DSA4212 Assignment 1: Portfolio Optimization

Da Weizhi, Lei Zhenyu, Tritham Jirawan, Wang Shijia

November 8, 2024

1 Introduction

In this project, we aim to construct an optimal investment portfolio by applying modern portfolio theory, inspired by the S&P 500 Index, which tracks 500 of the largest publicly traded companies in the United States. We have selected a subset of 75 stocks, representing various industries, and used their historical adjusted closing prices to calculate the optimal portfolio weights that balance risk and return.

To achieve this, we utilize the mean-variance optimization framework, a method developed by Harry Markowitz. By minimizing portfolio variance while targeting specific levels of return, we derive the efficient frontier—a curve representing portfolios with the best possible risk-return trade-offs. The optimization problem is formulated and solved using the Python library CVXPY, which allows us to implement convex optimization techniques to minimize risk.

Through this study, we will generate the efficient frontier, identify the optimal risk-return trade-offs, and explore the stability and robustness of our solutions across different time periods. Our findings will demonstrate the application of quantitative techniques to optimize portfolio returns and inform decision-making in real-world financial markets.

2 Data Collection and Preprocessing

To create a dataset for the portfolio optimization, we first selected 75 of the largest companies by market value from the S&P 500 Index. These companies are spread across various industries, including Technology, Consumer Discretionary, and Financials, allowing for diversified portfolio construction. The stock tickers of these companies were manually compiled into a list, as shown in Table 1, which was then used to download historical price data.

For each stock, we retrieved the adjusted closing price from Yahoo Finance using the yfinance Python library. We specified a date range from January 1, 2022, to October 1, 2024, to ensure we had recent and relevant data for our analysis.

The stock data was stored in a pandas DataFrame, where each column corresponds to a particular stock's adjusted closing price over the selected time period.

Stock Tickers				
MSFT	AAPL	NVDA	GOOG	AVGO
AMZN	META	BRK-B	LLY	TSLA
BA	JPM	V	WMT	XOM
UNH	MA	ORCL	COST	PG
CVX	PEP	HD	MRK	ABBV
KO	ADBE	NFLX	AMD	PFE
DIS	RTX	SPGI	GS	CMCSA
LOW	T	PGR	UNP	BKNG
HON	MS	ETN	BLK	COP
TJX	BSX	LMT	VRTX	SYK
C	PANW	MU	ADP	ADI
BX	KLAC	LRCX	CB	SBUX
ELV	MDT	MMC	PLD	FI
BMY	GILD	AMT	HCA	INTC
CI	PLTR	ICE	SHW	MDLZ

Table 1: List of 75 selected tickers for portfolio optimization

3 Methodology

3.1 Portfolio Return and Variance

To construct the optimal portfolio, we follow the mean-variance optimization framework (MVO) The portfolio return R_P is the weighted sum of the returns of individual assets:

$$R_P = \sum_{i=1}^{n} x_i R_i$$

where x_i represents the weight of stock i in the portfolio, and R_i is the return of stock i.

The portfolio variance is given by:

$$\text{Var}(R_P) = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j \text{Cov}(R_i, R_j)$$

This formula shows that the portfolio variance depends on the individual stock variances and covariances between stocks. These equations are adapted from Berk and DeMarzo (2023) [1].

3.2 Plotting the Efficient Frontier

The efficient frontier represents the set of optimal portfolios that offer the lowest risk for a given level of return. To plot the efficient frontier, we formulate an optimization problem to minimize the portfolio variance for a given expected return $R_P \in [0, 4]$.

The solution to the optimization problem provides the weights for each asset in the portfolio that minimize the portfolio variance. To assess the stability of the efficient frontier over time, we employ a rolling window approach. This allows us to observe how the efficient frontier evolves over time and whether there is need for re-balancing under different market conditions.

