MIE377: Financial Optimization Models

Project 2 (Winter 2024)

Group Members: Aarya Jha, Stephanie Lu, and Esther Zhou

Project Summary:

Our project aims to create an algorithmic trading system for a range of 15-40 assets, evaluating performance based on Sharpe ratio and average turnover. Initial models such as Mean Variance Optimization (MVO), CAPM, Risk Parity (RP), and others were trained and validated to identify best-performing models. We also explored MVO variants with dimensionality reduction (LASSO, PCA, BSS), min turnover, and cardinality. Additionally, we explored RP variants with dimensionality reduction. Utilizing optimal parameters obtained from the initial models, we developed a combined model, leveraging RP and PCA for parameter estimation. This model was trained and validated on the three given datasets and tested on three new datasets, showing competitive performance compared to the market capitalization-weighted portfolio when measured against Sharpe ratio and average turnover, making it a good choice as our final model.

Contents

1 Introduction

The purpose of this report is to discuss the process of creating and selecting an algorithmic trading system for financial optimization. The model performance is based on the Sharpe ratio

$$SR_i = \frac{E[r_i] - r_f}{\sqrt{var(r_i - r_f)}} = \frac{\mu_i - r_f}{\sigma_i}$$

and the average turnover

$$\sum |\mathbf{x_0} - \mathbf{x}|$$

where x are the previous weights and x_0 are the current weights. The ideal model would aim to maximize the Sharpe ratio as it indicates the amount of excess return received per unit of risk and minimize the average average turnover as it indicates a less variable portfolio over time.

The report will outline the purpose behind the various strategies implemented and how the Sharpe ratio and average turnover are compared for these strategies. Furthermore, the report will outline how the final strategy was selected.

1.1 Algorithmic Trading System Architecture

We decided to use an algorithmic trading system based on the Risk Parity model, using PCA for parameter estimation (RP-PCA) (see Section 5). During the 5-year calibration period, the model iterates through various values for the following parameters to be used in the model for the remaining investment horizon: **p** - the number of principal components (PCs) to select as factors for PCA and **NumObs** - the number of historical data points to use in the model's parameter estimation. The best parameters are determined based on the Sharpe ratio of each combination of **p** and **NumObs** that was tested (see Figure 1 for the model architecture). This report will further outline our process of selecting and creating this model.

2 Methodology

This section outlines the process we took to formulate our final model.

1. The market capitalization-weighted model was created as our baseline model to compare with the performance of our other models.

2. Then, we analyzed and evaluated simple individual models using Sharpe ratio and average turnover with the training and validation datasets. These models are outlined in Section 3.

3. Next, we analyzed and tuned the parameters for these initial models by analyzing trends in Sharpe ratio and average turnover.

4. Based on this initial analysis of basic individual models, we created compositions of these models and evaluated them with the training and validation datasets to propose a final algorithmic trading system.

5. Based on the analysis of the composition of models, we selected the risk parity model with PCA (RP-PCA) and evaluated its performance on test datasets. RP-PCA has the consistent performance across our training, validation, and testing datasets.

2.1 Data Processing

As part of our methodology, we trained and validated our models outlined in Section 3 with the given datasets to fine-tune its parameters (i.e., lower buy-in threshold, cardinality limit, λ, α, etc.).

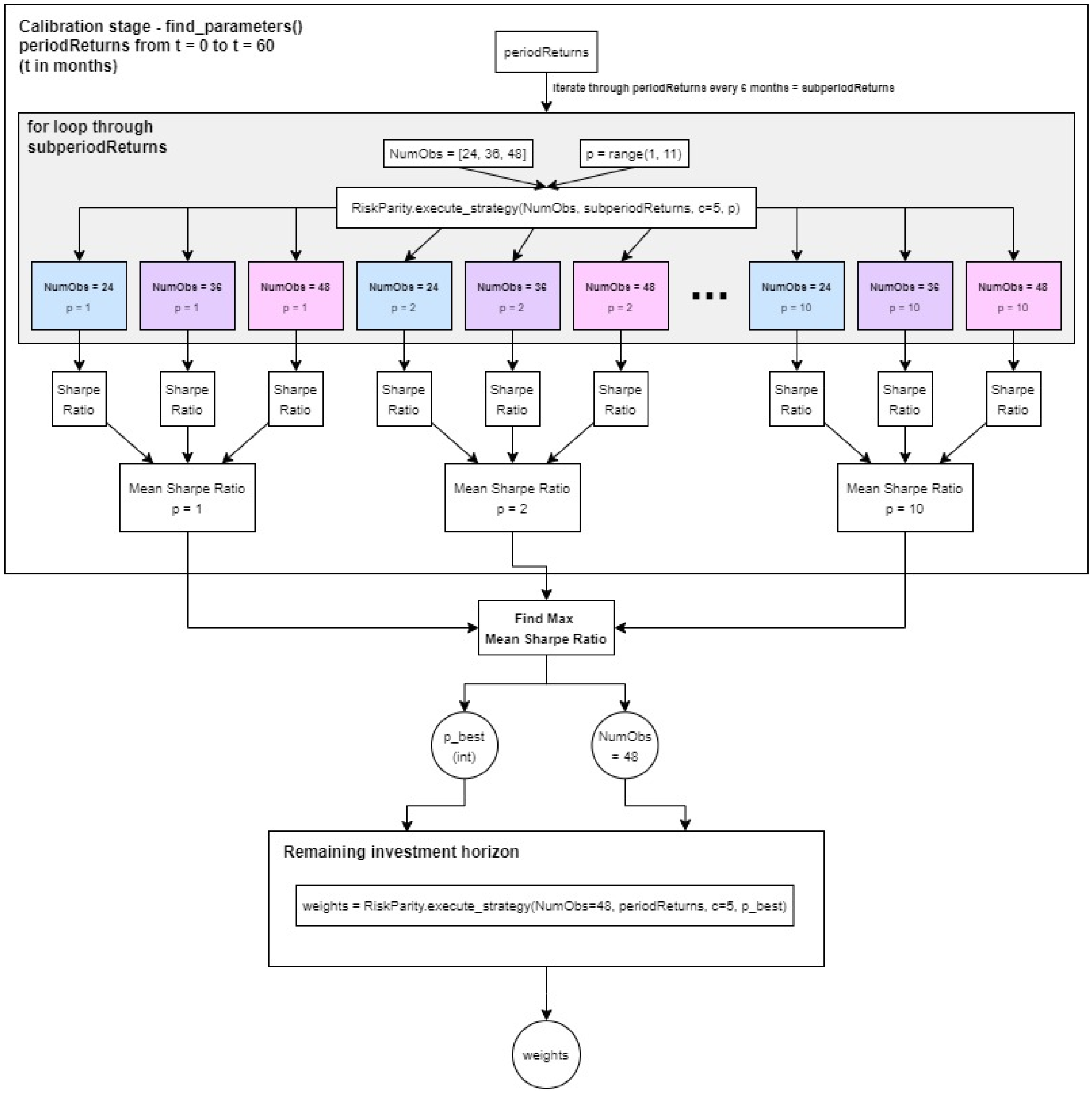


Figure 1: Final model (RP-PCA) architecture (calibration stage and weight calculation process for the remaining investment horizon).

During the testing phase, new datasets were created to test our final model.

To create the new datasets, 3 methods outlined below were implemented to ensure a larger range of time-periods were taken into account.

1. Adjusted closing prices were found for a random set of 35 stocks in the S&P 500 to mimic the original datasets [1]. This data was then used to create "MIE377_AssetPrices_A.csv".

2. A subset of dataset 2 was used to create "MIE377_AssetPrices_B.csv", where a random selection of 15 stocks were chosen from "MIE377_AssetPrices_2.csv".

3. A random selection of 35 stocks were chosen from both "MIE377_AssetPrices_1.csv" and "MIE377_AssetPrices_3.csv" to cover the time period 2002 - 2014. This new dataset is called "MIE377_AssetPrices_C.csv".

Throughout the report, the given datasets, "MIE377_AssetPrices_1.csv", "MIE377_AssetPrices_2.csv", and "MIE377_AssetPrices_3.csv" will be referred to as Dataset 1, Dataset 2, and Dataset 3, respectively. The newly created datasets, "MIE377_AssetPrices_A.csv", "MIE377_AssetPrices_B.csv", and

"MIE377_AssetPrices_C.csv", will be referred to as Dataset A, Dataset B, and Dataset C, respectively.

2.2 Training-Validation-Testing

We decided to use the Dataset 1, 2, and 3 to train and validate our algorithmic trading model.

The training and validation stages were an iterative process. The goal of the training stage is to capture patterns in the data that would allow us to fine-tune the model's parameters, as detailed in Section 2.2.1. We continued to train and validate our initial and combined models on all three datasets. The evaluation metric for both the training and validation stage is the Sharpe ratio and average turnover, with higher weighting given to Sharpe ratio.

We used our results to compare the different models, which will be used to select our final model. Lastly, to test the performance of the final model, we used Dataset A, B, and C to assess how well the model will perform on unseen data.

2.2.1 Parameter Tuning

Using the given dataset and new datasets, we analyzed and tuned the parameters for our different models. We looked at trends in the Sharpe ratio and average turnover with respect to how the model performed with different datasets. This helped us pick the best parameters to use for the models. We then scoped a range of values for each parameter which would be iterated through during the initial 5-year calibration period for the model when the best parameter values would be selected for the remaining investment horizon.

2.3 Baseline Model: The Market Capitalization-Weighted Portfolio

2.3.1 Rationale

We started developing our algorithmic trading system by replicating the returns of the market capitalization-weighted portfolio because data shows that it is very difficult to consistently outperform the market in the long-run [2]. Thus, we used the performance of this replicated market portfolio as a baseline to evaluate the performance of other portfolio optimization models that we explored.

2.3.2 Mathematical Model

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathbf{r}_{\text{mkt}} - \mathbf{R}\mathbf{x}\|_2 \\ \text{subject to} \quad & \mathbf{x} \geq 0 \\ & \mathbf{1}^T\mathbf{x} = 1 \end{aligned}$$

where:

$\mathbf{r}_{mkt} \in {}^{T}\mathbb{R}^1$ is the excess return of the market over T time periods

$\mathbf{R} \in {}^{T}\mathbb{R}^n$ is the excess return of n portfolio assets over T time periods

$\mathbf{x} \in {}^{n}\mathbb{R}^1$ is the portfolio weights for each of the n assets

3 Analysis: Initial Models

We evaluated the performance of individual optimization models on Dataset 1, 2, and 3 to determine which models could be top candidates for our algorithmic trading strategy. Table 1 in Appendix A summarizes the results for each of the individual optimization models we initially analyzed.

The following subsections describe each model in more detail, including a *rationale*, *mathematical model*, relevant *parameters*, and evaluation of its *performance*.

3.1 Equal Weight

As the name suggests, the equal weight portfolio distributes weights evenly across all assets in the portfolio. Compared to other models, the Sharpe ratio can vary depending on the dataset. For example, the Sharpe ratio was quite low compared to other models (i.e. Market Cap, MVO with regularization, and RP), whereas for Datasets 2 and 3 the Sharpe ratio was comparable to other models. This could suggest that the equal-weight portfolio performs well when market has low-moderate volatility. When there is high volatility in the market, other models that take advantage of greater movements in prices can perform better.

The average turnover for the equal-weight portfolio is significantly lower compared to the other models. This makes sense as there is less need for rebalancing, since the weights remain constant over time. Furthermore, because all assets have the same weight, the portfolio is less sensitive to price changes in any individual asset and thus there is less need to buy or sell assets, further reducing average turnover.

3.2 MVO with Historical Data

3.2.1 Mathematical Model

$$
\begin{aligned}
\min_{\mathbf{x}} \quad & \mathbf{x^T Q x} \\
\text{subject to} \quad & \boldsymbol{\mu}^T \mathbf{x} \geq R \\
& \mathbf{1}^T \mathbf{x} = 1 \\
& \mathbf{x} \geq \mathbf{0}
\end{aligned}
$$

where:

$\mathbf{x} \in {}^n\mathbb{R}^1$ is the portfolio weights for each of the n assets
$\mathbf{Q} \in {}^n\mathbb{R}^n$ is the estimated covariance matrix (from historical data)
$\boldsymbol{\mu} \in {}^n\mathbb{R}^1$ is the estimated asset returns (from historical data)
$R \in \mathbb{R}$ is the portfolio target return

Using the model given in the template code to compute the weights, the resulting Sharpe ratio and average turnovers vary from the three datasets. However, they tend to have higher average turnovers since MVO's main goal is to minimize variance which can be achieved by diversification. Therefore placing wealth into more assets and holding onto the wealth over time can lead to higher average turnovers. Furthermore, MVO with historical data can lead to overfitting since it places weights based on historical patterns of assets. If an asset price continues to increase, there is a possibility it may crash in the future and it would be risky to place a high amount of weight in the asset. Hence, since the model did not perform remarkably compared to other models we decided to not proceed with this model.

3.3 MVO with OLS

MVO with OLS uses ordinary linear regression (OLS) to estimate the covariance matrix and returns for a set of assets. A total of 8 factors were used in this initial MVO with OLS model to estimate the aforementioned parameters:

- Market, Profitability, Size, Investment, Value, Momentum, Short-term reversal, and Long-term reversal

3.3.1 Mathematical Model

OLS is an unconstrained optimization problem that aims to minimize the sum of squared residuals.

The mathematical model of OLS is as follows:

$$\min_{\mathbf{B_i}} \quad ||\mathbf{r_i} - \mathbf{X}\mathbf{B_i}||_2^2$$

where $\mathbf{r_i}$ is the returns of asset i, $\mathbf{B_i}$ is the vector of regression coefficients, and $\mathbf{X}$ is the data matrix of factor returns. Expanding the Euclidean norm, we get:

$$\begin{aligned} &\min_{\mathbf{B_i}} \quad (\mathbf{r_i} - \mathbf{X}\mathbf{B_i})^T(\mathbf{r_i} - \mathbf{X}\mathbf{B_i}) \\ = &\min_{\mathbf{B_i}} \quad \mathbf{r_i}^T\mathbf{r_i} - 2\mathbf{r_i}^T\mathbf{X}\mathbf{B_i} + \mathbf{B_i}^T\mathbf{X}^T\mathbf{X}\mathbf{B_i} \end{aligned}$$

The mathematical model of MVO that focuses on minimizing risk is as follows:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{x^T Q x} \\ \text{subject to} \quad & \boldsymbol{\mu}^T\mathbf{x} \geq R \\ & \mathbf{1}^T\mathbf{x} = 1 \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

where:
$\mathbf{x} \in {}^n\mathbb{R}^1$ is the portfolio weights for each of the n assets
$\mathbf{Q} \in {}^n\mathbb{R}^n$ is the estimated covariance matrix (from OLS)
$\boldsymbol{\mu} \in {}^n\mathbb{R}^1$ is the estimated asset returns (from OLS)
$R \in \mathbb{R}$ is the portfolio target return

Using the model given in the template code, the resulting Sharpe ratio and average turnover vary within the three training/validation datasets. MVO-OLS performed consistently with the baseline model for Dataset 1, but performed noticeably better for Dataset 2 and 3 (i.e., $> 5\%$).

