Evaluating Classical and State-of-the-Art Covariance Matrix Estimation Methods

Luiz Fernando F. Medeiros

October 22, 2019

Abstract

This work evaluates the performance of different estimators in the domain of Portfolio Optimization. Three different estimators, namely sample based (SCE), Ledoit-Wolf (LWE) [1], and Rotationally Invariant Estimators (RIE) [2] are evaluated on their ability to approximate Covariance matrices, as well as how these Covariance matrix approximation affect the portfolio optimization strategy result. The general conclusion is that LWE outperforms SCE and RIE under most conditions, while RIE outperforms in a few different scenarios.

Contents

1 Introduction

Covariance Matrix Estimation is important in a number of fields. The covariance matrix may be an explanation of a time window containing a time varying signal, data composed of different medical information from patient/s over time, or the returns of different assets over a specific time window. Additionally, covariance matrices must be computed in order to apply Principle Component Analysis (PCA) and decomposition to data, perform Exploratory Data Analysis (EDA), and optimization problems such as when one wants to minimize the variance of a set of assets or resources.

The general question surrounding covariance matrices is as follows: is it possible to produce an estimate Σ that is as close as possible to the true covariance matrix $\mathbf{\Sigma}$? The scenario is generally composed of a situation where there is limited knowledge or access to a data matrix of T observations and N variables.

Within this work the domain in focus is Portfolio Optimization (PO) [3]. This means that the information taken into account is matrix $\mathbf{X}$, composed of columns containing the returns $\mathbf{r}_i$, whereby each vector $\mathbf{r}_i$ (for $i \in [1, ..., N]$) represents and asset, and each row (or item in $\mathbf{r}_i$) represents an observed daily return (vector length $|\mathbf{r}_i| = T$). Finally, the general goal is then to find an effective allocation strategy defined by weight vector $\mathbf{w}$, where $\mathbf{w} = \{w_1, ..., w_N\}, w_k \in [0, 1]$, and $\sum_{k=1}^{N}\{w_k\} = 1$, indicates to the investor or portfolio manager what portion of the capital should be allocated to a specific asset.

The work produced here aims to evaluate three different estimators which are considered relevant on this problem at the moment of writing: Sample Covariance Estimator (SCE), Ledoit-Wolf (LW) [1], and Rotationally Invariant Estimators (RIEs) [2]. The latter being considered the state of the art.

The evaluation is done on two fronts. First, how well are these estimators reproducing the true covariance and correlation matrices? Second, what is the impact in the Portfolio Optimization phase?

For the first approach, the main is to investigate the Mean Squared Error (MSE) between the estimated and true. For the latter two approaches, the general method is to backtest strategies which contain these estimators(further details will be provided in section 2).

2 Background

The aim of this section is to present to the reader the definition of the co-varaiance matrix estimators used in this work, along with the perceived logic behind each of the estimators. Covariance Matrix Estimation is most widely performed as a Sample Covariance Estimate (SCE). Ledoit-Wolf Estimator (LWE) [1] and Rotationally Invariant Estimators (RIEs) [2] are highlighted and explored herein due to their relevance and state-of-the-art. However, as mentioned by the authors of both works, there exists other estimators that have been developed over time. The scope of the current work focuses on the three estimators mentioned applied to the problem definitions described below:

Definition 2.0.1. *A data matrix* $\mathbf{X} = \{\mathbf{r}_1, ..., \mathbf{r}_N\}$*, where* N *is number of assets, and* $\mathbf{r}_i = \{r_1, ..., r_T\}$ *where* T *is number of observations to be used for portfolio optimization.*

Definition 2.0.2. *Considering Definition 2.0.1,* T *and* N *are conditioned as follows:*

1. $T \approx N$
2. $q = \frac{N}{T}$ *is not sufficiently small, such to vanish [2].*

The conditions imposed by Definition 2.0.2 indicates that the classical SCE will not produce an estimate that is necessarily close to $\mathbf{\Sigma}$, since a large dimensional problem [1]. In PO, such situations emerge often when a strategy which uses N assets,say $N = 50$, and T observations, say $T = 60$ trading-days-prior (considering the three previous months).

Definition 2.0.3. *With Definition 2.0.2, the problem becomes to find an estimator that produces an estimate* $\mathbf{\Sigma}^\star$ *of the true covariance matrix* $\mathbf{\Sigma}$*, with the smallest expected mean-squared-error. More specifically, we want to:*

$$\text{minimize } \mathbb{E}\Big\{ \|\mathbf{\Sigma}^\star - \mathbf{\Sigma}\|_2^2 \Big\} \tag{1}$$

In order to investigate the problem described in Definition 2.0.3, the performance of SCE, along with LWE and RIE will be explored.