Mathematically, the optimization problem can be written as:

$$\text{Minimize } \mathbf{w}^T \mathbf{Q} \mathbf{w}$$

subject to:

$$\sum_{i=1}^{n} w_i = 1, \quad \mu^T \mathbf{w} = R, \quad -1 \leq w_i \leq 1 \text{ for all } i$$

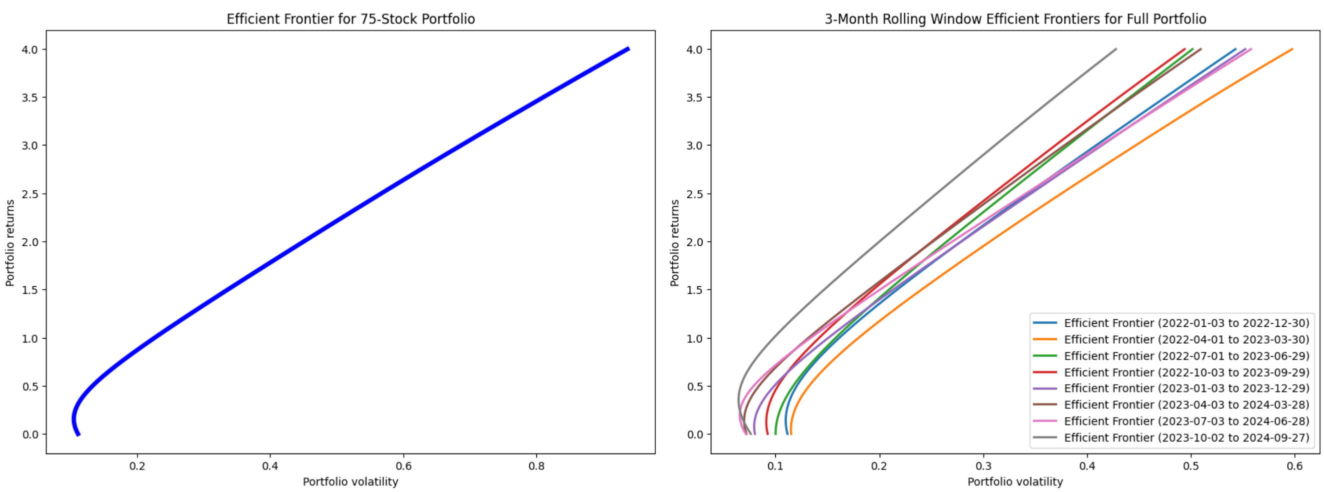
3.2.1 Efficient Frontier of Full Portfolio

The efficient frontier for the full portfolio is constructed using the complete dataset from January 1, 2022 to October 1, 2024. This gives us the baseline to assess the long-term trade-off between risk and return across the 75 selected stocks. This approach serves as a baseline, capturing the risk-return dynamics over an extended period and allowing us to evaluate portfolio performance under stable market conditions.

As shown in Figure 1a, the curve represents the optimal portfolios that achieve the highest expected return for a given level of risk (volatility). The risk-return relationship is convex, which indicates that the marginal increase in return diminishes as risk increases.

The analysis shows that portfolios with lower volatility provide relatively modest returns, while higher returns come with significantly more risk. For example, achieving a 100% increase in portfolio value would require accepting a volatility of approximately 0.25. This reflects typical market behavior, where achieving higher returns necessitates taking on more risk.

However, this static efficient frontier lacks sensitivity to short-term market fluctuations, leading us to explore a rolling window approach for a more dynamic analysis.



(a) Efficient Frontier for Full Portfolio

(b) 3-Month Rolling Window Efficient Frontiers

Figure 1: Efficient Frontiers

3.2.2 Rolling Window Analysis

To examine the stability of efficient frontiers over time, we conducted a rolling window analysis with 3-month intervals. Every 3 months, we update the input to reflect the previous 12 months' performance and reconstruct the efficient frontier. This approach reveals how the efficient frontier adapts to changing market conditions, capturing short-term shifts that a full-period analysis may overlook.

We chose a 3-month window to balance capturing meaningful changes in market behavior with retaining enough data for reliable portfolio optimization. As seen in Figure 1b, the efficient frontiers exhibit noticeable variation, indicating an unstable relationship between risk and return over time. Certain windows show significantly higher volatility for similar expected returns, pointing to periods of heightened market risk or instability.

This variability underscores the importance of periodically reassessing portfolio allocations. Later windows tend to show steeper efficient frontiers with better risk-return trade-offs, while earlier windows are flatter, possibly reflecting the impact of the COVID-19 pandemic on the US stock market.