With OLS, the objective is to minimize the residual sum of squares (RSS). However, this can be prone to overfitting, especially when there are a large number of predictor variables compared to the number of observations.

To further investigate the OLS method, we implemented this model in different ways.

- Selecting a subset of factors (CAPM, Fama-French, Carhart) in Section 3.4
- Reducing the dimensionality of the problem using different techniques (Lasso, BSS, and PCA) is further detailed in Section 3.5.1, 3.5.2, and 3.5.3

- Adding additional cardinality constraints to the MVO problem

Through implementing these models, we hoped to reduce overfitting, explain some of the variation between factors, and improve the interpretability of the model.

3.4 MVO with Common Factor Models

We compared different types of common factor models, including **CAPM**, **Fama-French-3-Factor**, and **Fama-French-5-Factor**. Because these factor models employ different types of factors that explain various characteristics of the assets, we felt that this is a good starting point to analyze which factors are the most important and will generalize well with other datasets.

From Table 1 in Appendix A, we can see that the multi-factor models resulted in a high Sharpe Ratio for Datasets 1, 2, and 3. According to Horstmeyer et. al, the RMW factor, which is included in the FF5 model, has continued to deliver excess returns since 1963, which indicates the value in investing in high quality stocks or profitable firms over low quality, unprofitable counterparts [3]. This suggests that the two extra factors in FF5 (profitability and investment) could be useful additions to include in a multi-factor model.

When evaluating Sharpe ratio with CAPM, we found that its performance against other models varied for different datasets. Although CAPM is a simple and well-known model, it has several limitations. For example, CAPM relies only on market risk to explain excess returns, but there are many other factors, such as size, value, profitability, and more than also play a significant role in determining an asset's returns. Due to the variability in our analysis we could not conclude whether a single-factor or multi-factor model performed better. Thus, we decided to test out factor models with different factors using dimensionality reduction to find the best ones.

3.5 MVO with Dimensionality Reduction

In reality there are over 300 factors to choose from when working with factor models. To include all factors would lead to overfitting and use large computational power. Thus, to promote model generalization and sparsity, different techniques, such as regularization (Lasso, BSS) and Principal Component Analysis (PCA), were used. Regularization can be used to promote sparsity and reduce statistical overfitting in factor models.

We decided to analyze how different techniques will impact our portfolio weights using MVO. The three methods we will consider are:

- Lasso
- BSS
- PCA

3.5.1 MVO with LASSO Regression

3.5.1.1 Mathematical Model

LASSO regression promotes generalization by adding a penalty term to the OLS objective function, which can help address to issue of overfitting in regular OLS. The penalty term makes some of the coefficients for the factors to be zero, effectively removing them from the model. The mathematical model can be written as follows:

$$
\begin{aligned}
\min_{x} \quad & \|\mathbf{r}_i - \mathbf{X}\mathbf{B}_i\|_2{}^2 \\
\text{subject to} \quad & \|\mathbf{B_i}\|_2 \leq S
\end{aligned}
$$

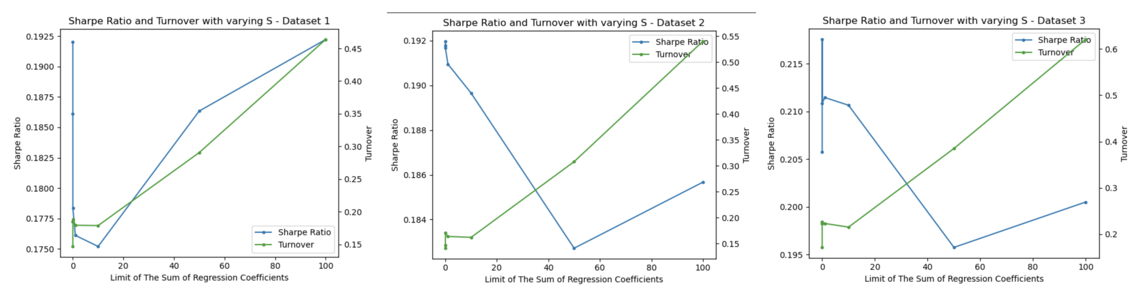
where S is the limit that the sum of the regression coefficients can sum up to.

3.5.1.2 Parameter: Sum of Regression Coefficients - S

We tested different values of S, refer to Table 1, in the constraint to see which value would give the best Sharpe ratio and average turnover for different datasets. Figure 2 illustrates that a lower S tends to have higher Sharpe ratio and lower turnover. With larger S (i.e. S = 50 and S = 100), tend to have lower Sharpe and higher turnover.

S	0.00001	0.001	0.01	1	10	50	100

Table 1: Parameters tested for varying S limits in constraint



(a) (b) (c)

Figure 2: Sharpe ratio and average turnover of the MVO with LASSO for Dataset 1 (a), 2 (b), and 3 (c) with varying S.

Table 2: Best parameters U and K for each dataset with respective Sharpe ratios and average turnovers

Model		**Dataset 1**	**Dataset 2**	**Dataset 3**
MVO-LASSO	**S**	0.001	0.01	1
	Sharpe	0.1861	0.1917	0.2176
	Turnover	0.1851	0.1465	0.227
Baseline	**Sharpe**	0.196	0.113	0.157
	Turnover	0.54	0.59	0.60

MVO-LASSO performs average to above-average compared to baseline model when measured against Sharpe ratio as shown in Table 2. It performs significantly better than baseline when measured against average turnover, with the ability to decrease turnover by more than 50%.

In general, by setting a smaller S value, this can help prevent overfitting by constraining what the regression coefficients can sum up to. This then reduces complexity and the variance of the model, as it prefers simpler models with smaller coefficients. We further tested whether this is the case when we combined Risk Parity with Lasso in section 4.2.2.

3.5.2 MVO with BSS

Sparsity during optimization is a data-driven approach to the factor selection process for the regression model; all factors are considered, but optimization selects the best-fitting subset.

The Best Subset Selection (BSS) regression model uses the l_0 norm while the LASSO regression model and the Ridge Regression model use the l_1 and l_2 norm respectively. All 3 models provide advantages such as promoting true sparsity (i.e., BSS and LASSO) or having a continuous and convex least-squares objective (i.e., LASSO and Ridge Regression). However, for BSS, the l_0 norm behaves like a cardinality constraint that allows us to control the number of non-zero elements (i.e., the number of factors to select for OLS). Thus, we decided to further analyze the performance of an MVO model with BSS since we can more explicitly control sparsity.

3.5.2.1 Mathematical Model

The l_0 norm is as follows:

$$\|\mathbf{B_i}\|_0 = \mathbb{1}_{\alpha_i \neq 0} + \sum_{k=1}^{p} \mathbb{1}_{\beta_{ik} \neq 0}$$

The constrained version of the BSS model is as follows:

$$\begin{aligned} \min_{\mathbf{B_i}} \quad & \|\mathbf{r_i} - \mathbf{X}\mathbf{B_i}\|_2^2 \\ \text{subject to} \quad & \|\mathbf{B_i}\|_0 \leq K \end{aligned}$$

The l_0 norm is discontinuous and non-convex. However, it is equivalent to a cardinality limit. Thus, BSS can be formulated as a mixed-integer problem:

$$\begin{aligned} \min_{\mathbf{B}_i} \quad & \|\mathbf{r}_i - \mathbf{X}\mathbf{B}_i\|_2^2 \\ \text{subject to} \quad & L\mathbf{y} \leq \mathbf{B}_i \leq U\mathbf{y} \\ & \mathbf{1}^T\mathbf{y} \leq K \\ & y_j \in \{0, 1\},\ j = 1, \ldots, p+1 \end{aligned}$$

where:

$\mathbf{r}_i \in {}^T\mathbb{R}^1$ is the returns of asset i
$\mathbf{X} = [\mathbf{1}\ \mathbf{f}] \in {}^T\mathbb{R}^{p+1}$ is the data matrix where $\mathbf{f} \in {}^T\mathbb{R}^p$ is the factor returns
$\mathbf{B}_i \in {}^{p+1}\mathbb{R}^1$ is the vector of regression coefficients
$L \in \mathbb{R}$ and $U \in \mathbb{R}$ are the lower and upper bounds of the regression coefficients
$K \in \mathbb{Z}_+$ is the cardinality of BSS

3.5.2.2 Parameter: U and K

Table 3: Variable parameters for MVO-BSS

Parameter	Range
L	[0]
U	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
K	[1, 2, 3, 4, 5, 6, 7]

After testing the MVO with BSS model using Dataset 1, 2, and 3, Figure 3 shows common patterns with respect to the behaviour of Sharpe ratio and average turnover as U and K vary.

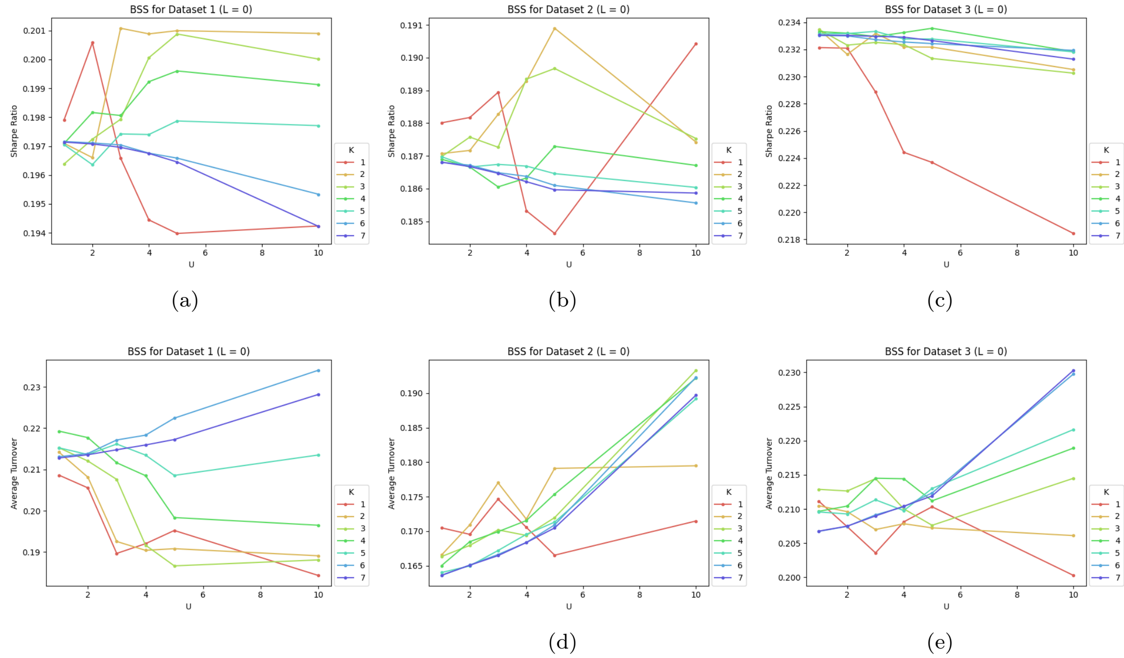


Figure 3: (a)-(c) Sharpe ratio and (d)-(e) average turnover of the MVO with BSS strategy for Dataset 1, 2, and 3. U and K are varied while L is fixed at $L = 0$

In general, a lower K results in a more noisy Sharpe ratio and average turnover trend as U increases, but the trends eventually become smoother at high U. $K = 1$ tends to be very noisy and while it might have very high Sharpe ratios for particular values of U, it will also have very low Sharpe ratios for other values of U. This effect is not as prominent for average turnover, but from Figure 3, it is evident that $K = 1$ also produces noise average turnover results that tend to differ from the general trend of the other K values.

To varying degrees for the Sharpe ratio trends at each K, Sharpe ratio increases as U increases until a maximum, then begins to plateau or slightly decrease as U continues to increase. In general, Sharpe ratio is higher for lower values of K, but all the trend lines are clustered together fairly closely.

For average turnover, there appears to be an opposite effect. As U increases, average turnover decreases until a minimum, then begins to plateau or increase as U continues to increase. In general, average turnover is lower for lower values of K.

Table 4: Sharpe ratio and average turnover of MVO-BSS for best L, U, and K

		Dataset 1	**Dataset 2**	**Dataset 3**
Parameters	L	0	0	0
	U	3	5	5
	K	2	2	4
BSS	**Sharpe**	0.2011	0.1909	0.2336
	Turnover	0.1926	0.1791	0.2112
Baseline	**Sharpe**	0.1958	0.1133	0.1574
	Turnover	0.5376	0.5949	0.6009

From Table 4, it is evident that for the best set of parameters for each dataset, the MVO-BSS strategy outperforms the baseline model for Dataset 1, 2, and 3 in terms of both Sharpe ratio and average turnover. It is also important to note that since Sharpe ratio and average turnover do not vary significantly with changes in the parameters within the range that we analyzed, the MVO-BSS strategy outperforms the baseline model for all combinations of these parameters.

3.5.3 MVO with PCA

Using Principal Component Analysis (PCA) to select principal components (PCs) rather than using factors from a traditional factor model can offer several advantages. PCA allows for the reduction of the dimensionality of the dataset by transforming the original variables into a smaller set of uncorrelated factors that capture the maximum variance in the data. By selecting factors based on these principal components, we can extract meaningful information to more accurately estimate parameters for MVO.

Furthermore, PCA can be relatively robust to outliers and noise in the data compared to some other factor methods. It focuses on capturing the overall structure of the data rather than individual data points, making it less sensitive to extreme values.

3.5.3.1 Mathematical Model

To extract factors to be used in OLS from PCA, the following mathematical model was used:

$$\mathbf{P} = \overline{\mathbf{R}}\mathbf{\Gamma}$$

where:

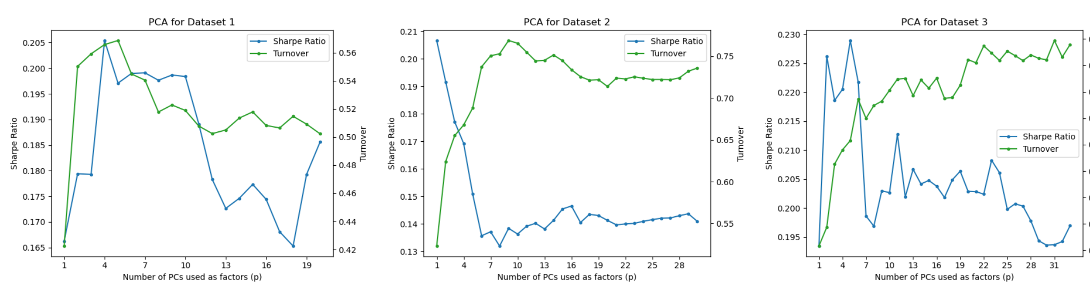
$\mathbf{\Gamma}$ is from the eigendecomposition of the biased covariance matrix, $\mathbf{Q}_{biased} = \frac{1}{T}\overline{\mathbf{R}}^T\overline{\mathbf{R}} = \mathbf{\Gamma\Lambda\Gamma}^T$

$\overline{\mathbf{R}}$ is from centering the data matrix of observed asset returns $\mathbf{R}$, $\overline{\mathbf{R}} = \mathbf{R} - 1(\frac{1}{T}\mathbf{R}^T 1)^T$

Since $\mathbf{Q}_{biased}$ is a symmetric matrix, eigenvalues and eigenvectors are real; any small complex components that result from Python were ignored, as indicated in the code.