2.1 Sample Covariance Estimator

A sample covariance estimate is one which is based solely on the sampled data. In the current context of interest, it is based solely on the data matrix $\mathbf{X} \in \mathbb{R}^{T\times N}$ containing the return information for a set of assets.

$$\mathbf{S} = \mathbb{E}\{\mathbf{X}^*\mathbf{X}\} - \mathbb{E}\{\mathbf{X}^*\}\mathbb{E}\{\mathbf{X}\} \tag{2}$$

Equation 2 describes SCE [4], where $\mathbb{E}\{\cdot\}$ represents the expectation of the argument, $\mathbf{X}^*$ is the transpose of data matrix $\mathbf{X}$, and $\mathbf{S} \in \mathbb{R}^{N\times N}$ is the respective sample estimate. As discussed in [1] and [2], and justified by the results presented in section 3 SCE is a good estimator as q approaches 0. In such cases, one has enough data such that the method is able to produce a good estimate of the underlying covariance matrix.

However, when such is not the case, SCE becomes unreliable thereby leading users to produce erroneous conclusions about $\mathbf{X}$. In fact, [5] even suggests as the reason for unfortunate portfolio choices.

2.2 Ledoit-Wolf Large-Dimensional Covariance Estimator

LWE builds on the SCE, and the general idea to minimize the difference between a true covariance matrix $\mathbf{\Sigma}$ and an estimate $\mathbf{\Sigma}^\star$. In fact, it is a more precise, constrained description of the optimization problem mentioned in Equation 1. Namely:

$$\begin{aligned} &\underset{\zeta_1,\zeta_2}{\text{minimize}}\ \mathbb{E}\Big\{\|\mathbf{\Sigma}^\star - \mathbf{\Sigma}\|_F^2\Big\} \\ &\text{s.t. } \ \mathbf{\Sigma}^\star = \zeta_1\mathbf{I} + \zeta_2\mathbf{S} \end{aligned} \tag{3}$$

where $\|\cdot\|_F$ illustrates the Frobenius norm, $\mathbf{I}$ is the identity matrix, $\mathbf{S}$ is the sample covariance estimate, and ζ_1 and ζ_2 are not random variables.

Equation 3 is a reflection of Theorem 2.1 in [1]. It demonstrates the thought processing of Ledoit and Wolf, where there exists an optimal linear combination that is able to provide an optimum estimation for the covariance matrix. The basic assumption for Equation 3 is as follows:

$$\begin{aligned}
&\alpha^2 + \beta^2 = \delta^2 \\
&\text{where: } \alpha^2 = ||\mathbf{\Sigma} - \mu\mathbf{I}||_F^2, \\
&\qquad\quad \mu = \mathbf{\Sigma} \cdot \mathbf{I}, \\
&\qquad\quad \beta^2 = \mathbb{E}[||\mathbf{S} - \mathbf{\Sigma}||_F^2], \\
&\qquad\quad \delta^2 = \mathbb{E}[||\mathbf{S} - \mu\mathbf{I}||_F^2]
\end{aligned} \tag{4}$$

Now, Equation 4 reflects Lemma 2.1 provided in [1] where the solution looks as follows:

$$\mathbf{\Sigma}^\star = \frac{\beta^2}{\delta^2}\mu\mathbf{I} + \frac{\alpha^2}{\delta^2}\mathbf{S} \tag{5}$$

$$\mathbb{E}[||\mathbf{\Sigma}^\star - \mathbf{\Sigma}||_F^2] = \frac{\alpha^2\beta^2}{\delta^2} \tag{6}$$

One way of interpret the work of LWE is by means of considering a balancing between bias and variance for our objective. In other words, when pursuing to minimize the objective in Equation 3, we are essentially minimizing the mean squared error, in which the shrinkage target $\mu\mathbf{I}$ is purely bias, and the sample term $\mathbf{S}$ is purely variance [1].

2.3 Rotationally Invariant Estimators (RIEs)

In contrast to the previous two methods, RIE builds on the idea of attempting to estimate a true correlation matrix $\mathbf{C}$ [2]. In a similar fashion to LWE, RIE also uses the sample based estimation. This is done in order to attempt to reach an estimate that produces a smaller error compared to a true correlation matrix $\mathbf{C}$. Bun et al defines the sample based estimate for the correlation matrix as follows:

$$\begin{aligned}
\mathbf{E} &:= \frac{1}{T}\mathbf{X}^*\mathbf{X} \\
&:= \sum_{k=1}^{N} \lambda_k \mathbf{u}_k \mathbf{u}_k^* \in \mathbb{R}^{N\times N}
\end{aligned} \tag{7}$$

In Equation 7 $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N \geq 0$. In addition, vectors $\mathbf{u}_k$ represent the eigenvectors of $\mathbf{E}$. With that efforts are then focused on understanding

[1] If we further ponder over the intuition behind LWE, considering the shrinkage term to be purely bias makes sense from the perspective of having an established baseline. Then, based on how we sample the data, we have an arbitrary, uncontrolled variance, deriving from the sample covariance matrix.

the eigenvalues that compose $\mathbf{E}$. In the case of interest, Definition 2.0.2, the eigenvalues become distorted (compared to the eigenvalues of $\mathbf{C}$). As such, if a method can be conceived where these values are fixed, than the final result should help reconstruct an estimate $\mathbf{\Xi}$ of the correlation matrix that is closer to the true correlation matrix $\mathbf{C}$.