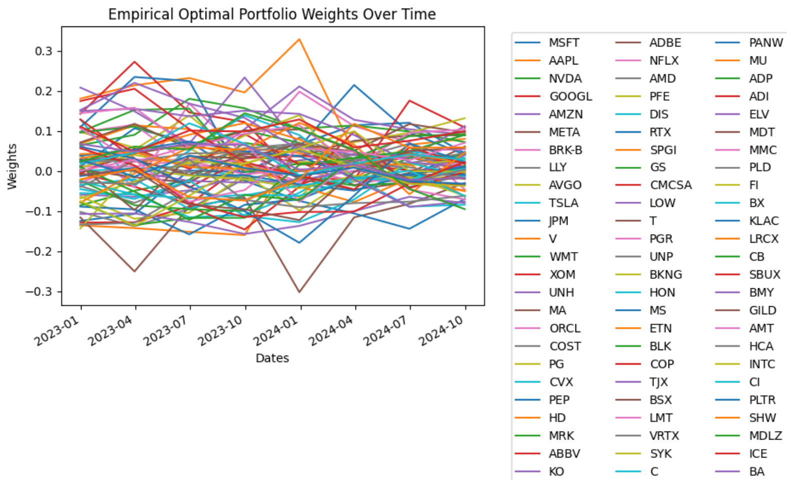


Figure 2: Optimal Portfolio Weights Over Time

The fluctuations in portfolio weights over time, as shown in Figure 2, further point to the instability of the efficient frontier under shifting market conditions. These weights were derived from portfolios with an expected return of 50%, illustrating how different market environments impact optimal allocations. While frequent portfolio reallocations might help adapt to these changes, they also incur significant transaction costs, which can erode returns.

To address this instability, we employ robust covariance estimators—specifically Ledoit-Wlf and a combination of MCD (Minimum Covariance Determinant) with Ledoit-Wolf. These estimators provide a more reliable covariance matrix, which is highly sensitive to outliers and plays a crucial role in portfolio optimization. By achieving a more stable efficient frontier with reduced fluctuations across rolling windows, we aim to minimize the need for frequent rebalancing, ultimately helping to control transaction costs and improve portfolio consistency.

3.3 Robust Covariance Estimation

In portfolio optimization, accurate estimation of the covariance matrix is critical, as it influences risk assessment and asset allocation. Traditional estimation methods are often sensitive to sample size limitations and outliers, which can lead to instability in optimization results.

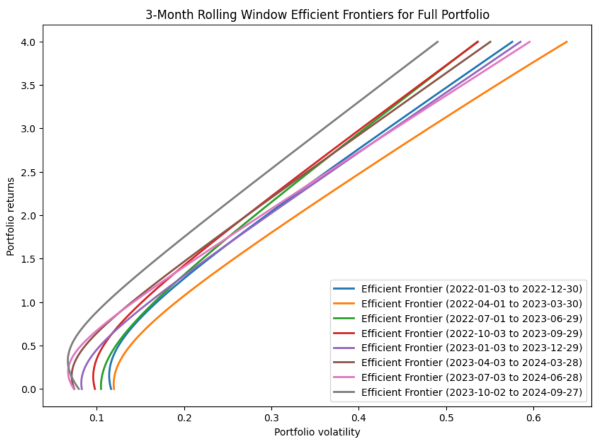
In the following sections, we apply two robust covariance estimators and compare their efficient frontiers and portfolio weights over time.

3.3.1 Ledoit–Wolf Shrinkage [2]

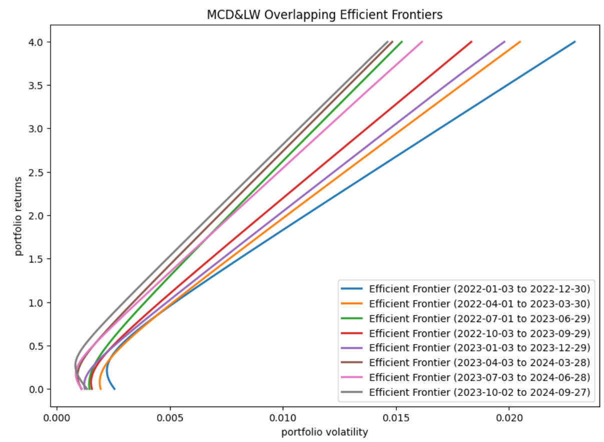
Ledoit-Wolf shrinkage enhances the accuracy of the covariance matrix by reducing estimation error. It does this by shrinking the empirical covariance matrix toward a more structured target, such as the identity matrix. This shrinkage process helps to stabilize covariance estimates, mitigating the effects of sampling variability and outliers, both of which can distort mean-variance optimization outcomes.