Furthermore, to ensure that OLS is feasible, we ensured that the number of factors selected (p) satisfied $T - p - 1 > 0$.

3.5.3.2 Parameter: Number of PCs Selected as Factors (p)



(a) (b) (c)

Figure 4: Sharpe ratio and turnover with respect to the number of PCs selected as factors for (a) Dataset 1, (b) Dataset 2, and (c) Dataset 3.

After testing the PCA model with the 3 different datasets, Figure 4 shows that the Sharpe ratio and turnover follow a similar trend for each dataset. Sharpe ratio first increases as p increases, then Sharpe ratio decreases sharply and plateaus. Turnover first increases as p increases, then plateaus. The Sharpe ratio reaches a maximum at $p = 4$ for Dataset 1 and $p = 5$ for Dataset 3, whereas it reaches a maximum at $p = 1$ for Dataset 2. This is reasonable because selecting too many PCs could result in overfitting; by selecting only a small subset of PCs, we are thus able to reduce the dimensionality of the dataset while preserving as much of the most significant variance as possible.

Table 5: Sharpe ratio and average turnover of MVO-PCA with best p

		Dataset 1	**Dataset 2**	**Dataset 3**
Parameters	p	4	1	5
MVO-PCA	**Sharpe**	0.2055	0.2066	0.2290
	Turnover	0.5660	0.5230	0.6181
Baseline	**Sharpe**	0.196	0.113	0.157
	Turnover	0.54	0.59	0.60

From Table 3.4, we can see that PCA results in a higher Sharpe ratio than the other common factor models for Dataset 1, 2, and 3.

From Table 5, it is evident that the MVO-PCA model outperforms the baseline model for Dataset 1, 2, and 3 in terms of Sharpe ratio when the best p is selected. However, this comes at the cost of a high average turnover that is comparable to the average turnover of the baseline model, unlike the MVO-BSS model which had relatively lower average turnover. It is also interesting to note that the Sharpe ratios for the MVO-PCA model vary a lot more as p varies than the Sharpe ratios for the MVO-BSS model as U and K vary. This means that the MVO-PCA model does not outperform the baseline model for all possible parameters p.

3.5.4 General Takeaways of Reducing Dimensionality

When we combined regularization with MVO we found some general trends for all methods, summarized as follows:

1. We found that with regularization Sharpe ratio performed better than baseline for most cases. For cases where Sharpe ratio was lower than baseline model, it was not significantly lower. This suggests that adding regularization helps generalize our model so that it does not overfit to a specific dataset.

2. Lasso and BSS helped significantly reduce average turnover by over 50% compared to the baseline. However, the average turnover for MVO-PCA remained relatively the same compared to baseline. This can be explained by the fact that PCA relies on a reduced set of principal components, which can be more sensitive to changes in the underlying data and thus lead to frequent rebalancing of weights. On the other hand, Lasso and BSS promotes sparsity, which tends to lead to more stable models.

3.5.4.1 Portfolio Weights

Figure 5 illustrates the benefits of reducing dimensionality and adding sparsity to the model. The weights are more evenly distributed across a wider range of assets and remain relatively constant over time, leading to lower average turnovers and more diverse portfolio, compared to the baseline model.

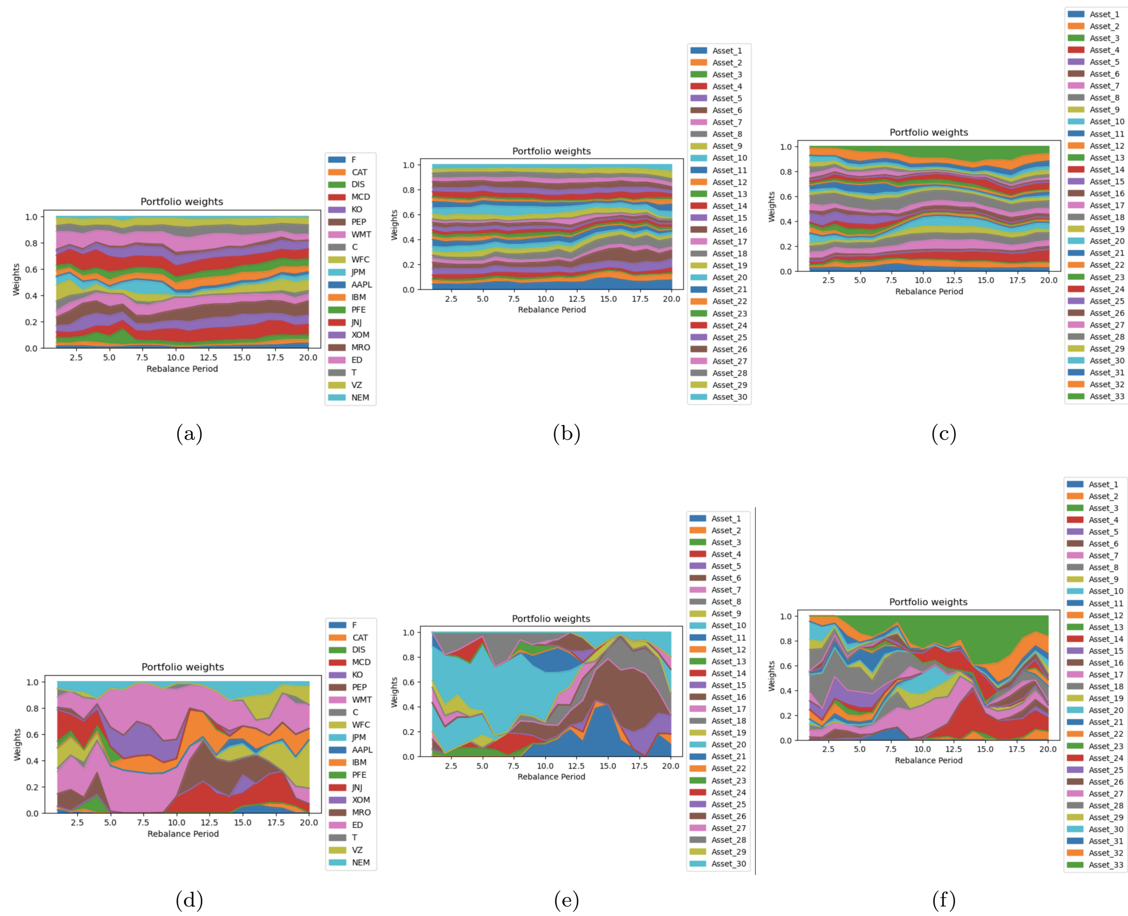


Figure 5: Portfolio weights with dimensionality reduction and MVO using Dataset 1(a), 2(b), and 3(c). Portfolio weights of baseline model for Dataset 1(d), 2(e), and 3(f).

3.6 MVO with Cardinality Constraints (MVO-CC)

MVO was implemented with cardinality constraints, specifically a lower (L) and upper (U) buy-in threshold for each asset, and a cardinality limit on the total number of assets (K).

3.6.1 Mathematical Model

The mathematical model can be written as follows:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{x^T Q x} \\ \text{subject to} \quad & \boldsymbol{\mu}^T \mathbf{x} \geq \mathbf{R} \\ & \mathbf{1}^T \mathbf{x} = 1 \\ & \mathbf{1}^T \mathbf{y} \leq K \\ & x_i \geq L_i y_i \quad \text{for } i = 1, \ldots, n \\ & x_i \leq U_i y_i \quad \text{for } i = 1, \ldots, n \\ & y_i \in \{0, 1\} \quad \text{for } i = 1, \ldots, n \\ & x_i \geq 0 \quad \text{for } i = 1, \ldots, n \end{aligned}$$

3.6.2 Parameter: U and K

As outlined in Appendix B, from Project 1, we found that the best parameter for lower buy-in threshold was $L = 0.05$.

Iterating through U's and K's, we found that the best values for these parameters vary between datasets. A summary of best parameters along with the Sharpe ratio and average turnover is summarized in Table 6. The best parameters were chosen based off of highest Sharpe ratio summarized in Appendix B.3.2, since more weighting is given to higher Sharpe ratio compared to lower average turnover.

Table 6: Best parameters U and K for each dataset with respective Sharpe ratios and average turnovers

Model		Dataset 1	Dataset 2	Dataset 3
MVO-CC	**U**	0.25	0.8	0.2
	K	20	13	15
	Sharpe	0.1823	0.181	0.2355
	Turnover	0.54	0.67	0.55
Baseline	**Sharpe**[1]	0.196	0.113	0.157
	Turnover	0.54	0.59	0.60

We found that MVO-CC gave an average to high Sharpe ratio compared to the baseline model. However, compared to average turnover, MVO-CC performed relatively the same compared to the baseline model. To try and minimize the average turnover, we implemented MVO-CC with a min turnover term in the objective function detailed in section 4.1.3.

3.7 Risk Parity (RP)

Risk parity is a type of optimization approach that allows us to maximize risk diversification through equalizing the risk contribution per asset.

[1]Baseline model results

3.7.1 Mathematical Model

The mathematical model for Risk Parity Optimization can be written as follows:

$$
\begin{aligned}
\min_{\mathbf{y}} \quad & \frac{1}{2}\mathbf{y}^{\mathbf{T}}\mathbf{Q}\mathbf{y} - c\sum_{i=1}^{n}\ln(y_i) \\
\text{subject to} \quad & \mathbf{y} \geq 0
\end{aligned}
$$

To find the weights of the portfolio $\mathbf{x}$, we can normalize $\mathbf{y}$ as follows:

$$
x_i^* = \frac{y_i^*}{\sum_{i=1}^{n} y_i}
$$

Overall, we found that risk parity performs average to above-average across different market conditions with varying volatility, summarized in Table 7, when measured against Sharpe ratio. For example in dataset 2, where we have low volatility compared to Dataset 3, where we have high volatility, risk parity outperforms the baseline model. In the case of Dataset 1 and 3, where there is moderate-high volatility, there is more opportunity for returns of assets to be significantly greater or lower than their expected returns. Thus, we can exploit the assets that would result in higher portfolio returns. However, a downside of this is that there are more chances for greater loss as well.

Risk parity performs significantly better in terms of minimizing turnover compared to the baseline. This could be because of the diversification benefits of risk parity. Because risk parity ensures the risk contribution of all assets is equal, there would be some weight assigned to each asset, compared to market capitalization-weighted portfolios which can have weights heavily concentrated in assets that dominate the market. As a result, risk parity can reduce the impact of extreme movements in the market, which is an advantage of this model.

Other advantages of the risk parity model can be summarized as follows:

1. Risk parity leverages less volatile assets to achieve higher returns without proportionately increasing risk. This can enhance returns relative to risk and improve the Sharpe ratio.

2. We do not need to use estimated expected returns, which are harder to estimate and more prone to error.

Table 7: Sharpe ratios and average turnovers of RP compared to Baseline Model

Model	**Metric**	**Dataset 1**	**Dataset 2**	**Dataset 3**
	Volatility [2]	0.0233	0.00929	0.0306
RP	**Sharpe**	0.1847	0.1996	0.2080
	Turnover	0.1538	0.1284	0.1935
Baseline	**Sharpe**	0.196	0.113	0.157
	Turnover	0.54	0.59	0.60

[2] The number presented in the table is representative of the standard deviation of the volatilities of all assets in the given dataset

3.8 Mean-Variance Tracking Error (MV-TE)

Index tracking is an example of a passive investment strategy. Under the assumption that the market is very difficult or even impossible to outperform in the long run, index tracking would provide us with a portfolio that behaves like the market but with fewer constituent assets.

One way to track an index is to minimize the variance-based tracking error (TE). TE is a variance measure because it depends on $\mathbf{Q}$:

$$TE(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_{mkt})^T \mathbf{Q}(\mathbf{x} - \mathbf{x}_{mkt})$$

where:
$\mathbf{x} \in {}^n\mathbb{R}^1$ is the portfolio weights for each of the n assets
$\mathbf{x}_{mkt} \in {}^n\mathbb{R}^1$ is the weights of the benchmark portfolio (i.e., the market portfolio)
$\mathbf{Q} \in {}^n\mathbb{R}^n$ is estimated covariance matrix

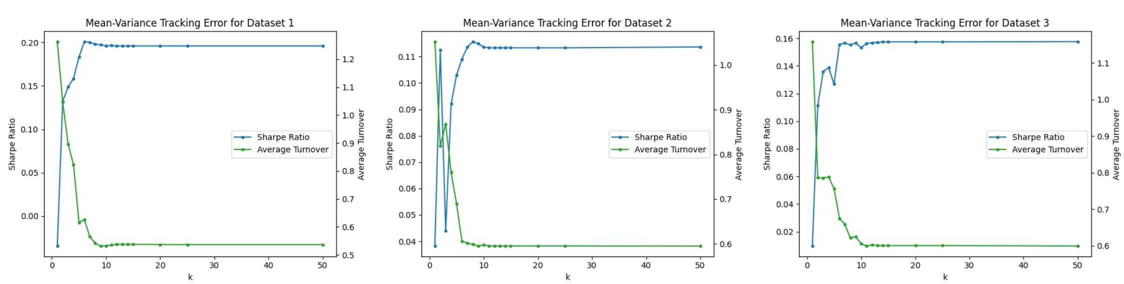
3.8.1 Mathematical Model

Mean-variance tracking error is similar to cardinality-constrained MVO with $k \leq n$ constituents.