With the above in mind, the following may be stated:

$$\mathbf{\Xi}^{\mathbf{RIE}} := \sum_{\mathbf{k=1}}^{\mathbf{N}} \xi_{\mathbf{k}}^{\mathbf{RIE}} \mathbf{u_k} \mathbf{u_k^*} \tag{8}$$

Equation 8 suggests an estimate $\mathbf{\Xi}^{\mathbf{RIE}}$ which depends on a modified eigenvalue ξ_k^{RIE}. The new eigenvalue is further defined as:

$$\xi_k^{RIE} = \frac{\lambda_k}{|1 - q + q z_k s_k(z_k)|^2} \tag{9}$$

where $q = N/T$ (as in Definition 2.0.2), $z_k = \lambda_k - \frac{i}{\sqrt{N}}$, and $k \in [1, N]$.

Now, Equation 9 suggests that one should re-scale the eigenvalues. The work of Marcenko & Pastur [6] propose that $\mathbf{E}$ contains a spectrum that is broader than $\mathbf{C}$. Due to this scaling, one should expect small eigenvalues to be smaller than those of the true matrix $\mathbf{C}$ (and large eigenvalues to be larger).

Next, one of the main weighing components which considers the difference between eigenvalues is defined:

$$\begin{aligned} s_k(z_k) &= \frac{1}{N} \sum_{\substack{j=1 \\ j \neq k}}^{N} \frac{1}{z_k - \lambda_j} \\ &= \frac{1}{N} \sum_{\substack{j=1 \\ j \neq k}}^{N} \frac{1}{\lambda_k - \lambda_j - \frac{i}{\sqrt{N}}} \end{aligned} \tag{10}$$

Although the achieved estimate in Equation 9 may work for a set of cases, and therefore can be the final RIE, it generally adds a bias to the final solution. This bias can be significant if the number of variables in being considered are not larger than 400 [5]. As such, Bun et al proposed a de-biasing term Γ_k for eigenvalue estimate ξ_k^{RIE}. It is defined as follows:

$$\Gamma_k = \sigma^2 \frac{|1 - q + q z_k g_{mp}(z_k)|^2}{\lambda_k} \tag{11}$$

where $\sigma^2 = \frac{\lambda_N}{(1-\sqrt{q})^2}$, and λ_N is the smallest eigenvalue found through through the sample based eigenvalue decomposition.

Now, the de-biasing term also depends on another variable that has not yet been defined: $g_{mp}(\cdot)$. This is the Stieltjes transform of the argument. This term is to bring information about how the current eigenvalue λ_k measures in the underlying distribution. This highlights once again the authors emphasis on the idea of re-balancing eigenvalues in order to reach a better estimation. Article [5] defines g_{mp} as:

$$g_{mp}(z) = \frac{z + \sigma^2(q-1) - \sqrt{z-\lambda_N}\sqrt{z-\lambda_+}}{2qz\sigma^2} \tag{12}$$

where $\lambda_+ = \lambda_N \left(\frac{1+\sqrt{q}}{1-\sqrt{q}}^2 \right)$

With the de-biasing term defined, the final de-biased estimator is defined as follows:

$$\hat{\xi_k} := \xi_k^{RIE} \times \max(1, \Gamma_k) \tag{13}$$

The final de-biased estimation of the true correlation matrix **C** becomes:

$$\mathbf{\Xi^{RIE}} := \sum_{\mathbf{k=1}}^{\mathbf{N}} \hat{\xi}_{\mathbf{k}} \mathbf{u_k} \mathbf{u_k^*} \tag{14}$$

It should be noted, however, that in order to reach an estimation $\mathbf{\Xi^{RIE}}$ which produces good results, one need to perform an extensive deal or pre-processing for data **X**. Namely, one must remove the sample mean the assets and normalize the result by the sample volatility estimate. Further details about this pre-processing procedure is presented in [2] and [5].

3 Test Results

This section will present the results achieved while testing the different covariance estimators.

3.1 Global Mean Variance Porfolio

In this section, the performance achieved by the multiple estimators will be presented in the context of a Global Mean Variance Portfolio (GMVP) strategy [3]. The goal with GMVP is to minimize the risk of a portfolio. A GMVP which follows a long-only strategy, along with with no borrowed margin may be described as the following optimization problem:

$$\begin{aligned} \underset{\mathbf{w}\in\mathbb{R}^T}{\text{minimize}} \ & \mathbf{w}^\top \mathbf{\Sigma} \mathbf{w} \\ \text{subject to} \ & \mathbf{1}\cdot\mathbf{w} = 1 \\ & \mathbf{w} \geq \mathbf{0} \end{aligned} \tag{15}$$

Equation 15 contains a $\mathbf{\Sigma}$ which indicates the covariance matrix for a group of assets, and $\mathbf{w}$, the portion of capital allocated to a specific set of assets. The goal with the constraints is to simplify the problem, since adding ideas such as shorting and margin investing would further complicate the experiment.