By applying the Ledoit-Wolf estimator, we obtain a more reliable risk profile that is less prone to overfitting. As shown in Figure 3a, the Ledoit-Wolf efficient frontier demonstrates greater stability across varying market conditions compared to the baseline approach. The frontiers appear more closely aligned, thereby reducing the need for frequent portfolio rebalancing.

Additionally, Figure 4a illustrates the portfolio weights over time under the Ledoit-Wolf estimator. The weights exhibit reduced volatility and gradually converge toward moderate values, suggesting that robust covariance estimation leads to a more consistent allocation strategy.

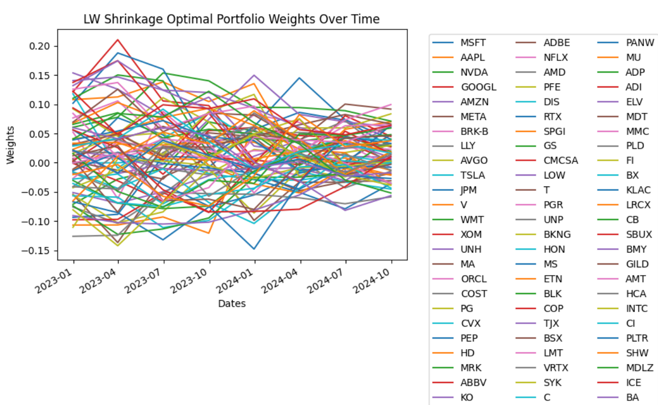


(a) Efficient Frontier with Ledoit-Wolf

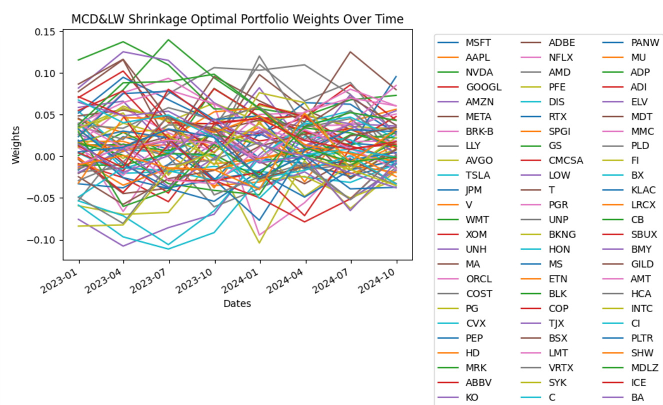


(b) Efficient Frontier with MCD X Ledoit-Wolf

Figure 3: Efficient Frontiers with Robust Covariance Estimators



(a) Portfolio Weights with Ledoit-Wolf



(b) Portfolio Weights with MCD X Ledoit-Wolf

Figure 4: Portfolio Weights with Robust Covariance Estimators

3.3.2 Minimum Covariance Determinant (MCD) [3]

To further enhance the robustness of our covariance estimation, we apply the Minimum Covariance Determinant (MCD) estimator followed by the Ledoit-Wolf (LW) shrinkage. The MCD estimator identifies the subset of data points with the smallest covariance determinant, effectively down-weighting outliers and yielding a stable initial covariance matrix that captures the underlying data structure with minimal influence from extreme values.

After obtaining the covariance matrix from MCD, we apply Ledoit-Wolf shrinkage to further stabilize this estimate. This combined approach leverages MCD's outlier-resilience along with the shrinkage benefits of Ledoit-Wolf, providing a covariance matrix that is both robust and adaptable across diverse market conditions.

By combining MCD with Ledoit-Wolf, we obtain the efficient frontier shown in Figure 3b. The efficient frontier converges more tightly around the global minimum variance point, where volatility is lowest, and diverges gradually as volatility increases. As shown on the axes, the efficient frontiers under this approach exhibit significantly lower volatility for similar levels of expected returns compared to the Ledoit-Wolf approach alone, highlighting an improvement in stability.