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y}} \quad & (\mathbf{x} - \mathbf{x}_{\text{mkt}})^T \mathbf{Q}(\mathbf{x} - \mathbf{x}_{\text{mkt}}) \\ \text{subject to} \quad & \boldsymbol{\mu}^T \mathbf{x} \geq \boldsymbol{\mu}^T \mathbf{x}_{\text{mkt}} \\ & \mathbf{1}^T \mathbf{x} = 1 \\ & \mathbf{1}^T \mathbf{y} \leq k \\ & \mathbf{x} \leq \mathbf{y} \\ & \mathbf{x} \geq 0 \\ & \mathbf{y} \in \{0, 1\}^n \end{aligned}$$

where:
$\mathbf{x} \in {}^n\mathbb{R}^1$ is the portfolio weights for each of the n assets
$\mathbf{x}_{mkt} \in {}^n\mathbb{R}^1$ is the weights of the benchmark portfolio (i.e., the market portfolio)
$\mathbf{Q} \in {}^n\mathbb{R}^n$ is estimated covariance matrix
$\boldsymbol{\mu} \in {}^n\mathbb{R}^1$ is estimated asset returns
$k \in \mathbb{R}$ is the number of constituents in the tracking portfolio

3.8.2 Parameter: k



(a) (b) (c)

Figure 6: Sharpe ratio and average turnover of the Mean-Variance Tracking Error strategy for Dataset (a) 1, (b) 2, and (c) 3.

After testing the mean-variance tracking error model using Dataset 1, 2, and 3, Figure 6 shows that all 3 datasets follow a similar trend. As k initially increases, Sharpe ratio also increases until it reaches a maximum, then plateaus for further increasing k whereas average turnover first decreases then plateaus. Sharpe ratio is maximized at $k = 6$, $k = 8$, and $k = 7$ for the 3 datasets respectively. This is reasonable because beyond a certain number of constituents in the portfolio, increasing the number of assets does not significantly alter its ability to track the market portfolio since the most representative assets have already been captured.

The maximum Sharpe ratio achieved by the mean-variance tracking error model almost identically mimics the market-capitalization weighted portfolio (see Section 2.3). Overall, this model does not appear to provide significant advantages about the objective of maximizing the Sharpe ratio and minimizing average turnover.

Table 8: Sharpe ratio and average turnover of Mean-Variance Tracking Error for best k

Model	**Metric**	**Dataset 1**	**Dataset 2**	**Dataset 3**
Parameters	k	6	8	50
MV-TE	**Sharpe**	0.2008	0.1157	0.1576
	Turnover	0.6254	0.5981	0.6000
Baseline	**Sharpe**	0.196	0.113	0.157
	Turnover	0.54	0.59	0.60

3.9 CVaR

CVaR is a risk measure that assesses the expected loss that would occur if the worst-case scenario threshold for variance is ever crossed. CVaR quantifies the expected loss that would occur past the VaR break-point. We aim to minimize only the downside risk. Historical scenarios are often used to model the risk distribution of the assets.

3.9.1 Mathematical Model

$$\begin{aligned} \min_{\mathbf{x},\mathbf{z},\gamma} \quad & \gamma + \frac{1}{(1+\alpha)S} \cdot \sum_{s=1}^{S} \mathbf{z_s} \\ \text{subject to} \quad & \mathbf{z_s} \geq 0, s = 1, 2...., S \\ & \mathbf{z_s} \geq \mathbf{f}(\mathbf{x}, \hat{\mathbf{r_s}}) - \gamma, s = 1, 2, ..., S \\ & x \in X \end{aligned}$$

In the model above γ is the loss , $\mathbf{x}$ is the vector of portfolio weights $\mathbf{r}$ is the vector of random asset returns $\mathbf{f}(\mathbf{x}, \hat{\mathbf{r_s}})$ is the loss of portfolio for a realization of our random asset return, z_s is an auxiliary variable and α represents the probability we are accounting for. The model performed relatively well compared to the other models we tried. Compared to other models implemented, the Sharpe ratio obtained is relatively average. Hence, CVaR has been effective in reducing average turnover due to it being a risk measure.

3.9.2 Distributionally Robust CVaR

A variation of CVaR is the Distributionally Robust CVaR. This model was inspired by $skfolio$ [4]. This model constructs a Wasserstein ball in the space of multivariate and non-discrete probability distributions. This is centered on the uniform distribution of the training samples. It also finds the allocation that minimizes the CVaR of the worst-case distribution within this Wasserstein ball.

Table 9: Sharpe ratio and average turnover of CVaR

Model	Metric	Dataset 1	Dataset 2	Dataset 3
CVaR	**Sharpe**	0.1759	0.1917	0.2011
	Turnover	0.1290	0.1067	0.1510
Baseline	**Sharpe**	0.2336	0.1733	0.1882
	Turnover	0.4532	0.4683	0.4901

The mathematical formulation of this problem is as follows:

$$\min_{\mathbf{x},\mathbf{z},\gamma} \quad \gamma + \frac{1}{(1+\alpha)S} \cdot \sum_{s=1}^{S} \mathbf{z_s}$$

$$\begin{aligned} \text{subject to} \quad & \mathbf{z_s} \geq 0, s = 1, 2...., S \\ & \mathbf{z_s} \geq \mathbf{f}(\mathbf{x}, \hat{\mathbf{r_s}}) - \gamma, s = 1, 2, ..., S \\ & \mathbf{E}(\hat{\mathbf{r}}) - \text{VaR}(\mathbf{f}(\mathbf{x}, \hat{\mathbf{r_s}})) \leq \beta \text{ where } \beta \text{ is ball radius} \\ & x \in X \end{aligned}$$

The model provided a very low average turnover, however, due to the ball constraint, it provided a much lower Sharpe ratio than average for each of the datasets. For this reason, the model was not selected since it is the team's priority to maximize the Sharpe ratio as much as possible. Furthermore, other models also provided comparatively low average turnovers with higher Sharpe ratios. If the objective of the model was to only minimize turnover, it would be worth considering this model.

Table 10: Sharpe ratio and average turnover of Distributionally Robust CVaR ($\alpha = 0.95$ and $radius = 0.5$)

Model	Metric	Dataset 1	Dataset 2	Dataset 3
Distributionally Robust	**Sharpe**	0.1487	0.1660	0.2011
CVaR	**Turnover**	0.0786	0.08162	0.1446
Baseline	**Sharpe**	0.2336	0.1733	0.1882
	Turnover	0.4532	0.4683	0.4901

Parameters alpha and the radius were varied but provided very little change in Sharpe and turnover values.

3.10 Monte Carlo

The Monte Carlo model is a model which can help predict unknown events. This is can be done through a geometric or arithmetic random walk, where asset values are normally distributed and follow a Gaussian process. Hence they can be changed by a small delta value that follows this metric. In the Monte Carlo process the first step is to estimate returns and variance, then select several time steps, we then repeat this process multiple times. Our version of Monte Carlo was tested on generating 100, 1000, and 10000 samples however they varied slightly in terms of performance. The sample size of 10000 seemed to generate on average better Sharpe and Turnover values. However, compared to other models this did not perform remarkably hence why it was not considered. Due to the randomness of this model, although it is less likely to overfit, it does provide results less favorable than other models.

Table 11: Sharpe ratio and average turnover of Monte Carlo (sample size = 10000)

Model	Metric	Dataset 1	Dataset 2	Dataset 3
Monte Carlo	**Sharpe**	0.1700	0.2057	0.1894
	Turnover	0.5437	0.3638	0.5474
Baseline	**Sharpe**	0.2336	0.1733	0.1882
	Turnover	0.4532	0.4683	0.4901

3.10.1 Monte Carlo with CVaR

The Monte Carlo method was also implemented with CVaR. Keeping an $\alpha = 0.95$, the sample size value of 10000 performed better, compared to lower values in terms of Sharpe ratio and average turnover.

Table 12: Sharpe ratio and average turnover of Monte Carlo with CVaR $\alpha = 0.95$ and sample size = 10000

Model	Metric	Dataset 1	Dataset 2	Dataset 3
Monte Carlo with CVaR	**Sharpe**	0.2037	0.1991	0.1979
	Turnover	0.4077	0.3650	0.5102
Baseline	**Sharpe**	0.2336	0.1733	0.1882
	Turnover	0.4532	0.4683	0.4901

Compared to just the Monte Carlo model, with CVaR the turnover was reduced for Dataset 1 and 3 and the Sharpe Ratio increased. However, for Dataset 2 the opposite effects occurred but the increase in Sharpe and decrease in turnover were slight. This can be attributed to the randomness of Monte Carlo. However, effects from CVaR can be visible as well. CVaR, aims to minimize extreme loss, providing a reason for the changes noted above. Dataset 2 could likely have had more riskier weightings in Monte Carlo, that with CVaR was prevented. Regardless, this model was not considered since other models in the report provided Sharpe values in the same range for much lower turnover values.

4 Analysis: Combining Initial Models

After we analyzed the initial individual optimization models, we decided that there were 2 major items we wanted to further investigate:

1. Targeting **average turnover** because many of our models with the highest Sharpe ratios would also have the highest average turnovers.

2. The **risk parity** model because it has very consistent performance across all datasets.

To investigate these 2 items, we evaluated different combinations of our initial models to determine whether these combinations would result in better Sharpe ratios and average turnovers.

4.1 Min Turnover (MVO-T)

A common theme that we observed in our initial models was that they had consistently high turnover rates. Thus, we decided to add an average turnover term to our MVO formulation to attempt to minimize average turnover.

We re-formulated MVO as follows:

$$
\begin{aligned}
\min_{\mathbf{x},\mathbf{z}} \quad & \mathbf{x}^T\mathbf{Q}\mathbf{x} + \lambda \sum_{i=1}^{n} z_i \\
\text{subject to} \quad & \mathbf{x} - \mathbf{x_0} \leq \mathbf{z} \\
& \mathbf{x} - \mathbf{x_0} \geq -\mathbf{z} \\
& \mathbf{z} \geq \mathbf{0}
\end{aligned}
$$

where:
$\mathbf{x} \in {}^n\mathbb{R}^1$ is the portfolio weights for each of the n assets
$\mathbf{x}_0 \in {}^n\mathbb{R}^1$ is $\mathbf{x}$ in the previous rebalancing period
$\mathbf{Q} \in {}^n\mathbb{R}^n$ is the estimated covariance matrix
$\lambda \in \mathbb{R}$ is a penalty factor for average turnover
$\mathbf{z} \in {}^n\mathbb{R}^1$ is an auxiliary variable

And other relevant constraints as described in the following subsections.

We evaluated the performance of this formulation of MVO, which will hence be referred to as MVO-T, on the best performing individual models from Section 3:

- MVO-T with BSS
- MVO-T with PCA
- MVO-T with Cardinality Constraints

4.1.1 MVO-T with BSS

Figure 7 shows that with the additional turnover term in the MVO objective, there is a significant decrease in average turnover, which is now consistently on the order of 0.001. However, the impact of MVO-T on Sharpe ratio is variable across Dataset 1, 2, and 3. For Dataset 1, the MVO-T Sharpe ratios are comparable to the Sharpe ratios without the turnover term. However, for both Dataset 2 and 3, there is a noticeable decrease in Sharpe ratio ($\leq 1\%$ for Dataset 2 and $\approx 4\%$ for Dataset 3. While the MVO-T with BSS model still outperforms the baseline model for almost all parameter combinations evaluated, it is evident that there is a tradeoff that must be made between Sharpe ratio and average turnover.

Table 13: Sharpe ratio and average turnover of MVO-T with BSS for best U and K

		Dataset 1	**Dataset 2**	**Dataset 3**
$\lambda = 1$	U	5	10	1
	K	2	6	1
	Sharpe	0.2016	0.1840	0.1958
	Turnover	0.0036	0.0021	0.0010
Baseline	**Sharpe**	0.1958	0.1133	0.1574
	Turnover	0.5376	0.5949	0.6009

From Figure 8, it is evident that even a very small λ has a significant impact on average turnover, with average turnover on the order of 0.001 for all values of λ evaluated. However, there are no clear trends with respect to the effect of λ on Sharpe ratio. For Dataset 1, Sharpe ratio is not

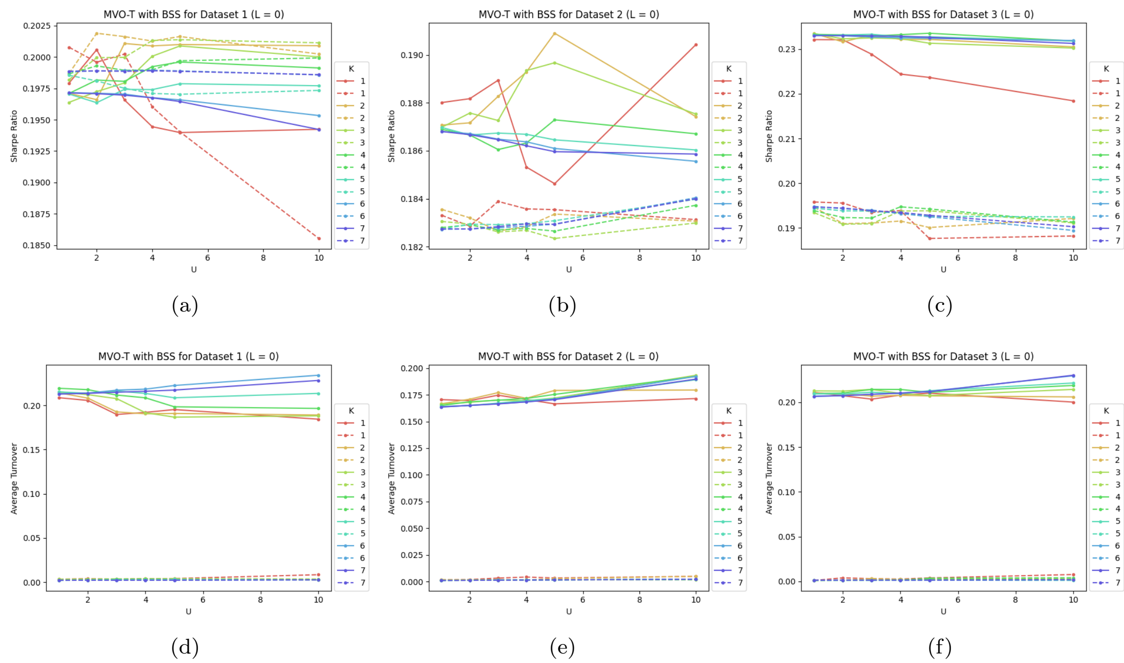


Figure 7: (a)-(c) Sharpe ratio and (d)-(f) average turnover with respect to U for varying K using MVO-T with BSS. Note that the solid lines indicate that MVO was used whereas dashed was MVO-T ($\lambda = 1$).

significantly impacted for $\lambda \leq 40$, but drops by $\geq 1\%$ for larger λ. For Dataset 2, Sharpe ratio is relatively unaffected by λ. For Dataset 3, an opposite effect to Dataset 1 is observed, where Sharpe ratio is not significantly impacted for $\lambda \geq 5$, but drops by $\approx 1\%$ for small λ. Thus, it appears that it will be difficult to predict the impact of λ on the Sharpe ratio of a dataset if we selected the MVO-T with BSS algorithmic trading strategy.