The overall performance of the estimators are presented is presented in the images below. The first image demonstrates the results for total capital return of the GMVP strategy, re-balanced monthly. This means that every month, the portfolio was analyzed, a covariance matrix was computed (on the basis of a 6 month lookback period), and finally re-weighted according the newly achieved weights. In addition, the graphs demonstrate results from varying number of covariance estimation methods, and number of assets. This was specifically done to analyze the different covariance estimations under a chosen number of assets (10, 50, 100, 150, and 200).

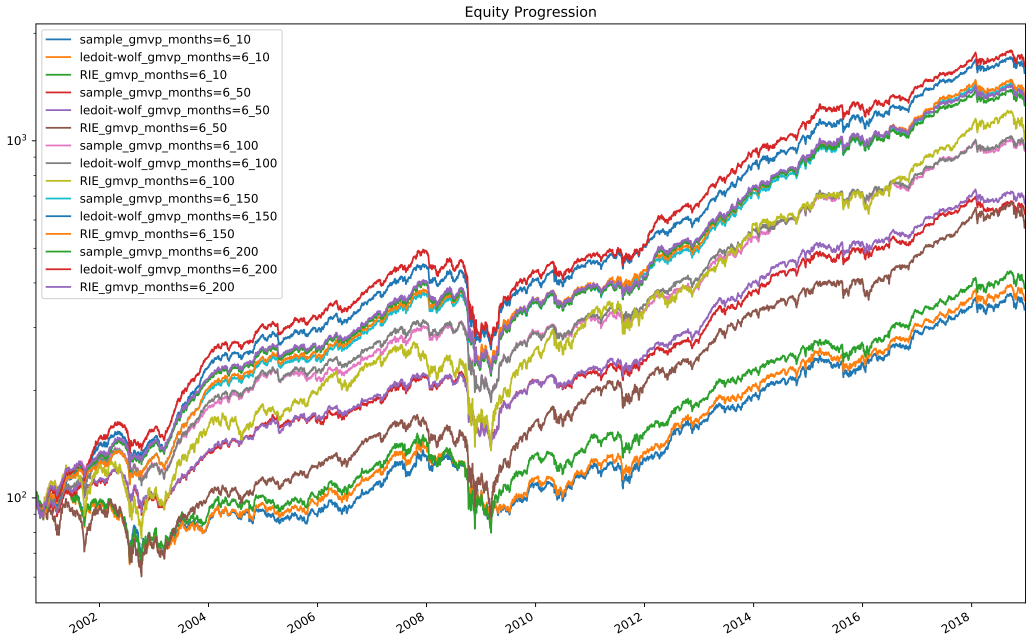

Figure 1: Equity progression for the different covariance estimation methods tested. Note that the y axis is on the logarithmic scale.

The general message from Figure 1 is that there is a positive trend. The methods, when used in the strategy to produce the covariance matrices, produce a sufficient level of estimation such to maintain positive returns.

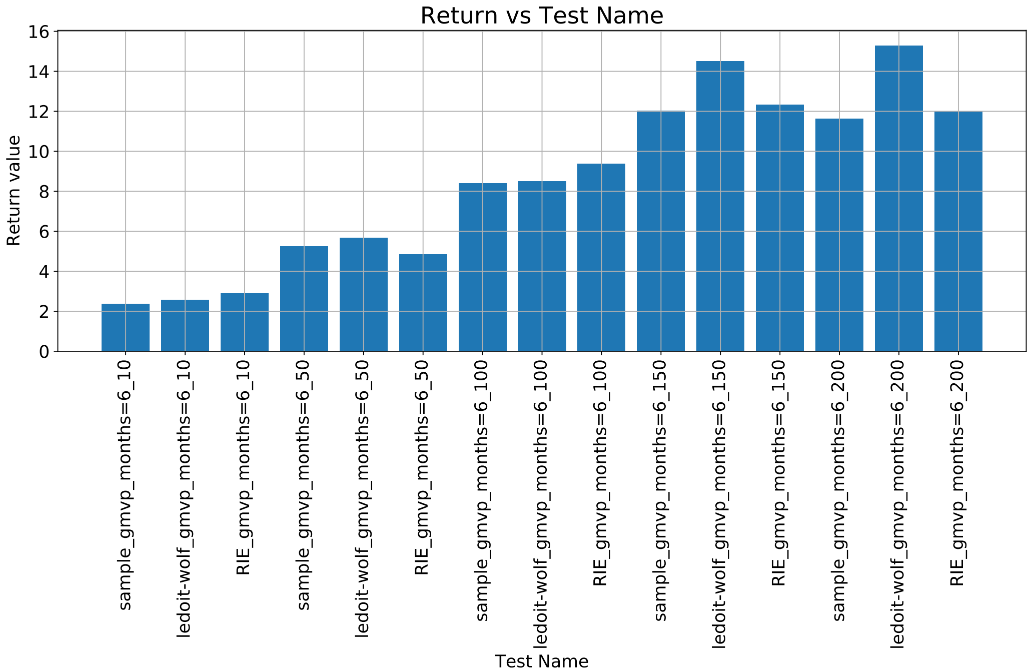

Figure 2: Demonstrating the returns. The y axis represents raw returns.

