Portfolio Optimization Strategies Using Covariance Matrix Shrinkage

Issy Anand, Robbie Cook, George Kepertis, Bhavin Shah, Gauri Sharma

Abstract

When doing portfolio optimization, the most common way to estimate the covariance matrix of a group of assets is to use the sample covariance matrix. Although very simple to calculate, it introduces several problems. Examples include overweighting assets which contain high estimation error and outputting unstable portfolio weights, leading to frequent rebalancing and high transaction costs. Ledoit and Wolf (2003) suggest an alternative way to estimate a covariance matrix using a technique called *shrinkage*. They argue that it should mitigate the problems that arise when using the sample covariance matrix. Therefore in this project, we aim to construct portfolio optimization trading strategies using this alternative, and we will compare the results to those same strategies used with the sample covariance matrix.

Introduction

Mean-variance portfolio theory is one of the most well-known methods for optimizing the returns of a group of assets at a given level of risk. Typically, we use a covariance matrix to quantify that risk between the assets. The most obvious choice of a covariance matrix is to use the sample covariance matrix, which comes from the historical returns of the assets. The benefit of doing this is that it is simple, intuitive, and very easy to implement. Unfortunately there are also costs. The biggest one is that using historical stock returns can lead to high estimation error, especially if our group of assets is very large compared to the amount of data that we have. When we input this sample covariance matrix into an optimizer, it may assign the largest weights to those assets with the most error, which is clearly undesirable. Another cost is that a large sample covariance matrix may lead to unstable weights given to us by the optimizer. If the weights are constantly moving around, we will need to trade very frequently, which will lead to higher transaction costs that eat into our profits.

Ideally we would like a covariance matrix with as little estimation error as possible, and we would like our portfolio weights to be as stable as possible. How do we find a covariance matrix that will accomplish these for us if the sample covariance matrix will not? Ledoit and Wolf (2003) suggest that we use a technique called *shrinkage* and apply it to the sample covariance matrix (along with a *shrinkage target* which we will define later). By taking a linear combination of the sample covariance matrix and the shrinkage target, we will “shrink” the most extreme error terms in the sample covariance matrix, both on the positive and negative ends, towards some center.

In the following paper, we will outline exactly how to perform the shrinkage technique. We will justify why it works, and we will test it on some portfolio trading strategies.

Methodology

Covariance matrix shrinkage:
For this method, we closely followed Ledoit & Wolf, 2003 who claim that "nobody should be using the sample covariance matrix for the purpose of portfolio optimization". The paper suggests that transforming ("shrinking") the covariance matrix using the described method reduces estimation error thus substantially increasing the information ratio of the portfolio.

The main idea behind this shrinkage can be summed up in the following equation:

$$\hat{\Sigma}_{Shrink} = \hat{\delta}^* F + (1 - \hat{\delta}^*) S$$

where F is the shrinkage target matrix, S is the sample covariance matrix calculated based on T observations, and $\hat{\delta}^*$ is the optimal shrinkage intensity. Both S and F have dimension NxN
Calculating the shrinkage target F:
The main idea behind calculating the shrinkage target F, is the assumption of constant and identical pairwise correlation between all assets. This constant $\bar{r}$ is simply assumed to be mean of all the sample correlations. Then, the matrix F can be constructed by

$$F_{ii} = S_{ii} \text{ and } F_{ij} = \bar{r}\sqrt{S_{ii}S_{jj}}$$

Estimating the optimal shrinkage constant:
While the shrinkage constant δ can be chosen by the investor, Ledoit & Wolf, 2003 provide a method to find the optimal shrinkage constant. The estimation is quite involved, and the full details can be found in the Appendix but here we provide an overview of the method:

- A loss function $L(\delta)$ is defined as the Frobenius norm of the matrix $(\delta F + (1 - \delta)S) - \Sigma$, where Σ is the true covariance matrix
- The goal is to find $\delta^* = argmin_{\delta}(E(L(\delta)))$
- Then assuming that δ^* asymptotically behaves like a constant $\varkappa$ for a fixed N as T $\rightarrow \infty$, $\varkappa$ can be expressed as

$$\varkappa = \frac{\pi - \varrho}{\gamma}$$

- Since none of the 3 components of $\varkappa$ are known, consistent estimators $\hat{\pi}$, $\hat{\varrho}$, and $\hat{\gamma}$ are provided (see Appendix)
- Once those are computed, $\hat{\varkappa}$ is obtained, and the optimal shrinkage intensity is given by

$$\hat{\delta}^* = max\{0, min\{\frac{\hat{\varkappa}}{T}, 1\}\}$$

Principal Component analysis:
With this method, we look at the principal components that explain some predetermined share of the total variance and discard the rest as "noise". The method for doing this is as follows:

- The sample covariance matrix S is diagonalized as $S = UDU^{-1}$, where matrix D is a diagonal matrix and its entries correspond to the eigenvalues of S in descending order.
- Now the eigenvalues can be thought of as the amount of variance explained by each principal component.
- Let $\beta = \sum_{i=1}^{N} \lambda_i$ be the sum of the eigenvalues. Then $\psi_i = \frac{\lambda_i}{\beta}$ is the percentage of variance explained by the i[th] principal component.
- The goal is to be able to explain at least x% of the variance. To do this, the first n principal components are taken, where $n = inf\{k \mid \sum_{i=1}^{k} \psi_i \geq \frac{x}{100} \}$.
- In practice, this means that we define a new diagonal matrix C that has entries $C_{ii} = \lambda_i \ if \ i \leq n,$ and $C_{ii} = 0 \ otherwise$.
- We notice that C is not actually invertible, so a pseudo-inverse B can be defined that has entries $B_{ii} = \frac{1}{C_{ii}} \ if \ C_{ii} \neq 0$ and $B_{ii} = 0 \ if \ C_{ii} = 0$.
- The two new matrices that should be used for the optimization are $P = UCU^{-1}$ and $Q = UBU^{-1}$ where Q is pseudo-inverse of P.

Trading Strategy:

We used three different methods to compute the covariance matrix.

- Sample Covariance Matrix
- Shrunk Covariance Matrix found using methods described above
- PCA Covariance Matrix found by doing eigenvalue decomposition on the sample covariance matrix.

Next, we did two types of optimizations.

- Mean Variance: where the expected return is maximized for a given level of risk.
- Minimum Variance: where the variance of the portfolio is minimized.

Hence, in total we had six different weight matrices.

For monthly data, we used a training period of 36 months and a prediction period of 12 months from January 2019 to December 2019. For each month, weights were calculated using 36 months prior data and the portfolio was rebalanced every month. Weights for each month were recorded in seven separate weight matrices corresponding to the different optimization methods.

Using stock data from January 2019 to December 2019 and weights matrices, we calculated monthly portfolio return for the year 2019 for each of the seven optimization methods.

We used the monthly S&P 500 return for 2019 as a benchmark for calculating information ratio.

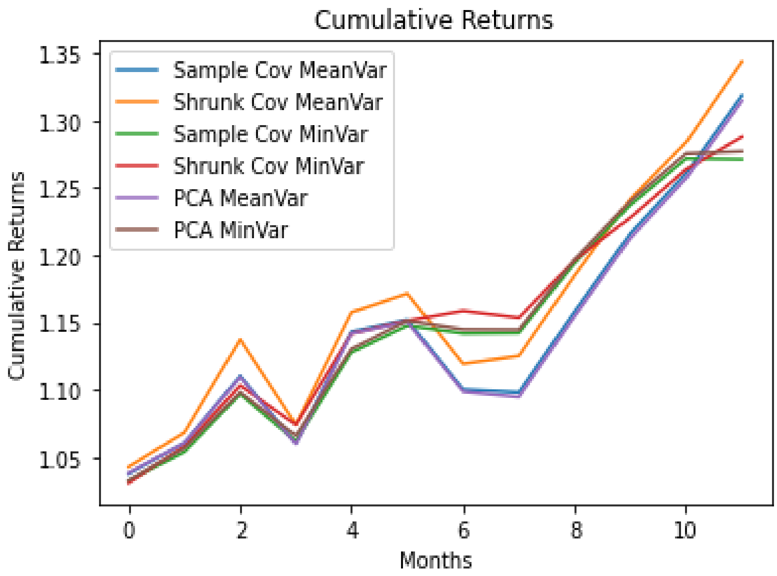
$$Information\ Ratio\ =\ \frac{n\,(R_p - R_b)}{\sqrt{n}\sigma}$$

Here R_p is the portfolio return. R_b is the return of the benchmark i.e. S&P 500. $R_p - R_b$ is the excess return. σ is the standard deviation of the excess return and n is the annualizing factor.