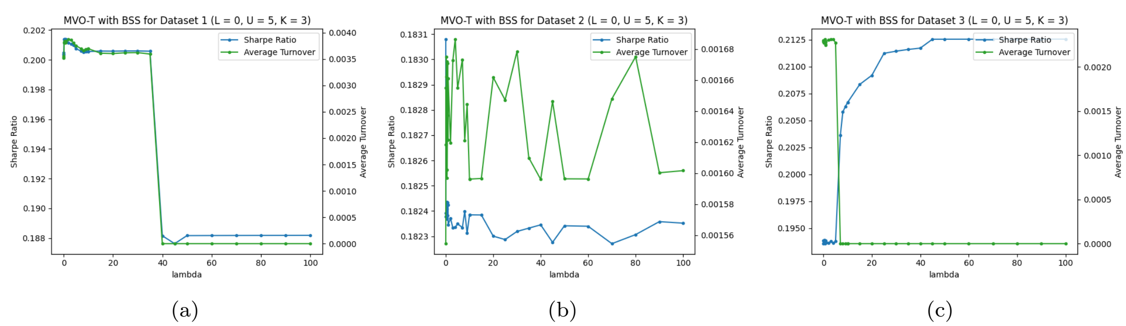
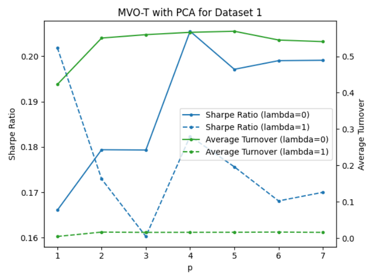


Figure 8: Sharpe ratio and average turnover with respect to lambda (a) Dataset 1, (b) Dataset 2, and (c) Dataset 3 for MVO-T with BSS.

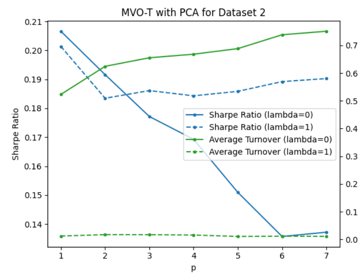
4.1.2 MVO-T with PCA

From Figure 9, it is evident that similarly to MVO-T with BSS, average turnover decreases significantly when the turnover term is added to the MVO objective. However, the new average turnovers for MVO-T with PCA are on the order of 0.01, which is one order of magnitude larger than with BSS.

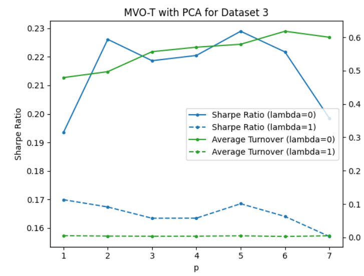
Furthermore, MVO-T with PCA results in unpredictable behaviour with respect to Sharpe ratio as p varies. Across the datasets, the Sharpe ratio using MVO-T with PCA is not consistently above or below the Sharpe ratio using regular MVO as p varies, and the difference between the Sharpe ratios is often $> 4 - 5\%$. Thus, it appears that it will be difficult to predict the impact of on the Sharpe ratio of a dataset and the Sharpe ratio has the potential to drastically decrease if we selected the MVO-T with PCA algorithmic trading strategy.



(a)



(b)



(c)

Figure 9: Sharpe ratio and average turnover with respect to the number of PCs selected as factors for (a) Dataset 1, (b) Dataset 2, and (c) Dataset 3. Note that lambda = 0 indicates that MVO was used whereas lambda = 1 was MVO-T.

Table 14: Sharpe ratio and average turnover of MVO-T with PCA for best p

		Dataset 1	**Dataset 2**	**Dataset 3**
$\lambda = 1$	p	7	1	2
	Sharpe	0.1926	0.2106	0.2137
	Turnover	0.1718	0.1437	0.2250
Baseline	**Sharpe**	0.1958	0.1133	0.1574
	Turnover	0.5376	0.5949	0.6009

4.1.3 MVO-T with Cardinality Constraints

The goal of MVO-T-CC is to create an optimization model that would allow us to take more control over our portfolio using cardinality constraints, while minimizing average turnover.

The mathematical model of MVO-T-CC can be outlined as follows:

4.1.3.1 Mathematical Model

$$
\begin{aligned}
\min_{\mathbf{x},\mathbf{z}} \quad & \mathbf{x^T Q x} + \lambda \sum_{i=1}^{n} z_i \\
\text{subject to} \quad & \boldsymbol{\mu}^T \mathbf{x} \geq \mathbf{R} \\
& \mathbf{1}^T \mathbf{x} = 1 \\
& \mathbf{1}^T \mathbf{y} \leq K \\
& x_i \geq L_i y_i \quad \text{for } i = 1, \ldots, n \\
& x_i \leq U_i y_i \quad \text{for } i = 1, \ldots, n \\
& y_i \in \{0, 1\} \quad \text{for } i = 1, \ldots, n \\
& x_i \geq 0 \quad \text{for } i = 1, \ldots, n \\
& \mathbf{x} - \mathbf{x0} \leq \mathbf{z} \\
& \mathbf{x0} - \mathbf{x} \geq -\mathbf{z} \\
& \mathbf{z} \geq \mathbf{0}
\end{aligned}
$$

, where z_i is the average turnover for each asset.

Different values of parameters L, U, K, and λ, were tested as outlined in the following sections.

4.1.3.2 Parameter: Lower Buy-in Threshold (L)

From Figure 2 in Appendix B.3.1, we can see that the lower the buy-in threshold is better, making the model less restrictive. We can also see that the Sharpe ratio and turnover rate has an inverse relationship. This makes sense because if the lower buy-in threshold is high, we need to put more wealth in an asset, which results in lower number of assets being chosen each time. As a result, when rebalancing the portfolio, there are more drastic changes to which assets we choose each time. Thus, we will keep L = 0.05 fixed, while varying U and K.

4.1.3.3 Parameter: U, K, and λ

From our previous parameter adjustments detailed in Section 4.1.3.2, we found that L = 0.05 performs most optimal. Thus, keeping L constant, we varied U according to Table 15 and K between 10 and 30 assets for dataset 1, 2, and 3. This process was done for $\lambda = 1$ and $\lambda = 100$.

Upper threshold	0.1	0.15	0.2	0.25	0.3	0.5	0.8	1

Table 15: Parameters tested for upper buy-in threshold constraint with L = 0.05 and varying K

The best parameters were based off the porfolio's performance when measured against Sharpe ratio and average turnover. As per client's request, we give higher weighting to Sharpe ratio. Thus, we decided to choose the best U and K that would maximize Sharpe, while giving average to low turnover. A summary of best parameters based off of the graphs summarized in Appendix C.1 for each dataset is summarized in Table 16.

Comparing regular MVO-CC with MVO-CC and min turnover, the parameters vary significantly between the two models and for different λ.

From Table 16, U can vary significantly between datasets. This suggests that we would have to iterate through numerous parameters during the 5-year calibration stage to get the most optimal parameters for different datasets. From Figure 6 detailed in Appendix C.1, the Sharpe ratio and average turnover remains consistent with varying K. However, from analyzing the average turnover

Table 16: Best parameters U and K for each dataset with resulting Sharpe ratio and average turnover

		Dataset A	Dataset B	Dataset C
$\lambda = 1$	**U**	1	0.2	0.1
	K	15	13	12
	Sharpe	0.2119	0.2082	0.2218
	Turnover	0.0094	0.0252	0.05988
$\lambda = 100$	**U**	0.5	0.1	0.8
	K	15	10	12
	Sharpe	0.2515	0.2281	0.2687
	Turnover	0.0284	0.0610	0.0177

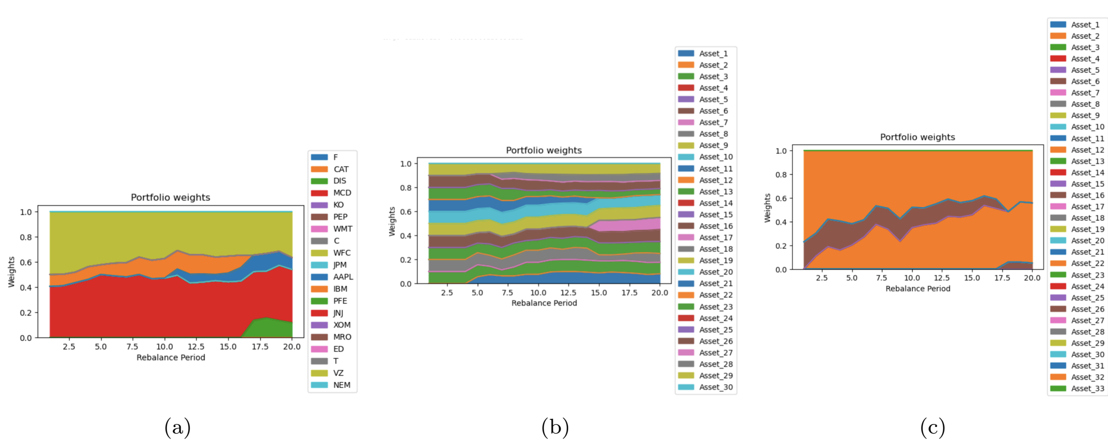
in Figure 4, we found that a more diversified portfolio tends to have lower turnover because changes in individual asset prices have a smaller impact on the overall portfolio allocation. Thus, we thought it would be best to set a higher K to allow for more diversification. We also found that a smaller K (i.e. K $\leq$ 5) can give an infeasible solution with a fixed U = 0.1 or U = 0.15. This suggests that investing in higher number of stocks would be better.

Because the parameters can vary quite a bit, one disadvantage of this model would be the computational runtime, as it can take more than 5 minutes to run through 3 different parameters (U, K, λ) each with 5-10 options (i.e. up to 1000 combinations).

However, we can conclude that MVO-T-CC performs average to above-average compared to MVO-CC with respect to the Sharpe ratio, while significantly decreasing average turnover by more than 50%.

4.1.3.4 Portfolio Weight Distribution

As seen in Figure 10, we can see that the diversification of assets vary between datasets. For datasets 1 and 3, the model chose to put weights in only a few assets (i.e. 3-5 assets), whereas dataset 2 more diversified with weights over 12 assets. Although these portfolio weights did result in the highest Sharpe ratio, the diversification of the porfolio is lacking compared to some of the other models. This further emphasizes the need to iterate through numerous parameters during the 5 year calibration period, as the most optimal parameters can vary significantly between datasets and our model may be overfit if we choose fixed parameters based on training.



(a) (b) (c)

Figure 10: Portfolio weights with MVO-T-CC optimization using best parameters chosen for Dataset 1 (a), Dataset 2 (b), and Dataset 3 (c)

4.2 Risk Parity with Dimensionality Reduction

One thing we observed about the risk parity model was that it performed very consistently across Dataset 1, 2, and 3. While it often did not have the best Sharpe ratio or average turnover, it always performed in the top quartile of the models we evaluated (see Table 1). Furthermore, in Section 3.5, we found that adding dimensionality reduction techniques to MVO often improved Sharpe ratio and average turnover. Thus, we decided to further analyze how different parameter estimation techniques for $\boldsymbol{\mu}$ and $\mathbf{Q}$ would impact the performance of the risk parity model using three different dimensionality reduction techniques:

- BSS
- LASSO
- PCA

By reducing dimensionality, this helps promote sparsity and reduce overfitting, while diversifying risk across all assets.

4.2.1 Risk Parity with BSS (RP-BSS)

In general, $L = 0$ performs the best (see Appendix D) for the RP-BSS strategy. For U and K, although there are parameters that evidently perform better than others (see Figure 8), there is overall little variation in both Sharpe ratio and average turnover as U and K vary. However, Dataset 1, 2, and 3 have different Sharpe ratio trends as U varies. Dataset 1 and 3 have a Sharpe ratio that exhibits concave down behaviour as U increases, but Dataset 2 has a Sharpe ratio trend that exhibits concave up behaviour as U increases. For all datasets, Sharpe ratio tends to decrease as K increases, but at low values of U, Sharpe ratio for all K is relatively identical. In general, $K = 2$ results in the best Sharpe ratio performance across all the datasets.

One important thing to note is that RP-BSS decreases average turnover when compared to raw risk parity and raw MVO with BSS without significantly impacting Sharpe ratio, unlike with the strategies that were combined with MVO-T in Section 4.1. Moreover, using BSS with risk parity also increased Sharpe ratio from raw risk parity for Dataset 1 and 3 (see Table 17).

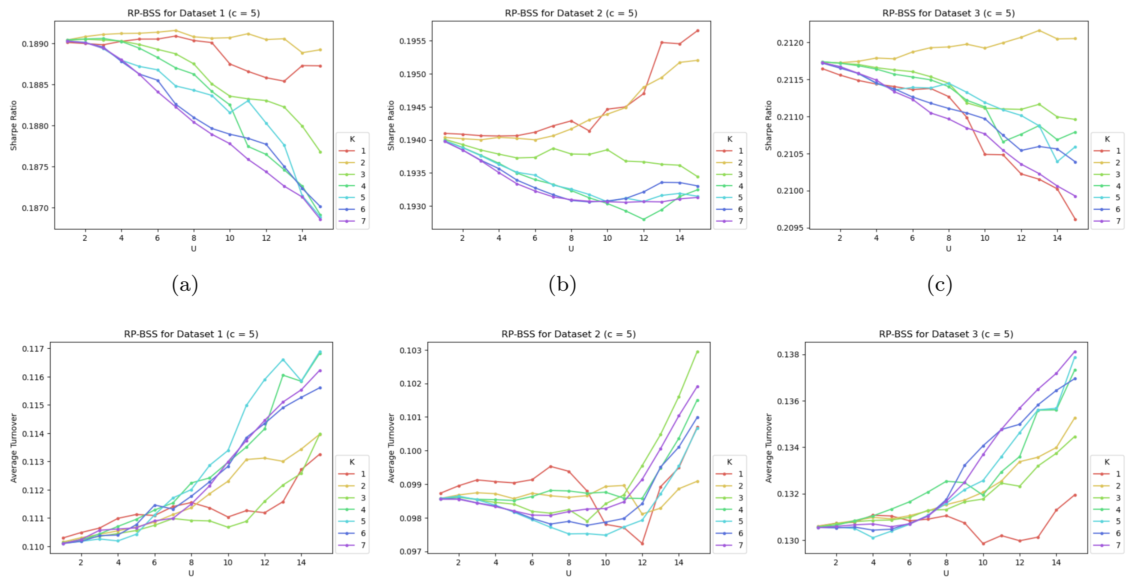
Table 17: Sharpe ratio and average turnover of RP with BSS for best U and K

		Dataset 1	**Dataset 2**	**Dataset 3**
RP-BSS	U	7	15	13
	K	2	1	2
	Sharpe	0.1892	0.1957	0.2122
	Turnover	0.1111	0.1007	0.1336
RP	**Sharpe**	0.1848	0.1996	0.2080
	Turnover	0.1538	0.1284	0.1934
Baseline	**Sharpe**	0.1958	0.1133	0.1574
	Turnover	0.5376	0.5949	0.6009

4.2.2 Risk Parity with LASSO (RP-LASSO)

Risk parity was combined with regularization technique LASSO to determine if Sharpe ratio and average turnover would perform better compared to raw risk parity. Evaluation metrics of these models are summarized in Table 18.