Figure 2 shows in more detail the added value between the different covariance method estimators. The returns are calculated by considering the capital achieved at the end of the investment period, with the initial capital as reference. The largest return was achieved by the LWE, with 200 assets. The general trend shows that the larger the number of assets, the larger the returns.

Finally, by building on the above findings, its then possible to investigate which lookback period would produce the smallest risk (since that is what GMVP aims to minimize).

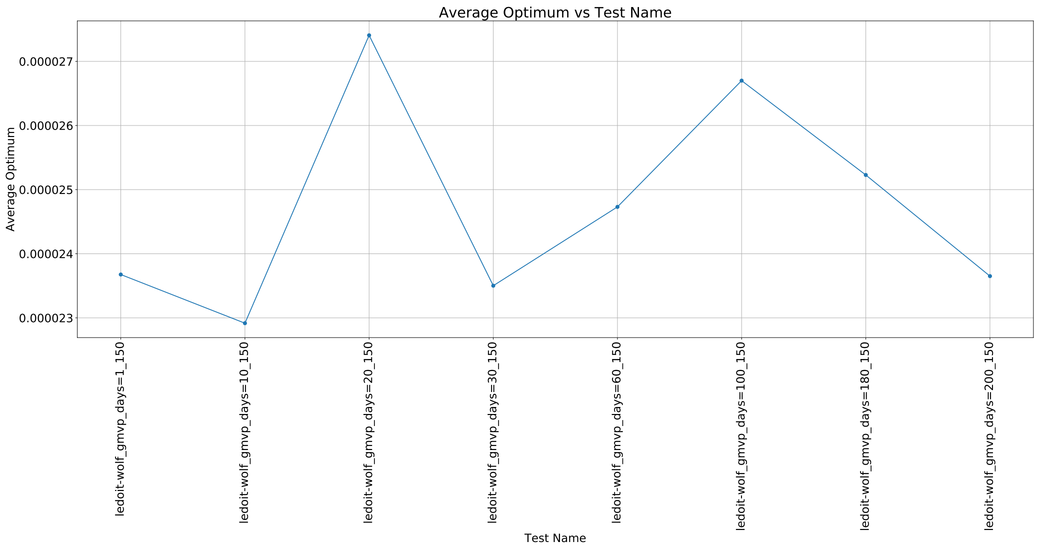

Figure 3: The average optimum illustrates the average minimum risk calculated over the period of investment. The Test Name in this case refers to the varying lookback periods that was investigated.

Figure 3 provides insight into how many days worth of data one should consider in order to produce the smallest risk in a GMVP portfolio. It turns out that 10 days is the right amount of information to take into consideration, while 20 produces the worst results.

3.2 Most Diversified Portfolio

This section brings about the results achieved for the Most Diversified Portfolio (MDP) strategy [7]. MDP builds on the idea that it is possible to produce a ratio of that indicates how diversified a portfolio is based on its volatility vector and covariance matrix description. By maximizing this description, one achieves a portfolio that assures the maximum diversification. The diversification ratio $D(\mathbf{w})$ for a portfolio $\mathbf{w}$ can be described as follows:

$$D(\mathbf{w}) = \frac{\mathbf{w}^T \mathrm{diag}(\mathbf{\Sigma})}{\sqrt{\mathbf{w}^T \mathbf{\Sigma} \mathbf{w}}} \tag{16}$$

It is not possible to directly compute the diversification ratio with the constraints imposed for the GMVP strategy. As a consequence, for this strategy the constraints are dropped, and therefore attempt to maximize $D(\mathbf{w})$ without weight constraints. This means that it is possible to have negative weights. The results which follow reflect the performance of different covariance estimators applied to the MDP problem.