Additionally, Figure 4b illustrates the portfolio weights under the MCD&LW approach. The weights display a stronger convergence, with less dispersion than those obtained using the standalone Ledoit-Wolf method, occupying only about half the range seen with Ledoit-Wolf alone.

3.3.3 Transaction Cost

In high-frequency rebalancing portfolio practices, transaction costs are a critical factor to consider. In real-world applications, these costs can become significant, especially in portfolios involving options, commodities, or futures trading, as they directly impact portfolio returns. To address this, we incorporate transaction costs as a penalty term in our optimization model. This penalty term discourages frequent and substantial portfolio adjustments, helping to balance the trade-off between achieving optimal returns and minimizing transaction expenses.

At each rebalancing step, we assume that each adjustment in portfolio weights incurs a proportional transaction cost based on the absolute change in weights. This adjustment is multiplied by a transaction rate, represented as τ, to quantify the total transaction cost for rebalancing. Consequently, our optimization problem now includes a penalty term for transaction costs, where the objective is to minimize the portfolio variance while accounting for these additional costs. With a transaction rate τ, the optimization problem is formulated as follows:

$$\text{Minimize} \quad w^T Q w + \tau \sum_{i=1}^{n} |w_i - w_{i,\text{prev}}|$$

subject to:

$$\sum_{i=1}^{n} w_i = 1, \quad \mu^T w = R, \quad -1 \le w_i \le 1 \quad \text{for all } i$$

In this formulation, the transaction cost penalty term, $\tau \sum_{i=1}^{n} |w_i - w_{i,\text{prev}}|$, serves to limit drastic weight changes, thus reducing transaction costs and promoting a more stable portfolio structure.

4 Analysis

In this analysis, we examine the stability and performance of optimal portfolios constructed using three covariance estimation methods: the empirical method, Ledoit-Wolf shrinkage, and combined MCD&LW approach with the same rolling window method.

4.1 Weights fluctuation

Figure 5 illustrates the standard deviation of portfolio weights over time for each covariance estimation method, based on an annualized expected return of 50% over 3-month rolling windows. The empirical efficient frontier shows high fluctuations in weight standard deviations, indicating its sensitivity to changes in the covariance matrix estimation. This sensitivity reflects the empirical method's vulnerability to estimation noise and market variability. In contrast, the Ledoit–Wolf shrinkage method reduces these fluctuations by introducing a shrinkage factor that mitigates estimation error, resulting in a more stable portfolio. The combined MCD&LW approach further stabilizes the weights, achieving the lowest weight standard deviation among the methods. This combination leverages the robustness of the MCD against outliers and the Ledoit–Wolf's shrinkage adjustment, producing a portfolio that effectively manages both estimation noise and data variability.

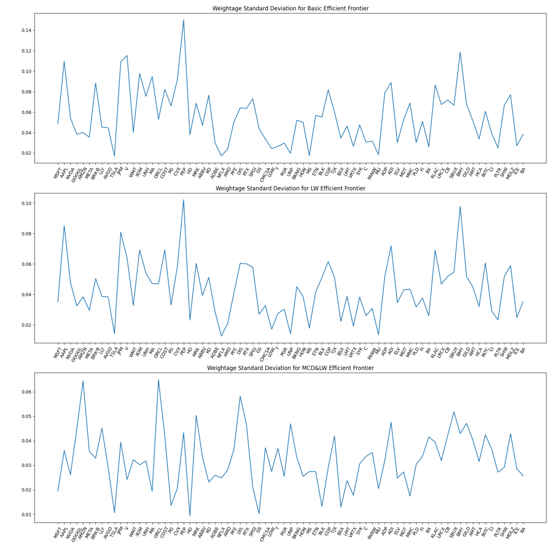


Figure 5: Weight Standard Deviation for Different Efficient Frontiers

In summary, the Ledoit–Wolf shrinkage and MCD&LW methods both demonstrate enhanced stability in portfolio weights compared to the empirical approach, with the combined method achieving the greatest reduction in fluctuations. This stability is essential, as it minimizes transaction costs and reduces the risk of abrupt weight changes in response to market volatility. The impact of these stabilized weights is further reflected in the cumulative returns, as we examine next in the portfolio performance analysis.