This trading strategy was repeated for weekly data with training data of 53 weeks and prediction period of 52 weeks.

Results

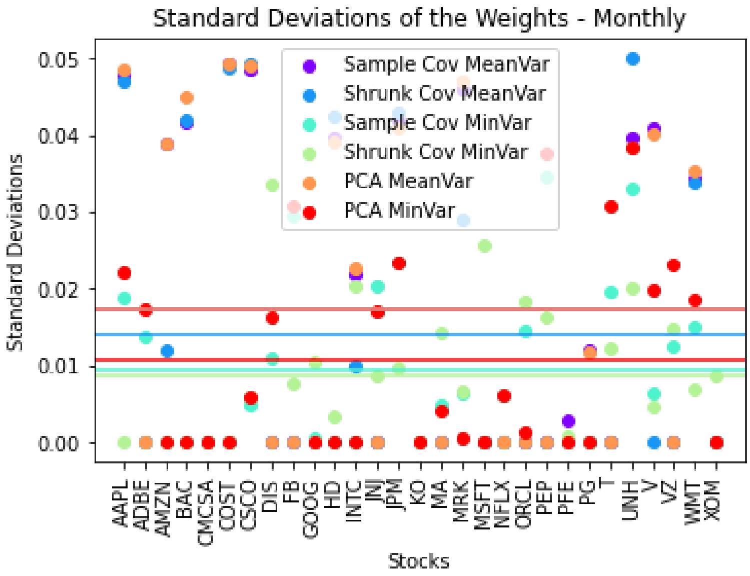
After implementing the aforementioned trading strategy, we observed the cumulative returns after rebalancing on a monthly basis. Based on this, we can see the portfolio constructed using the shrinkage estimator not only has the best cumulative returns at the end of our testing period, but also significantly outperforms most of the other portfolios.



To further analyse these portfolios, we compute their respective information ratios. They are as shown below.

Portfolio	Information Ratio
Sample Mean Variance	1.53
Shrinkage Mean Variance	1.92
Sample Minimum Variance	1.38
Shrinkage Minimum Variance	1.35
PCA Mean Variance	1.47
PCA Minimum Variance	1.42

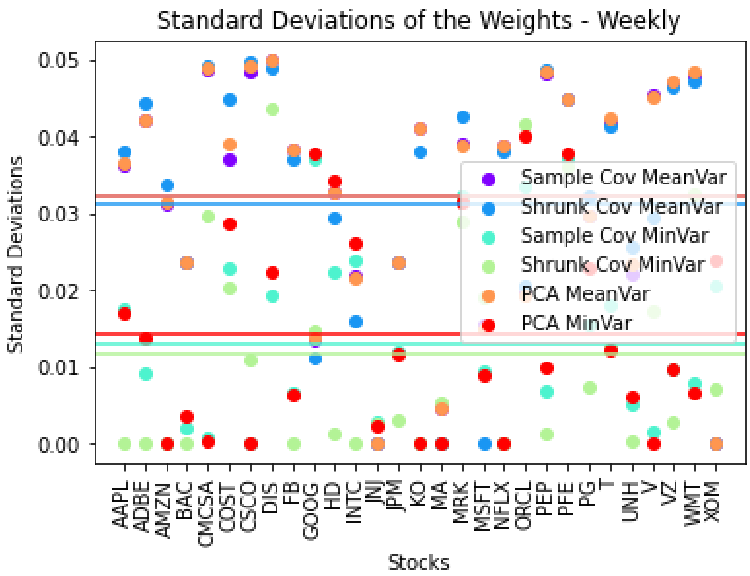
From the table, it is evident that our shrinkage portfolio has a much higher information ratio than it's sample covariance matrix counterpart. For a minimum variance optimization, the performances are near identical. As it is seen that the portfolio based on the shrinkage estimator has the highest information ratio in comparison with the remaining portfolios, we can say that it has the best performance within our investment analysis universe.