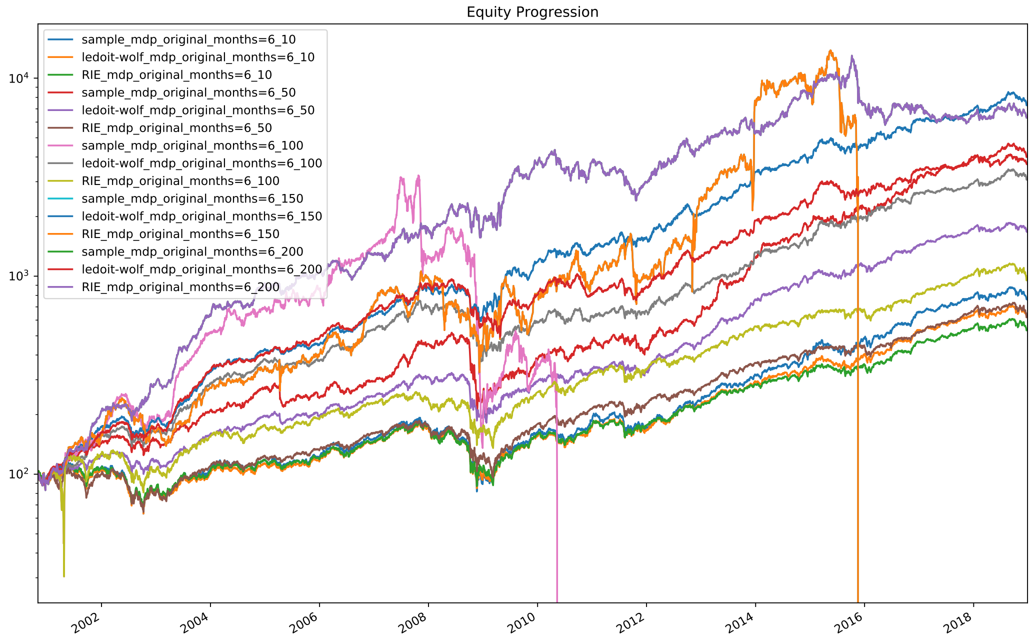

Figure 4: The Equity Progression plot for the MDP strategy shows returns higher the GMVP. However, it also demonstrates that some of the strategies had very significant low peaks.

For most cases, MDP provides superior returns compared to the GMVP. However, for a number of tests, the returns turned negative. the results below demonstrates the achieved returns in more details:

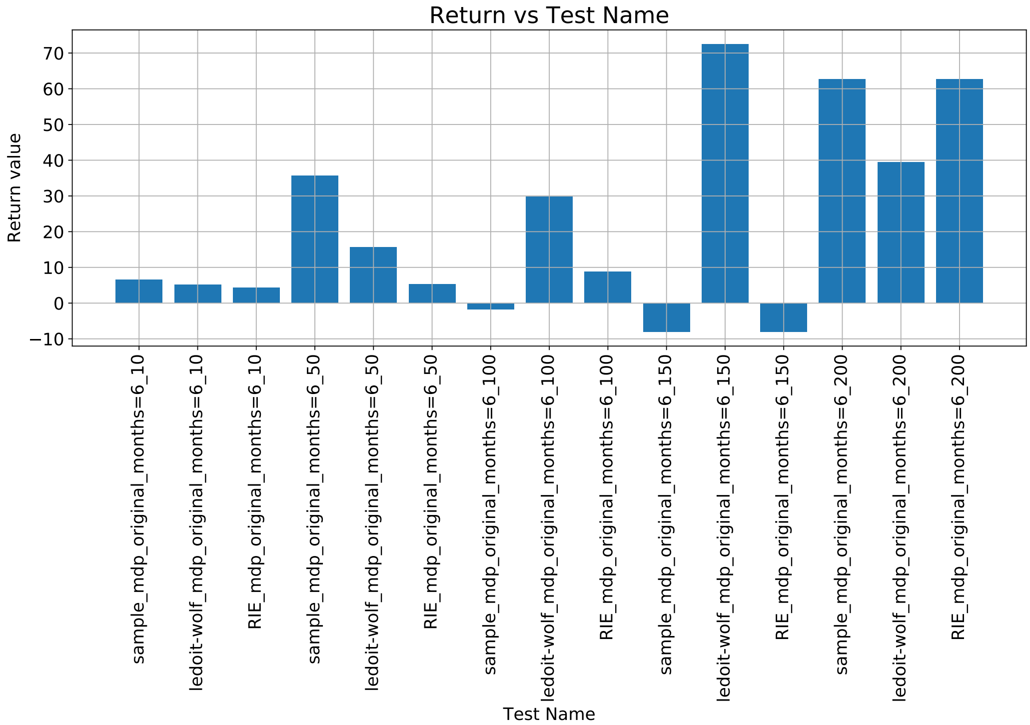

Figure 5: Detailed returns for MDP strategy.

Figure 5 shows that for 150 assets, SCE and RIE provided negative returns when applied to the MDP strategy. Moreover, SCE also provided negative returns for 100 asset strategy. However, in comparison, for 150 assets, LWE provided returns that surpassed 70 times the initial investment. That is significantly better than the best GMVP return.

Now, let us consider how average optima changed over different lookback periods. Figure 6 shows that the highest diversification ratio achieved was for a lookback period of 60 days, while the lowest was achieved for 200 days.