4.2 Portfolio Performance

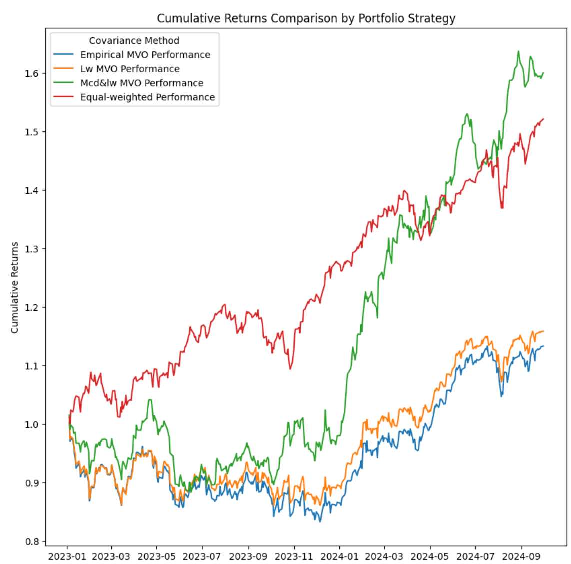


Figure 6: Cumulative Returns Comparison

Figure 6 illustrates the cumulative returns for each portfolio strategy, comparing optimized portfolios using different covariance estimators with a baseline equal-weighted portfolio. Each strategy follows a quarterly rebalancing schedule alongside rolling window estimation, allowing for adaptive responses to shifting market conditions while managing transaction costs. In this bullish market environment, the equal-weighted portfolio achieves a cumulative return exceeding 50% over a 21-month period, setting a benchmark target annual return of 50%.

The empirical MVO portfolio experiences more pronounced fluctuations, with cumulative returns stabilizing around a 13% increase. This reflects its sensitivity to market changes, resulting in moderate underperformance relative to the equal-weighted strategy. In comparison, the Ledoit–Wolf MVO portfolio consistently generates higher returns, resulting in a 15% cumulative increase, attributed to its stabilized weight adjustments that reduce estimation error. This small 2 percentage point improvement over the empirical method demonstrates Ledoit–Wolf's abilities in capturing market trends more reliably.

Notably, the MCD&LW strategy yields the highest performance, achieving approximately 60% cumulative returns over the same period. This result surpasses both the empirical and Ledoit–Wolf methods by 47 and 45 percentage points, respectively. The enhanced returns suggests that MCD&LW effectively balances risk and return, particularly in high-volatility market phases where empirical estimates struggle.

While the MCD&LW strategy does not consistently outperform the equal-weighted portfolio across all months, it maintains a stable advantage, with higher cumulative returns over the period. This outcome highlights MCD&LW as an effective approach to achieving both robustness and performance. When markets are favorable, the combined method capitalizes on gains more effectively, whereas during downturns, its risk management attributes help preserve value better than purely empirical estimates.

4.3 Metrics for Robustness

Metric	Empirical	LW	MCD&LW	Equal-Weighted
Cumulative Return	1.131	1.159	1.601	1.521
Risk (Annualized)	0.124	0.126	0.007	0.121
Sharpe Ratio	0.278	0.391	1.629	1.773
Max Drawdown	-0.169	-0.143	-0.153	-0.071
Turnover Rate	4.957	4.602	1.410	0.000

Table 2: Comparison of New Portfolio Metrics by Approach

The table 2 presents various metrics to evaluate the robustness of different portfolio optimization strategies, including cumulative return, annualized risk, Sharpe ratio, max drawdown, and turnover rate. The MCD&LW strategy demonstrates a balance between high returns and effective risk management, achieving the highest end return (1.601) and an impressive Sharpe ratio (1.629). This suggests that combining robust covariance estimators, such as MCD with LW shrinkage, enhances the portfolio's resilience and adaptability to market conditions, resulting in better risk-adjusted performance.

The equal-weighted approach, which avoids optimization and simply distributes weights evenly, offers a stable alternative with minimal rebalancing costs (turnover rate of 0.000) and a competitive Sharpe ratio (1.773). By not depending on estimated covariance matrices, this approach reduces sensitivity to estimation errors, making it a resilient choice in volatile or uncertain markets. Although its risk (0.121) is higher than that of MCD&LW, the equal-weighted portfolio remains robust due to its simplicity and reduced reliance on complex covariance adjustments.