Another measure of the stability of the shrinkage method is by analysis of the standard deviations of the weights for each stock. This analysis can show us how stable our weights allotted to stocks are, for different approaches.

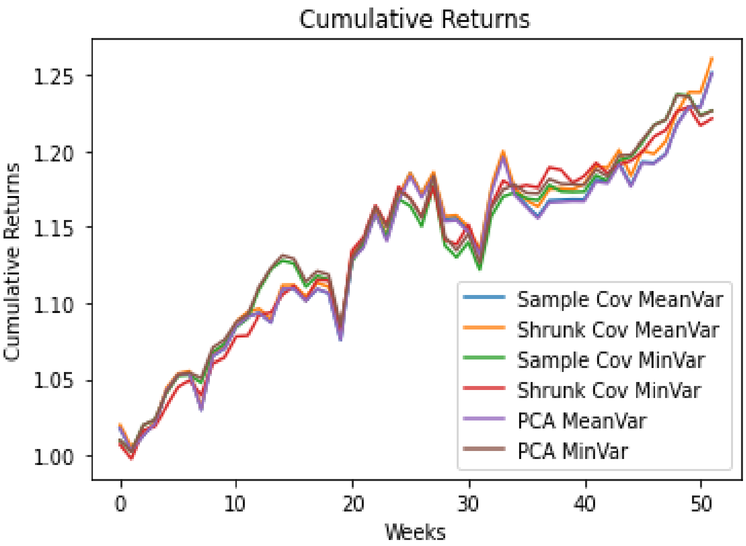
The graph shows the deviations of the weights allotted to each stock and it is evident that our portfolio using shrinkage has one of the lower standard deviations of the lot. On average, the shrinkage estimator minimum variance portfolio has lower weight standard deviation than the mean-variance portfolio using PCA and the sample covariance matrix portfolios. This implies that our weights using shrinkage are more robust compared to those obtained using the sample covariance matrix.

We have seen that the portfolios using shrinkage led to better performance on a monthly rebalancing basis. However to add more weight to our claim we investigated the performance of our portfolios under varying constraints and additionally with weekly data.



We will begin by discussing the performance when using weekly testing and training data. What we found from this was an overall greater performance of shrinkage portfolios on all our criteria.

Looking at the standard deviations of the weights, we can see that shrinkage portfolios have lower average weight standard deviations than both their sample covariance and PCA counterparts (the average from the sample mean variance portfolio is the same as the PCA mean variance average which is why it cannot be seen on the graph). This means that in the weekly scenario, similarly to in the monthly case, the shrinkage portfolios are more stable and therefore will require less rebalancing, which will minimize broker fees and transaction costs.



We additionally considered the cumulative return of the portfolios at the end of the testing period, and found the shrinkage portfolios had a higher cumulative return for both the mean variance and minimum variance portfolios than the sample portfolios. Looking at the graph, the shrinkage portfolios are not only outperforming by the end of the testing period but at each interval too.

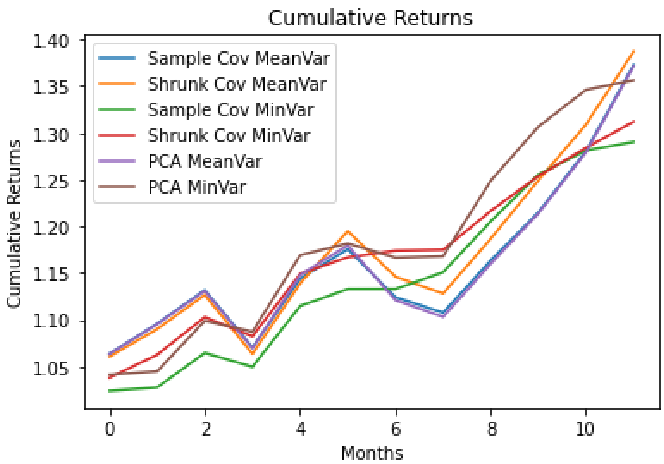
We then took into consideration the information ratio, and though the results overall were much lower than our monthly information ratios, we still saw outperformance by the shrinkage portfolio in the mean variance case. The comparatively lower results may be due to the data being much more noisy when looked at on a weekly rather than monthly basis.