Figure 12 illustrates how the pattern for Sharpe ratio and average turnover greatly depends on the dataset, which are similar to the results in section 3.5.1. However, similar to section 2, we see



(a) (b) (c)

(d) (e) (f)

Figure 11: (a)-(c) Sharpe ratio and (d)-(f) average turnover for risk parity ($c = 5$) with BSS ($L = 0$) for parameter estimation with varying U and K using Dataset 1, 2, and 3.

Table 18: Sharpe ratio and average turnover of RP with LASSO using best parameters

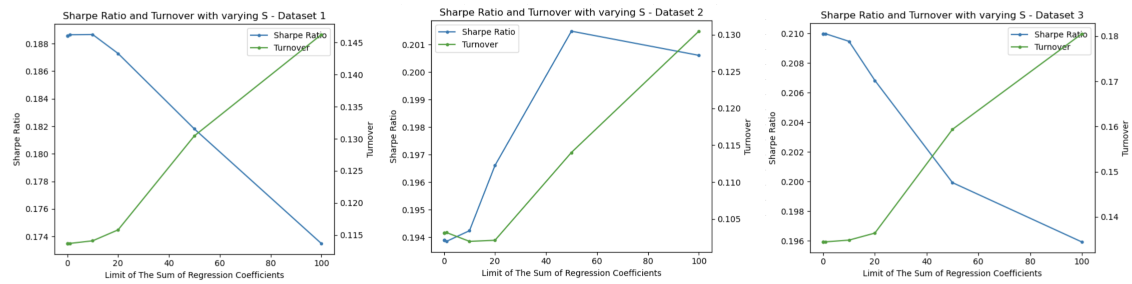
		Dataset 1	**Dataset 2**	**Dataset 3**
RP-LASSO	S	1	1	0.001
	Sharpe	0.1878	0.1940	0.2036
	Turnover	0.1071	0.0971	0.1269
RP	**Sharpe**	0.1848	0.1996	0.2080
	Turnover	0.1538	0.1284	0.1934
Baseline	**Sharpe**	0.1958	0.1133	0.1574
	Turnover	0.5376	0.5949	0.6009

that a smaller S gives higher Sharpe and lower turnover. Compared to similar models, RP-LASSO performs similarly in terms of Sharpe ratio with differences seen in the thousandths. However, the turnover significantly decreases by 3-5% with the RP-LASSO model. One reason for this could be because risk parity compared to MVO promotes diversification from a risk perspective, and thus weights are more evenly distributed over a larger asset class leading to lower average turnovers. Furthermore, adding in regularization techniques prevents overfitting and improves model stability leading to lower average turnovers as well.

4.2.3 Risk Parity with PCA (RP-PCA)

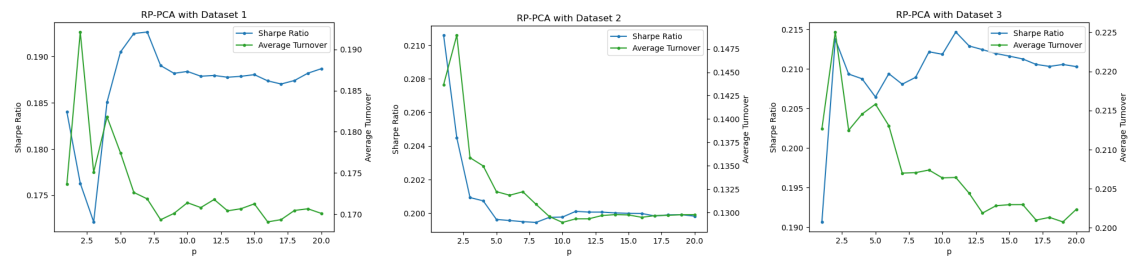
Figure 8 shows that as p varies, the Sharpe ratio varies within a larger range of values than with BSS. These larger overall variations result in the RP-PCA method having larger Sharpe ratios than the RP-BSS model not only when the best p for a dataset is selected, but also for the majority of ps. This can be seen from Figure 8 where while the range of the max and min Sharpe ratios is relatively large, only a small number of ps reach the extreme max and min Sharpe ratios, with the remaining ps achieving a plateau Sharpe ratio greater than the best Sharpe ratios per dataset for the best BSS parameters with RP-BSS.

Table 19 shows that although RP-PCA results in a higher Sharpe ratio than with just raw risk



(a) (b) (c)

Figure 12: Sharpe ratio and average turnover for risk parity ($c = 1$) with Lasso for parameter estimation with varying S using (a) Dataset 1, (b) Dataset 2, and (c) Dataset 3.



(a) (b) (c)

Figure 13: Sharpe ratio and average turnover for risk parity ($c = 1$) with PCA for parameter estimation with varying p (i.e., number of PCs selected as factors) using (a) Dataset 1, (b) Dataset 2, and (c) Dataset 3.

parity ($\approx 1\%$ increase), it also results in a higher average turnover ($\approx 1\%$ increase).

Table 19: Sharpe ratio and average turnover of RP with PCA for best p

		Dataset 1	Dataset 2	Dataset 3
RP-PCA	p	7	1	1
	Sharpe	0.1926	0.2106	0.2146
	Turnover	0.1718	0.1437	0.2064
RP	**Sharpe**	0.1848	0.1996	0.2080
	Turnover	0.1538	0.1284	0.1934
Baseline	**Sharpe**	0.1958	0.1133	0.1574
	Turnover	0.5376	0.5949	0.6009

4.2.4 General Takeaways from Risk Parity with Regularization

4.2.4.1 Comparing Evaluation Metrics

General trends we found when combining regularization techniques with risk parity is summarized below:

1. Although regularization with regular MVO can sometimes yield a slightly higher Sharpe ratio ($< 1.5\%$), we found that average turnover significantly decreased when using risk parity with a regularization technique. Thus, we can decrease average turnover without compromising Sharpe when compared to similar models (i.e., raw risk parity, MVO with regularization).

2. We can identify a general trend for how varying parameter values will change performance metrics for different datasets. This suggests that if we dynamically select specific parameters during the initial 5-year calibration period, the model will not suddenly perform really badly for the rest of the data since Sharpe ratio and average turnover do not change as suddenly as with other models when the optimal parameters change. This again goes back to the idea that regularization techniques promote sparsity and risk parity promotes diversification across risk, resulting in more stable and diverse portfolios.

Other advantages of using RP instead of MVO include:

1. There is more opportunity for greater returns with RP, as it does not focus on minimizing risk, but instead focus on diversifying risk. Since, RP's main focus is not minimizing risk, it may choose some riskier assets with potential for greater return, while still averaging out risks.
2. Because MVO focuses on minimizing risk, the most optimal solution may be concentrated in a few assets (i.e. assets with least amount of risk). Compared to RP which ensures all risk contributions of each asset is equal, the portfolio weights are diversified across all assets. This can be demonstrated by the portfolio weight graphs outlined in Figure 14.
3. RP exhibits robust performance across diverse market conditions, as evidenced by testing across various timelines through the different datasets.

4.2.4.2 Portfolio Weights

From Section 3.5.4.1, we know that reducing the dimensionality of the problem and promoting sparsity will result in a more diversified portfolio. Comparing risk parity and MVO with a dimensionality reduction technique, we found that the portfolio weights in the RP model is more stable over time compared to the MVO model as shown in Figure 14. This also helps explain why average turnover was lower for RP model compared to the MVO model.

5 Final Model

The details of our final algorithmic trading model are outlined below. The final model is a combination of the risk parity model with PCA for parameter estimation.

5.1 Mathematical Model

$$\begin{aligned} \min_{\mathbf{y}} \quad & \frac{1}{2}\mathbf{y}^{\mathbf{T}}\mathbf{Q}\mathbf{y} - c\sum_{i=1}^{n} \ln(y_i) \\ \text{subject to} \quad & \mathbf{y} \geq 0 \end{aligned}$$

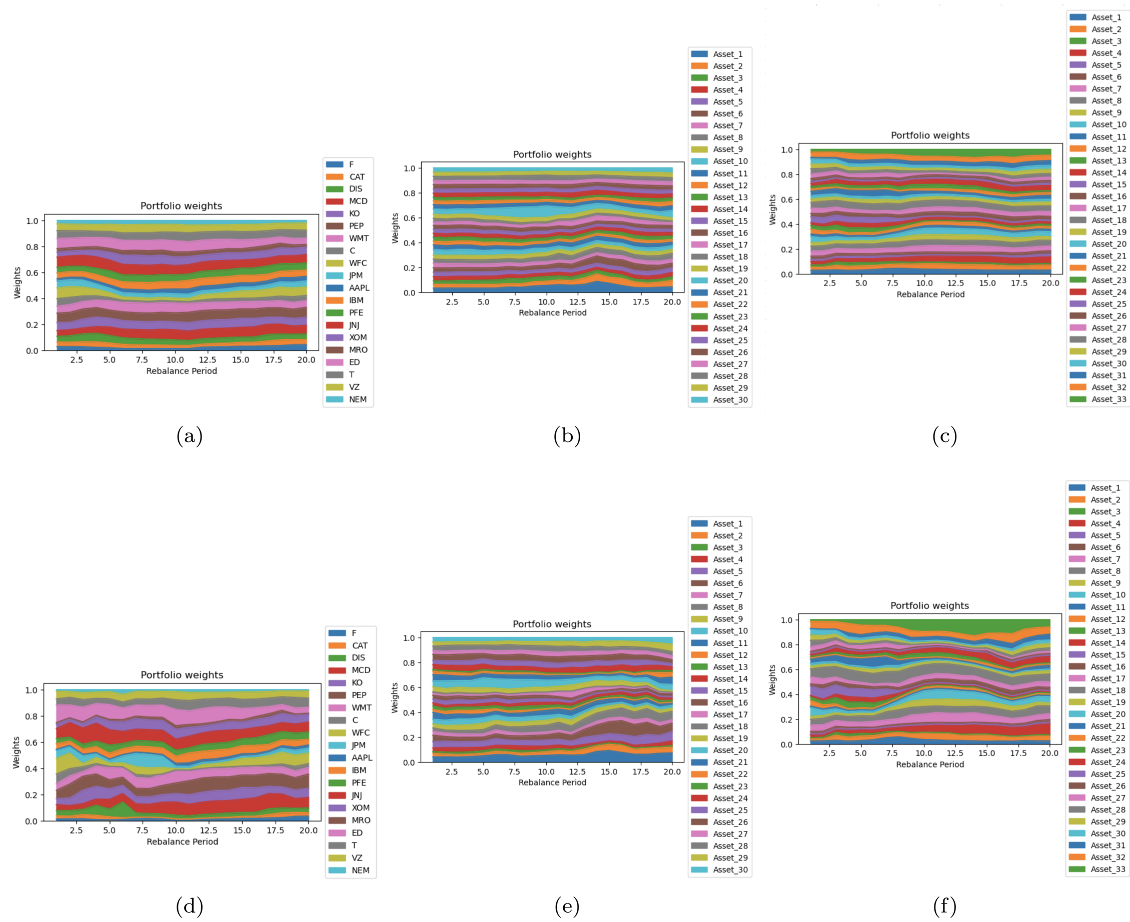
where:

$\mathbf{x} \in {}^n\mathbb{R}^1$ is the portfolio weights and can be calculated by normalizing $\mathbf{y}$: $x_i^* = \frac{y_i^*}{\sum_{i=1}^n y_i}$

$\mathbf{Q} \in {}^n\mathbb{R}^n$ is the covariance matrix estimated using p PCs from $\mathbf{P} = \overline{\mathbf{R}}\mathbf{\Gamma}$ (see Section 3.5.3).

$c = 5$ is a scaling constant for risk parity

$p \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15\}$ is determined during the 5-year calibration stage as described in Figure 1.



(a) (b) (c)

(d) (e) (f)

Figure 14: Portfolio weights of Risk Parity with regularization model using Dataset 1(a), 2(b), and 3(c). Portfolio weights of MVO with regularization model for Dataset 1(d), 2(e), and 3(f).

5.2 Results from Testing

Table 20 compares Sharpe ratio and average turnover results for all the RP-dimensionality reduction models using the parameter selection architecture described in 1 using our training, validation, and testing datasets. This confirmed that the parameters the strategy selected during the initial 5-year calibration period are good because the selected parameters, Sharpe ratio, and average turnover results are consistent with what we expected from our analysis of these models in Section 4.2. Thus, our implementation of a dynamic parameter selection strategy was successful; not having to set fixed parameters for our final algorithmic trading strategy is very important because different parameters perform well for different datasets.

For the testing datasets (i.e., Dataset A, B, and C), RP-PCA has consistently higher Sharpe ratios than the other strategies, including the baseline model (see Table 20). Although average turnover is also higher for the RP-PCA strategy compared to RP-BSS and RP-LASSO, it is still lower than the baseline model and MVO-PCA. Since the Sharpe ratio is the more important evaluation metric for this project, these results confirmed that the RP-PCA strategy should be our final model.

Our testing results give us the confidence that our model will be able to consistently outperform the market when the market performs poorly or achieve very similar results to the market when it performs poorly.

Table 20: Sharpe ratio and average turnover of RP-dimensionality reduction models with parameters determined based on the strategy described in Figure 1 (fixed parameters $c = 5$ and $L = 0$)

		***D1**	**D2**	**D3**	**DA**	**DB**	**DC**
RP-BSS	U	5	15	15	13	14	5
	K	2	7	2	2	4	6
	Sharpe	0.1877	0.1929	0.2049	0.4709	0.1518	0.1634
	Turnover	0.1040	0.0923	0.1249	0.1223	0.0925	0.1079
RP-LASSO	S	1	1	0.001	0.001	10	0.001
	Sharpe	0.1878	0.1940	0.2036	0.4691	0.1518	0.1637
	Turnover	0.1071	0.0971	0.1269	0.1283	0.0955	0.1096
****RP-PCA**	p	4	1	8	4	1	6
	Sharpe	0.1881	0.2044	0.2071	0.5052	0.1527	0.1690
	Turnover	0.1534	0.1166	0.1705	0.1541	0.1027	0.1507
Baseline	**Sharpe**	0.1958	0.1133	0.1574	0.5365	0.1335	0.1323
	Turnover	0.5376	0.5949	0.6009	0.6089	0.5205	0.5421

*Note that Dataset is abbreviated to **D** in Table 20, i.e., **D1** is Dataset 1.