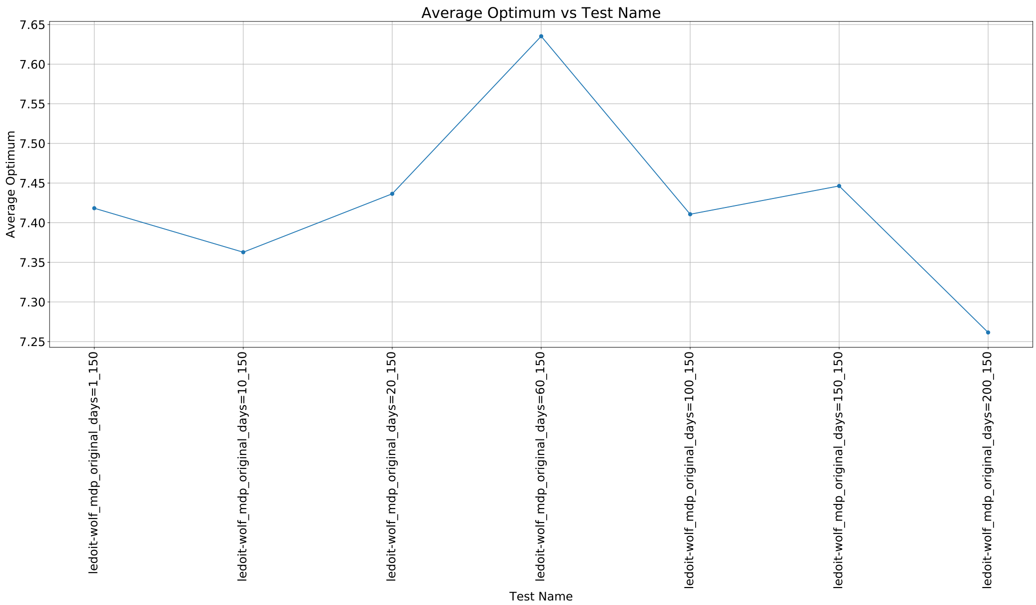

Figure 6: Average Optima for different lookback periods in the MDP strategy. In this case, the higher the better.

3.3 Modified MDP Strategy

During this experiment, a third strategy was devised and tested in order to evaluate the covariance estimation methods. This strategy was developed based on the ideas proposed by the MDP. We will call it Modified Maximum Diversification Portfolio strategy (MMDP). The idea is similar to MDP, in the sense of maximizing the diversification ratio, however, in this strategy the risk is constrained to be less than a certain established criteria. The logic behind the strategy is as follows:

$$\begin{aligned} \underset{\mathbf{w}\in\mathbb{R}^T}{\text{maximize}}\ & \mathbf{w}^\top \text{diag}(\mathbf{\Sigma}) \\ \text{subject to}\ & \mathbf{1}\cdot\mathbf{w} = 1 \\ & \mathbf{w} \geq \mathbf{0} \\ & \mathbf{w}^\top \mathbf{\Sigma}\mathbf{w} \leq \frac{1}{1+c}, \forall c \neq -1 \end{aligned} \tag{17}$$

For this experiment, the above varialbe c was chosen to be 0.50. In short, the constant c controls how much the user is willing to deviate from the optimum $D(\mathbf{w})$, with the constraint of risk, and an intuition that the higher returns may not necessarily lie in the optimum $D(\mathbf{w})$. The equity progression graph presented on Figure 7 demonstrates the results achieved for this strategy.

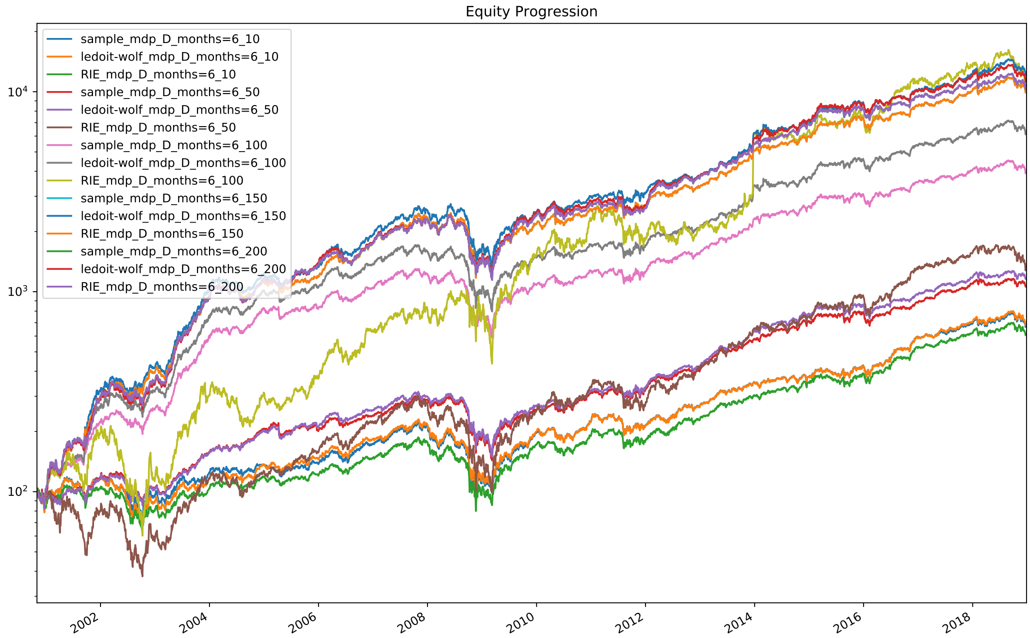

Figure 7: MMDP equity progression over all data.