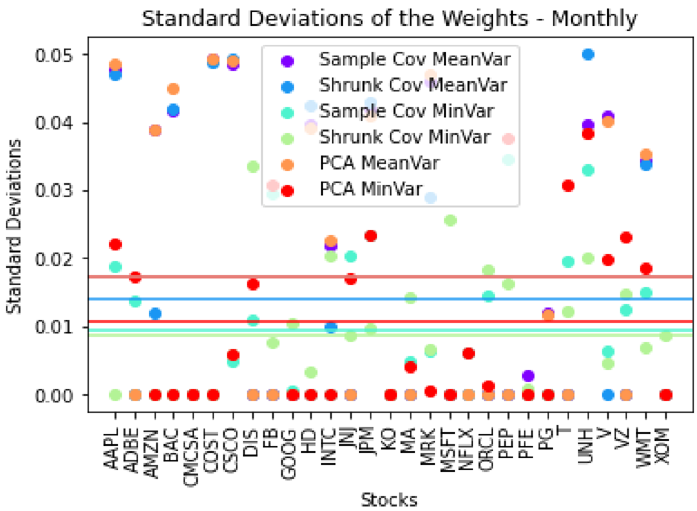
Portfolio	Information Ratio
Sample Mean Variance	-0.11
Shrinkage Mean Variance	-0.07
Sample Minimum Variance	-0.26
Shrinkage Minimum Variance	-0.28
PCA Mean Variance	-0.12
PCA Minimum Variance	-0.26

Therefore, though we are not getting strong performance in comparison to the S&P500 for our weekly data, we are still seeing portfolios using shrinkage perform better than those that do not.

To add further proof to the outperformance of shrinkage portfolios, we investigated the returns of the portfolios with varying constraints. Here we go back to using monthly returns to minimize run time.

Initially when calculating the monthly portfolio returns we had constrained our weights to allow no short selling – all weights must be between 0 and 1. We lifted this constraint to compare the performance of our 2 strategies when only requiring that the portfolio be fully invested.





We can see here that when we lessen the constraints and allow short selling, shrinkage portfolios perform substantially better than those not using shrinkage. Both than shrinkage portfolios have considerably lower average standard deviation of weights (again the average standard deviation for the sample mean variance portfolio is the same as for the PCA mean variance), meaning they are a lot more stable, and the cumulative return in both cases is higher. As shown below, the information ratio is higher for the shrinkage portfolios in both cases as well, and in all cases we have our portfolios beating the S&P500 return.

Portfolio	Information Ratio
Sample Mean Variance	1.589
Shrinkage Mean Variance	1.826
Sample Minimum Variance	1.249
Shrinkage Minimum Variance	1.625
PCA Mean Variance	1.536
PCA Minimum Variance	1.825

Lastly, we investigated the performance when varying the risk aversion coefficient. Standard risk aversion coefficients are generally considered to be between 1 and 3, so here we considered the performance of our portfolios under more extreme risk aversion coefficients, with 0.5 allowing for a risky portfolio and 5 resulting in a very risk averse portfolio.

Risk Aversion Coefficient	Portfolio	Standard Deviation	Cumulative Return	Information Ratio
0.5	Sample Mean Var	0.037	0.971	-0.947
	Shrunk Mean Var	0.032	0.998	-0.820
	Sample Min Var	0.013	1.244	0.749
	Shrunk Min Var	0.011	1.278	1.034
2	Sample Mean Var	0.039	1.202	0.125
	Shrunk Mean Var	0.032	1.219	0.249
	Sample Min Var	0.013	1.244	0.749
	Shrunk Mean Var	0.011	1.278	1.034
5	Sample Mean Var	0.037	1.314	0.962
	Shrunk Mean Var	0.025	1.322	1.018
	Sample Min Var	0.013	1.244	0.749
	Shrunk Min Var	0.011	1.278	1.034

What we can see from the results is that not only in standard cases are shrinkage portfolios outperforming. Regardless of the preferences of the investor, using shrinkage portfolios leads to higher cumulative returns, a higher information ratio and less deviation of weights.