**The model highlighted in red is the final algorithmic trading strategy selected.

6 Discussion

6.1 Strengths and Weaknesses of the Model

6.1.1 Strengths

Key strengths of our final model can be summarized as follows:

1. **Consistent Sharpe ratio across different datasets**

   Our final model performs very consistently in terms of Sharpe ratio across all our training, validation, and testing datasets, either significantly outperforming the baseline model when the baseline performs relatively poorly compared to our other initial models in Section 3 or achieving very similar results to the baseline model when the baseline performs relatively well. Thus, the final RP-PCA model would be attractive to risk-averse investors.

2. **Risk diversification**

   By definition, risk parity maximizes risk diversification (i.e., it equalizes the risk contribution per asset). Furthermore, since estimated returns are unreliable and the risk parity strategy ignores estimated returns, this is advantageous during portfolio construction because we avoid potential large estimation errors.

   On another note, using PCA for parameter estimation can provide more stable covariance estimates because PCA extracts factors representing common sources of systematic risk in the market. This can also help construct portfolios that are well-diversified across different sources of systematic risk and improve the robustness of the portfolio. By diversifying across different systematic risk factors, the resultant portfolio becomes less reliant on the performance of individual assets.

3. **Reduced problem dimensionality**

   Employing PCA in parameter estimation offers the advantage of reducing the dimensionality of the problem, making it more manageable, particularly when dealing with a large number of assets. By identifying a smaller set of principal components that capture the majority of the variation in asset returns, PCA simplifies the optimization process and reduces computational runtime. As a result, PCA improves the precision and effectiveness of risk parity by providing reliable parameter estimates, ultimately contributing to more robust portfolio management and risk mitigation.

6.1.2 Weaknesses

Key weaknesses of our final model can be summarized as follows:

1. **Does not explicitly target low average turnover**

   Our model does not explicitly optimize for low average turnover. In Section 4.1, we evaluated the impact of adding a term to penalize average turnover in the objective of MVO. However, while this did significantly decrease average turnover, it often also resulted in a non-insignificant decrease in Sharpe ratio and the impact of the penalization factor was unpredictable. In general, our strategies that included risk parity as the portfolio optimization objective had relatively low average turnovers, but we did not use any direct methods to reduce or constrain average turnover in our final algorithmic trading strategy.

2. **Difficult to predict convergence characteristics of *cvxpy* solvers**

   During our evaluation of different portfolio optimization models, we noticed that different *cvxpy* solvers could have different results. For the risk parity strategy, the *cvxpy.ECOS* solver would often be unable to solve the optimization problem with small c. This could be a result of how ECOS scales the problem using the conic optimization framework. On the other hand, while the *cvxpy.SCS* solver results in feasible solutions for risk parity for all c, it will often use the maximum number of iterations to solve risk parity because risk parity does not converge to values within the default required ϵ for small c.

   Thus, our algorithmic trading strategy does not guarantee a feasible solution for all possible sets of parameters for all possible datasets due to limitations of numerical solvers.

3. **Weakened interpretability of factors and portfolio**

   Implementing PCA introduces uncertainty regarding which factors it will select each time, making it challenging to interpret their impact on portfolio performance. Since PCA identifies principal components based solely on statistical properties of asset returns, without regard to economic or financial theory, understanding the economic rationale behind these factors can be difficult. Consequently, the interpretation of PCA-derived factors and their implications for portfolio outcomes may be difficult, hindering the ability to assess how the model aligns with economic principles.

6.2 Next Steps

Our next steps to further improve our model can be summarized as follows:

1. **Investigate numerical solvers**

   To improve the feasibility of our algorithmic trading strategy, we should continue to investigate the characteristics of numerical solvers and how to best tune their parameters (e.g., the maximum number of iterations, ϵ for convergence, etc.) for the risk parity optimization problem. Furthermore, it could be worthwhile to use a simpler Newton's method numerical solver where we can more explicitly control the convergence characteristics of the problem.

2. **Investigate advances in risk parity portfolio optimization to increase Sharpe ratio and decrease average turnover**

Risk parity solves a different optimization problem than MVO, and thus conventional techniques in portfolio optimization (e.g., robustness - see Appendix B.2) might not work as well for risk parity. Investigating new risk parity frameworks could help us further improve our algorithmic trading strategy.

For example, in Giorgio Costa Del Pozo's PhD thesis, he proposes a robust framework for risk parity portfolio optimization which seeks to address uncertainty in estimating asset risk contributions, a distributionally robust risk parity framework, and a regime-switching factor model for risk parity that introduces an additional dimension of risk into the estimated parameters [5]. All these frameworks tackle parameter estimation uncertainties from a risk parity perspective rather than from a variance minimization perspective, which is what the regularization techniques we explored aimed to target. Thus, using risk parity-specific techniques could help us improve the Sharpe ratio and average turnover of our algorithmic trading strategy.

3. **Improve computational runtime**

Our algorithmic trading strategy relies on iterating through many combinations of model parameters during the initial 5-year calibration period to determine the optimal parameters that should be used for the remainder of the investment horizon (see Figure 1). This iterative process is currently completed using nested for-loops for each parameter, which does not have an optimal computational runtime. To improve computational runtime, our next steps are to vectorize this process using functions such as *numpy.meshgrid* and *numpy.stack*.

7 Conclusion

Through our initial evaluation of various more basic models (e.g., MVO-OLS, factor models, index tracking, etc.) using Sharpe ratio and average turnover as performance metrics, we were able to identify combinations of these models that improved overall model performance across a large range of datasets. By tuning parameters in our initial models, we found that leveraging risk parity with PCA for parameter estimation, while keeping the number of principal components selected as factors dynamic, resulted in a good final algorithmic trading strategy when measured against the two aforementioned metrics. Training, validating, and testing on multiple datasets, including newly generated ones from the S&P 500, revealed competitive performance compared to the market capitalization-weighted portfolio, demonstrating promising potential for real-world application of this model.

Appendices

A Initial Results

Table 1: Dataset statistics for initial optimization models (more details in [6])

	Dataset 1		Dataset 2		Dataset 3	
	Sharpe	**Turnover**	**Sharpe**	**Turnover**	**Sharpe**	**Turnover**
Market Cap	**0.19584**	**0.53761**	**0.1133**	**0.5985**	**0.1574**	**0.6009**
Equal-Weight	0.1777	0.11697	0.1953	0.0882	0.2072	0.1379
MVO-Historical	0.17191	0.48752	0.1410	0.7357	0.1970	0.6700
MVO-OLS	0.18707	0.48650	0.1633	0.6469	0.2295	0.5505
CAPM	0.16504	0.44552	0.1935	0.5074	0.2210	0.4827
Fama-French-3	0.17096	0.46028	0.1727	0.5793	0.2467	0.5289
Fama-French-5	0.17928	0.49045	0.1685	0.6203	0.2319	0.5417
MVO-Lasso	0.1861	0.1851	0.1917	0.1465	0.2176	0.2270
MVO-BSS	0.2011	0.1926	0.1909	0.1791	0.2336	0.2112
MVO-PCA	0.2055	0.5660	0.2066	0.5230	0.2290	0.6181
MVO-Cardinality	0.1823	0.5400	0.1810	0.6700	0.2355	0.5500
Risk Parity[3]	0.1847	0.1538	0.1996	0.1284	0.2080	0.1935
Mean Variance Tracking[4]	0.2008	0.6254	0.1157	0.5981	0.1576	0.6000
CVaR	0.1759	0.1290	0.1917	0.1067	0.2011	0.1510
Robust CVar	0.1487	0.0786	0.1660	0.0816	0.2011	0.1446
Monte Carlo	0.1700	0.5437	0.2057	0.3638	0.1894	0.5474
Monte Carlo with CVaR	0.2037	0.4077	0.1991	0.3650	0.1979	0.5102

[3] Using fixed parameter of $c = 5$

[4] Observation: tracks market cap very accurately

B Project 1 Models and Test Data

B.1 MVO with Historical Data

Using the model given in the template code to compute the weights, the resulting Sharpe ratio and average turnovers vary from the three datasets. However, they tend to have higher average turnovers since MVO's main goal is to minimize variance which can be achieved by diversification. Therefore placing wealth into more assets and holding onto the wealth over time can lead to higher average turnovers. Furthermore, MVO with historical data can lead to overfitting since it places weights based on historical patterns of assets. If an asset price continues to increase, there is a possibility it may crash in the future and it would be risky to place a high amount of weight in the asset. Another flaw in the MVO model is that it assumes assets are normally distributed but this is not true in practice. Hence, since the model did not perform remarkably compared to other models and is flawed in its assumptions of the market we decided to not proceed with this model.

B.2 Robust MVO with Ellipsoidal Uncertainty Set

MVO was further improved by using an ellipsoidal uncertainty set to robustify standard MVO. The risk-return trade off model was selected and the ellipsoidal uncertainty set was included in the objective of the robust MVO so that we could readily add parameters that represent how much we care about return and the uncertainty set in the objective. Furthermore, by adding return and the ellipsoidal uncertainty set into the objective, we prevent unfeasible solutions to this version of MVO.

B.2.1 Mathematical Model

$$\begin{aligned} \min_{x} \quad & \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} - \lambda\boldsymbol{\mu}^T\mathbf{x} + \epsilon\|\boldsymbol{\Theta}^{1/2}\mathbf{x}\|_2 \\ \text{subject to} \quad & \mathbf{x} \geq 0 \\ & \mathbf{1}^T\mathbf{x} = 1 \end{aligned}$$

where:

$\mathbf{Q} \in {}^n\mathbb{R}^n$ is the estimated covariance matrix of the n assets

$\boldsymbol{\mu} \in {}^n\mathbb{R}^1$ is the estimated expected return of the n assets

$\mathbf{x} \in {}^n\mathbb{R}^1$ is the portfolio weights for each of the n assets

$\boldsymbol{\Theta} \in {}^n\mathbb{R}^n$ is the matrix of squared standard errors arising from the estimation of the T asset expected returns, where $\Theta = \frac{1}{T}diag(\mathbf{Q})$, and $(\Theta^{1/2})_{ii} = \frac{\sigma_i}{\sqrt{T}}$ and $(\Theta^{1/2})_{ij} = 0$ for $i \neq j$

λ is return-favouring parameter

ϵ is a scaling parameter that determines the size of the uncertainty set, where $\epsilon = \sqrt{\chi^2_n(\alpha)}$ and α is the confidence level

B.2.2 Parameters: λ and α

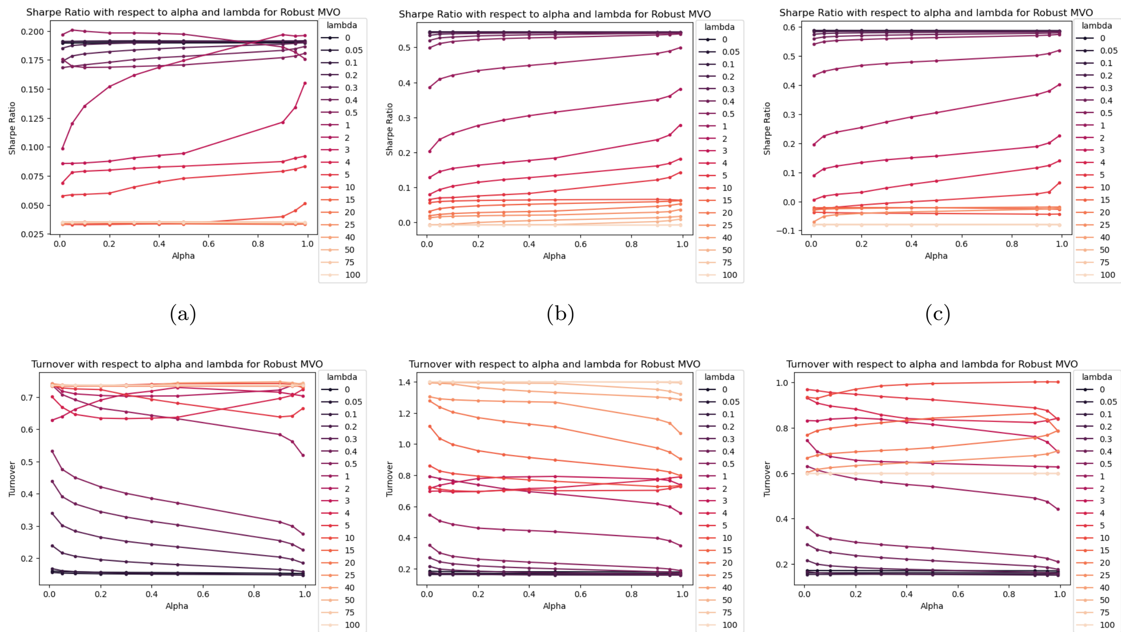
Table 2 summarizes the parameter values that were tested with the 3 different datasets. All combinations of the parameters were tested and Figure 1 shows that datasets A, B, and C had similar trends with respect to Sharpe ratio and turnover. In general, as λ increases, Sharpe ratio decreases and turnover increases. For high λ, α does not significantly impact the Sharpe ratio and turnover. For low λ, as α increases, Sharpe ratio increases and turnover decreases.

Furthermore, high values of λ resulted in over-concentrated portfolios.

From these datasets, it appears that α close to 1 and small λ produce the best results in terms of Sharpe ratio and turnover for robust MVO with an ellipsoidal uncertainty set.

λ	0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 1, 2, 3, 4, 5, 10, 15, 20, 25, 40, 50, 75, 100
α	0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.9, 0.95, 0.99

Table 2: Parameters tested for robust MVO with ellipsoidal uncertainty set



(a) (b) (c)

(d) (e) (f)

Figure 1: (a)-(c) Sharpe ratio and (d)-(f) turnover with rest to α and λ for robust MVO with an ellipsoidal uncertainty set

B.3 MVO with Cardinality Constraints

B.3.1 Varying Lower Buy-In Threshold (L)

Using the three datasets, different values of L and K were tested to find the best lower buy-in threshold and total number of assets. The upper buy-in threshold was set at 1 for all trials. First, different values for the lower buy-in threshold (see Table 3) were tested with U and K constants at 1 and 10, respectively.

Label	A	B	C	D	E	F	G	H
Lower threshold	0.01	0.02	0.03	0.04	0.05	0.1	0.2	0.3

Table 3: Parameters tested for lower buy-in threshold constraint with U = 1, K = 10

We found that lower buy-in thresholds of L = 0.03 and L = 0.05, perform the best across the 3 datasets.