It is important to note, however, that the risk (annualized) metric may not be entirely comparable across strategies due to differences in covariance adjustments. For example, the MCD&LW approach reports an exceptionally low risk (0.007), likely due to its robust adjustments that may smooth out volatility but not fully capture real-time fluctuations. Therefore, while robust optimizations like MCD&LW appear to improve risk-adjusted returns, the differences in covariance estimators make it advisable to assess portfolio robustness by focusing on cumulative return and Sharpe ratio rather than relying solely on the absolute risk values.

In conclusion, although the optimized strategies offer more tailored risk-return profiles, they come at the cost of higher transaction costs from periodic rebalancing. Among these, the MCD&LW strategy stands out for its ability to balance cumulative returns with lower volatility, supporting the case for robust covariance estimation in dynamic portfolio management. This analysis highlights the trade-offs between risk minimization, transaction costs, and cumulative returns, offering insights into the effectiveness of various robust covariance estimators in portfolio optimization.

5 Conclusion

In conclusion, this study examined various portfolio optimization strategies, highlighting the advantages and limitations of traditional and robust covariance estimation methods in managing risk and return. By comparing empirical MVO, Ledoit–Wolf shrinkage, and the combined MCD&LW approach, we demonstrated that robust estimators can significantly enhance portfolio stability and risk-adjusted performance. The MCD&LW method, in particular, exhibited the highest cumulative return and the lowest volatility, showcasing its resilience to outliers and its adaptability to changing market conditions. Despite incurring higher transaction costs than the equal-weighted strategy, its superior Sharpe ratio and controlled risk profile position it as a valuable tool for dynamic portfolio management.

However, the effectiveness of these robust adjustments across varying market scenarios remains uncertain. Given that MVO relies on historical performance data, it is inherently momentum-based and may yield suboptimal results in rapidly shifting markets/ Different approaches may be better suited to specific conditions. For instance, the EWMA method, which emphasizes recent performance, is more suitable for shorter rolling windows, such as a 1-month rebalancing period. This limitation highlights the potential need for further analysis into the behavior of robust covariance estimators under different market conditions, especially in high-volatility or bear markets.

The current quarterly rebalancing schedule may also be too infrequent to respond promptly to market fluctuations. Implementing a shorter rebalancing period, such as monthly adjustments, could allow the portfolio to better adapt to dynamic conditions. Additionally, exploring alternative techniques like LASSO for dimensionality reduction or Bayesian methods to incorporate prior knowledge could improve the robustness of portfolio optimization. Integrating MVO with advanced models, including neural networks or reinforcement learning, offers promising opportunities for capturing complex market patterns and enhancing adaptability.

Overall, our findings emphasize the trade-offs between simplicity, stability, and adaptability in portfolio construction. Robust covariance estimators like MCD&LW present a compelling option for investors seeking consistent returns in volatile markets, especially when combined with strategies for frequent rebalancing and advanced model integration to address the challenges of modern financial landscapes.

6 Appendix

We also experimented with using only the MCD covariance and an alternative approach known as Exponentially Weighted Moving Average (EWMA) covariance [4]. EWMA places greater emphasis on recent returns, allowing it to respond more quickly to changes in market conditions and better capture dynamic market volatility in rapidly shifting environments. However, both approaches exhibited high fluctuation in portfolio weights when applied with a 3-month rolling window, indicating instability in weight allocations. As a result, we chose to omit these results from the main report but have included them in the code files for reference.

References

[1] Jonathan Berk and Peter DeMarzo. *Corporate Finance*. Pearson, 6th edition, 2023.

[2] Olivier Ledoit and Michael Wolf. The power of (non-)linear shrinking: A review and guide to covariance matrix estimation. Working Paper 323, University of Zurich, Department of Economics, Zurich, 2020.

[3] Dedi Rosadi, Ezra Putranda Setiawan, Matthias Templ, and Peter Filzmoser. Robust covariance estimators for mean-variance portfolio optimization with transaction lots. *Operations Research Perspectives*, 7:100154, 2020.

[4] Zimo Zhu, Aerambamoorthy Thavaneswaran, Alexander Paseka, Julieta Frank, and Ruppa Thulasiram. Portfolio optimization using a novel data-driven ewma covariance model with big data. In *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, pages 1308–1313, 2020.