Therefore, we have found that using a shrunk covariance matrix leads to a better performing portfolio consistently. Regardless of the risk aversion of the investor, the reshuffling period, or the constraints on the weights, a shrinkage portfolio will lead to a more stable portfolio with higher returns.

Conclusion

Using the covariance matrix shrinkage technique had a positive effect on our portfolio optimization trading strategies. In all cases, the information ratios were higher than those resulting from the sample covariance matrix strategies. Additionally, the portfolio weights were much more stable for the shrinkage strategies, which means we could expect additional profit through transaction cost savings. These were the primary goals of our project. With respect to our secondary goals of analyzing the returns of these strategies, we observed positive effects in some cases but not all of them. Specifically, using monthly data and rebalancing, the shrinkage strategies outperformed both the sample covariance matrix strategies and the S&P 500 index over the same time periods. This includes both long-only portfolios and portfolios with short-sales permitted. However, this same outperformance was not observed when using weekly data and rebalancing, at least in regards to the S&P 500 index. To be clear, in all cases the strategies using shrinkage outperformed those that did not. We conjecture that the underperformance relative to the S&P 500 can be explained by the fact that the weekly data was simply more noisy than the monthly data.

Bibliography

Ledoit, O., & Wolf, M. (2003). Honey, I Shrunk the Sample Covariance Matrix
MATLAB code to implement the alternative covariance matrix - http://www.ledoit.net/

Appendix A (Shrinkage Intensity Calculation)

Formulas for calculating the transformed covariance matrix from Ledoit & Wolf 2003

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}}$$

$$\bar{r} = \frac{2}{(N-1)N}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} r_{ij}$$

If y is the TxN matrix of returns, then we have the following estimators:

$$\hat{\rho} = \sum_{i=1}^{N}\hat{\pi}_{ii} + \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\frac{\bar{r}}{2}\left(\sqrt{\frac{s_{jj}}{s_{ii}}}\,\hat{\vartheta}_{ii,ij} + \sqrt{\frac{s_{ii}}{s_{jj}}}\,\hat{\vartheta}_{jj,ij}\right)$$

$$\hat{\pi} = \sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\pi}_{ij} \qquad \text{with} \qquad \hat{\pi}_{ij} = \frac{1}{T}\sum_{t=1}^{T}\{(y_{it}-\bar{y}_{i\cdot})(y_{jt}-\bar{y}_{j\cdot}) - s_{ij}\}^2$$

where

$$\hat{\vartheta}_{ii,ij} = \frac{1}{T}\sum_{t=1}^{T}\left\{(y_{it}-\bar{y}_{i\cdot})^2 - s_{ii}\right\}\left\{(y_{it}-\bar{y}_{i\cdot})(y_{jt}-\bar{y}_{j\cdot}) - s_{ij}\right\}$$

$$\hat{\vartheta}_{jj,ij} = \frac{1}{T}\sum_{t=1}^{T}\left\{(y_{jt}-\bar{y}_{j\cdot})^2 - s_{jj}\right\}\left\{(y_{it}-\bar{y}_{i\cdot})(y_{jt}-\bar{y}_{j\cdot}) - s_{ij}\right\}$$

and

$$\hat{\gamma} = \sum_{i=1}^{N}\sum_{j=1}^{N}(f_{ij} - s_{ij})^2$$

With the above estimators,

$$\hat{\kappa} = \frac{\hat{\pi} - \hat{\rho}}{\hat{\gamma}}$$

and the optimal shrinkage intensity is given by

$$\hat{\delta}^* = \max\left\{0, \min\left\{\frac{\hat{\kappa}}{T}, 1\right\}\right\}$$