B.3.2 Varying U and K

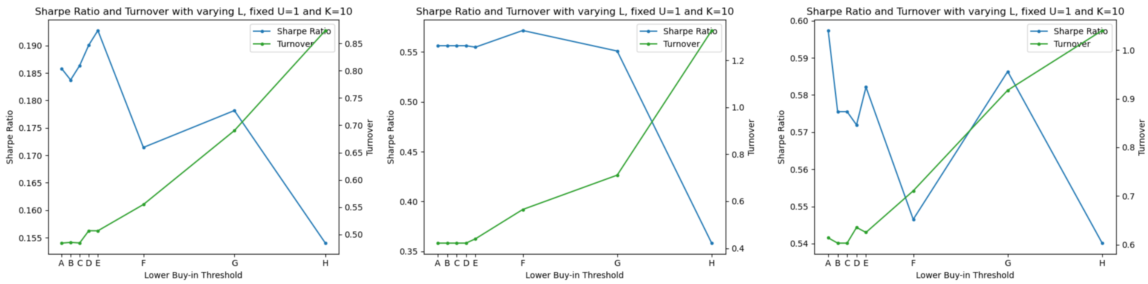
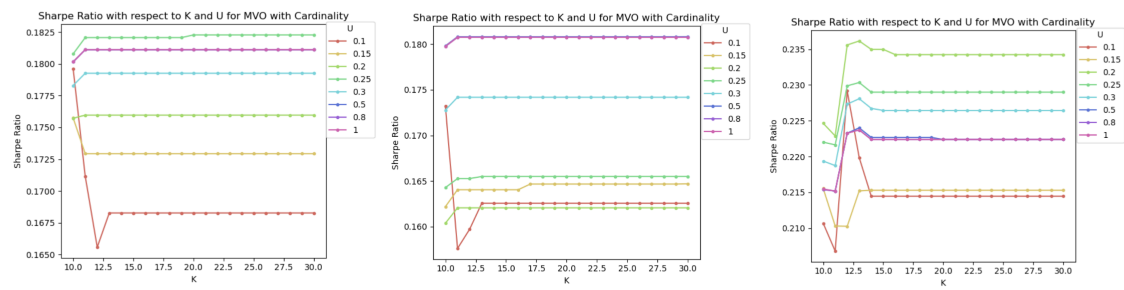
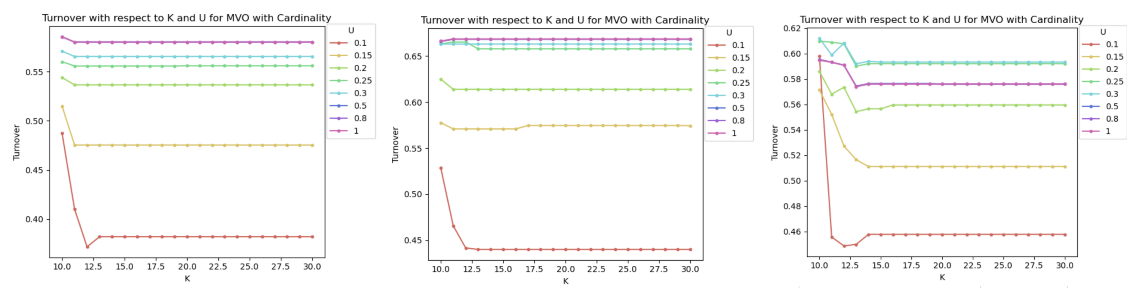


Figure 2: Sharpe ratio and average turnover for varying L, fixed U=1 and K=10 for Dataset A (left), B (center), C (right)



(a) (b) (c)



(d) (e) (f)

Figure 3: (a)-(c) Sharpe ratio and (d)-(f) turnover for datasets A, B, and C respectively, with varying U and K for regular MVO with cardinality

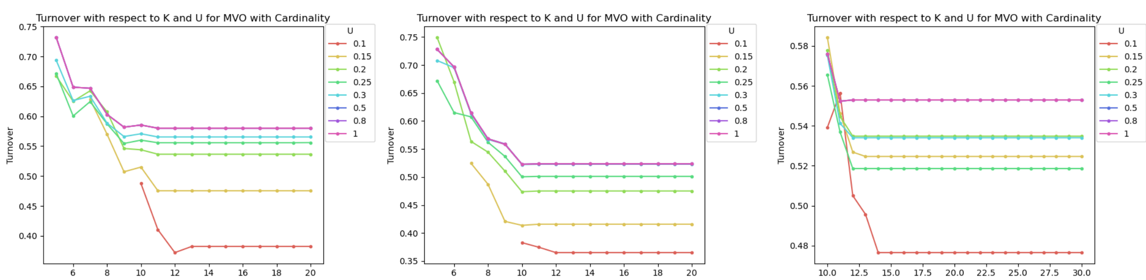


Figure 4: average turnover for MVO Cardinality for varying U and K for Dataset A (left), B (center), C (right)

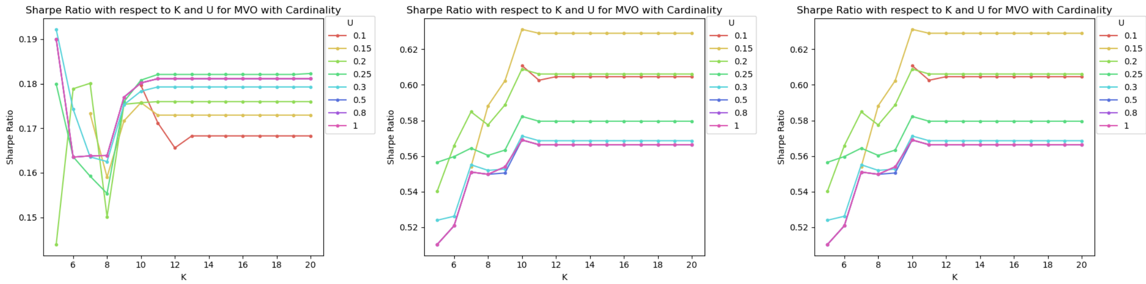
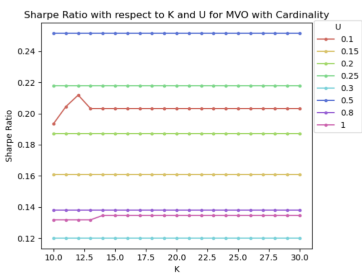


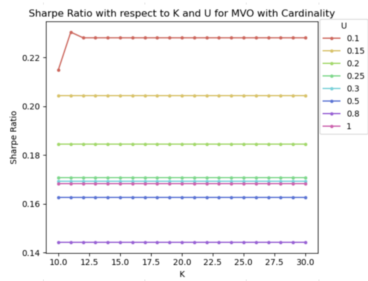
Figure 5: Sharpe rate for MVO Cardinality for varying U and K for dataset A (left), B (center), C (right)

C MVO with Cardinality Constraints and Min Turnover

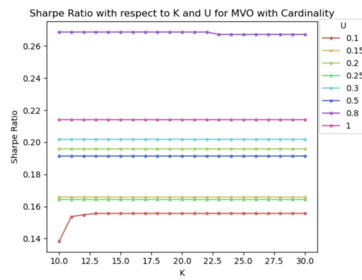
C.1 Varying U and K with different λ



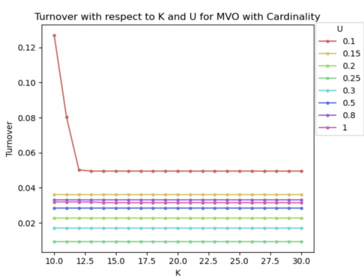
(a)



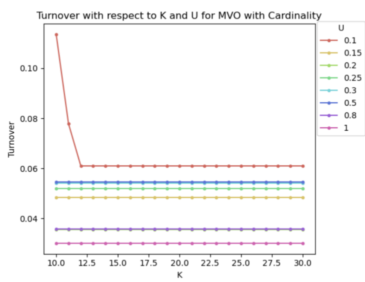
(b)



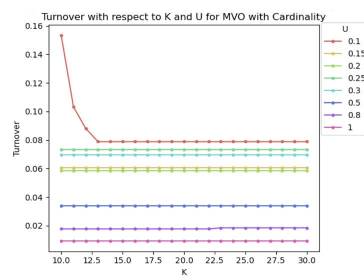
(c)



(d)

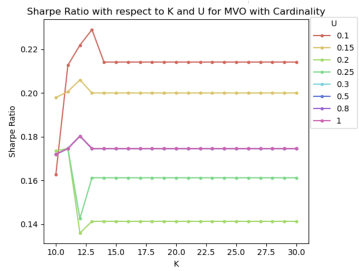


(e)

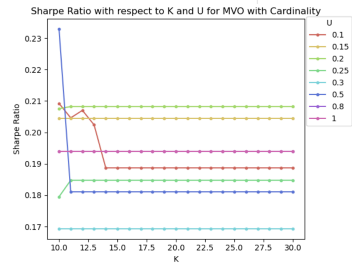


(f)

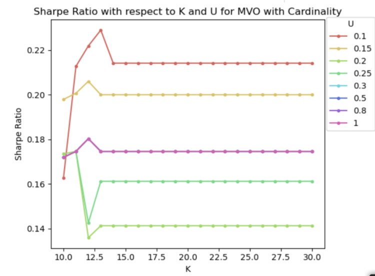
Figure 6: (a)-(c) Sharpe ratio and (d)-(f) turnover for datasets A, B, and C respectively, with varying U and K and fixed $\lambda = 100$ for MVO with cardinality and min turnover term



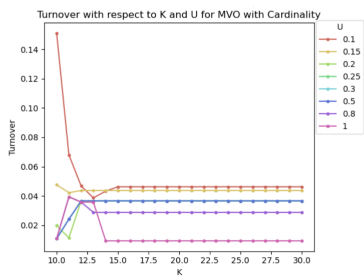
(a)



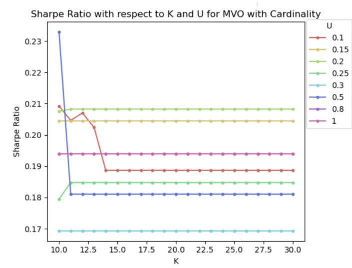
(b)



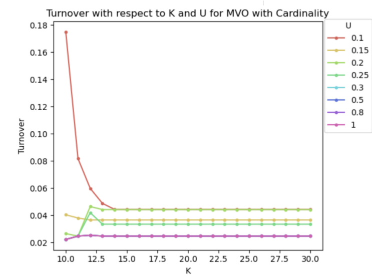
(c)



(d)



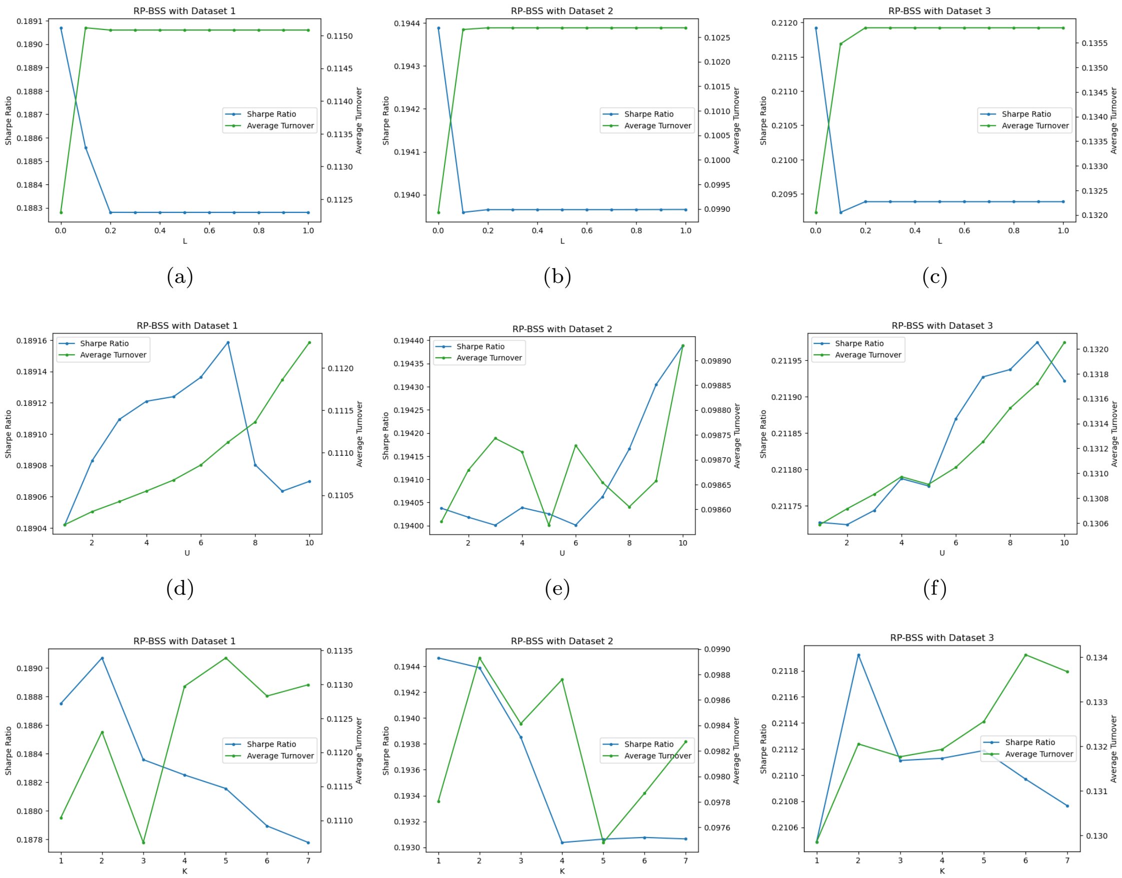
(e)



(f)

Figure 7: (a)-(c) Sharpe ratio and (d)-(f) turnover for datasets A, B, and C respectively, with varying U and K and fixed $\lambda = 1$ for MVO with cardinality and min turnover term

D Varying BSS-Related Parameters for RP-BSS



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 8: Sharpe ratio and average turnover for risk parity ($c = 1$) with BSS for parameter estimation with varying (a)-(c) L, (d)-(f) U, and (g)-(i) K using Dataset 1, 2, and 3. When fixed, $L = 0$, $U = 5$, and $K = 3$.

References

[1] P. Mooney. *Stock Market Data (NASDAQ, NYSE, SP500)*. https://www.kaggle.com/datasets/paultimothymooney/stock-market-data/data. [Online; accessed 17-Mar-2024]. 2023.

[2] R. Fink. *Why active managers have trouble keeping up with the pack*. July 2014. URL: https://www.chicagobooth.edu/review/why-active-managers-have-trouble-keeping-up-with-the-pack.

[3] A. W. Derek Horstmeyer Ying Liu. *Fama and French: The Five-Factor Model Revisted*. https://blogs.cfainstitute.org/investor/2022/01/10/fama-and-french-the-five-factor-model-revisited/. [Online; accessed 16-Mar-2024]. 2022.

[4] URL: https://github.com/skfolio/skfolio.

[5] G. Costa Del Pozo. "Advances in risk parity portfolio optimization". PhD thesis. 2021. URL: https://tspace.library.utoronto.ca/bitstream/1807/106376/4/Costa_Del_Pozo_Giorgio_202106_PhD_thesis.pdf.

[6] en-US. URL: https://docs.google.com/spreadsheets/d/1J95RtgpsTEIOGijhh5Rvp1MiK3ETFkkHIZg3s8E7NV8/edit?usp=sharing&usp=embed_facebook.