk measures. We can see from the graphs that all 4 of have AMMIX as the most dominant choice compared to others. Then we see that on second place for MV, MDD and MAD is DVIPX fund while its not for CVaR.This could mean that different risk measures prioritize different aspects: CVaR might be placing more emphasis on tail risk or extreme events compared to other measures like MV, MDD, or MAD. This difference in prioritization could lead to variations in asset allocation.

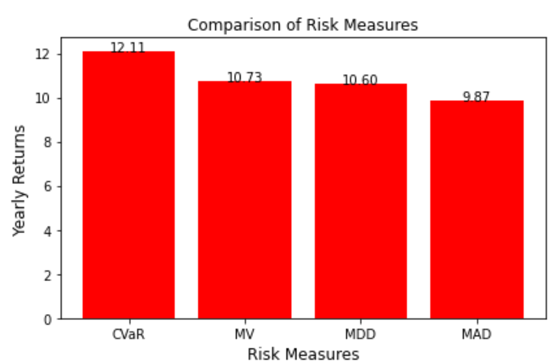

Figure 11: Comparative Analysis of All Risk Measures - Bar Plot

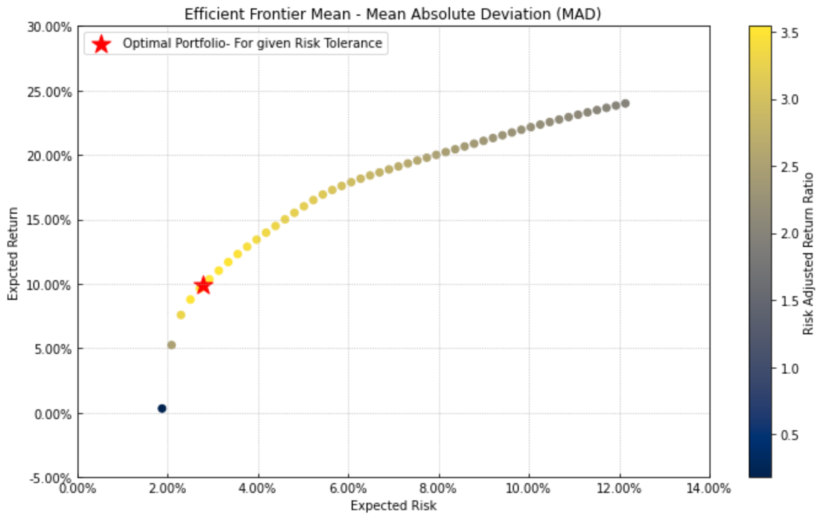

Figure 12: Visualized is the expected return in % for a given expected risk in %. The error used is the mean absolute deviation and the color represents the risk adjusted return ratio. Generally, we can see a decreasing slope for increasing risk. To get a 10% return we would need a risk of 2.9% which is highligthed by the red star

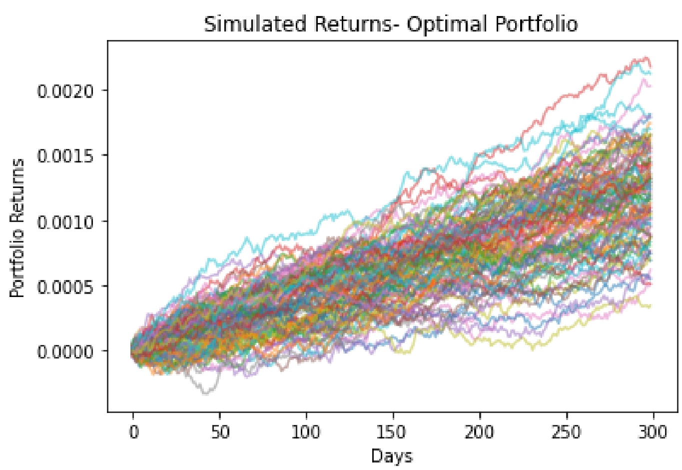

Figure 13: The graph depicts the simulated cumulative returns of an optimal portfolio over a 300-day period, generated through 2000 Monte Carlo simulations. Each simulation represents a hypothetical scenario of portfolio performance based on historical return distributions and specified weights. The graph highlights a visible upward trend in portfolio returns, indicating the potential for growth over time. Despite variations in individual simulation outcomes, the overall trajectory suggests a positive outlook for the portfolio's performance.

7 Conclusion

In this paper, we conducted a comprehensive analysis of mutual funds utilizing a dataset sourced from Yahoo Finance. Our exploration encompassed various dimensions of mutual funds, including fund size, investment types, asset allocations, sector distributions, and portfolio optimization strategies. Through exploratory analysis, we gained valuable insights into the composition and characteristics of mutual funds. Asset allocation analysis shed light on the distribution of investments across different asset classes, providing insights into diversification strategies and risk management approaches adopted by mutual funds. Additionally, sector distribution analysis unveiled the allocation of investments across various industries, indicating potential growth opportunities and risk exposures within fund portfolios.

Building upon the foundational understanding derived from EDA, we employed quantitative techniques for portfolio optimization. Linear programming (LP) was utilized to formulate optimization problems aimed at maximizing returns while adhering to constraints such as total allocation and sector allocations. LP solvers provided optimal portfolio compositions tailored to specific investment goals and risk management objectives. Additionally, we leveraged the Riskfolio library to conduct portfolio optimization based on historical data, utilizing the classical Mean-Variance model to maximize the Sharpe ratio. By visualizing the efficient frontier and comparing risk measures, we identified optimal portfolios offering the highest expected returns for given levels of risk. Furthermore, Monte Carlo simulations provided insights into the potential performance of optimized portfolios over time. By simulating hypothetical scenarios and analyzing cumulative returns, we demonstrated the robustness and growth potential of optimized portfolios, despite market uncertainties. While the simulation offers valuable insights into portfolio performance, several limitations warrant consideration. Firstly, the assumption of normality in asset returns may not always hold true, particularly during periods of market turbulence, potentially impacting the accuracy of the results. Additionally, the simulation assumes static correlations between asset returns, overlooking potential variations over time that could affect portfolio dynamics. Furthermore, focusing on a single 300-day period may not fully capture the long-term evolution of portfolio performance or account for diverse market conditions. Moreover, external factors such as economic indicators or geopolitical events are not incorporated into the analysis, limiting its predictive capability.

In conclusion, our study underscores the importance of integrating exploratory data analysis with quantitative methodologies in portfolio management and investment decision-making. By combining data-driven insights with rigorous optimization techniques, investors can navigate financial markets effectively, mitigate risks, and maximize returns, thereby achieving their investment objectives with confidence and agility.