MMDP produces positive returns, while maintaining a rising equity value trend. Figure 8 provides further details into the overall returns:

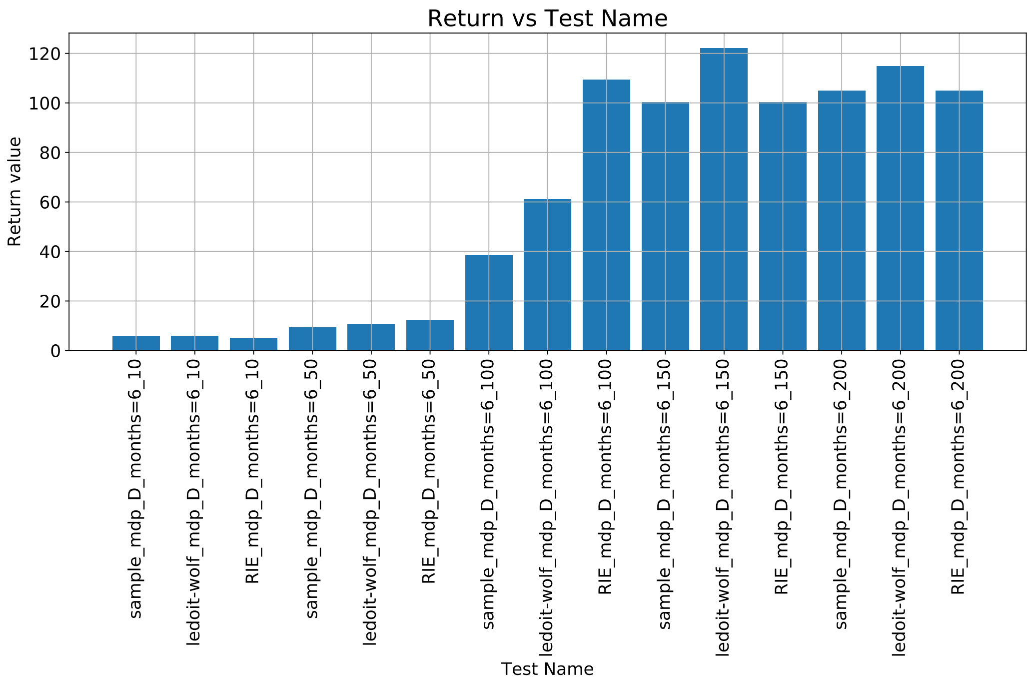

Figure 8: MMDP overall return for every covariance method tested.

It's possible to note that once again LWE provided the largest returns for a 150 asset portfolio, while maintaining consistency in all other asset numbers. RIE followed closely, while producing the largest returns for 100 asset portfolio. It's also possible to note that asset numbers below 100 provide significantly lower returns. Finally, not only doe MMDP provide higher returns than MDP, it also manages to bring about positive returns for all covariance matrix estimators.

Now, although this strategy does not aim to reach the maximum $D(\mathbf{w})$ or the minimum risk, it's still interesting to analyze the achieved diversification ratio for different lookback periods.

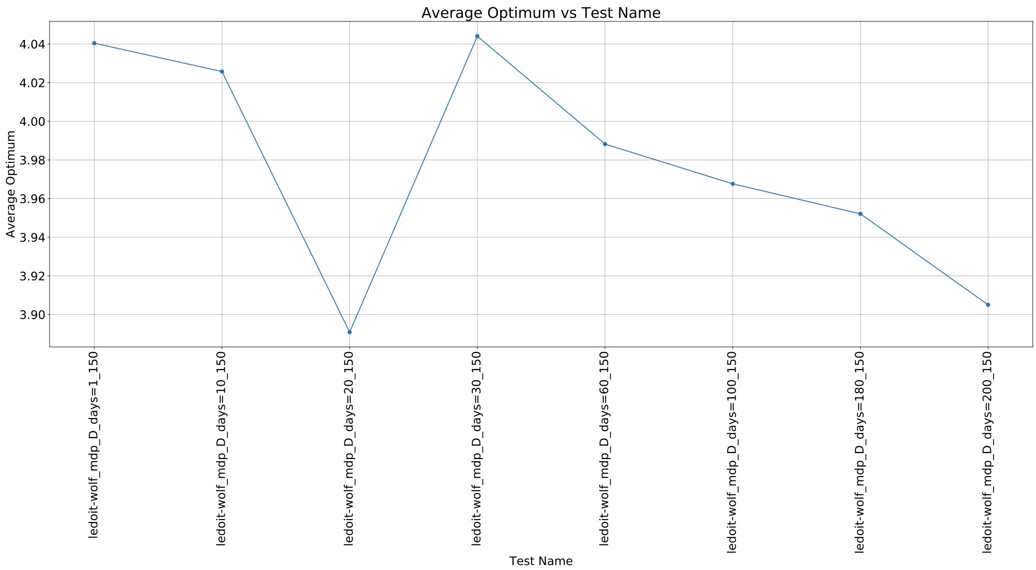

Figure 9: MMDP average optima achieved over different lookback periods. In this case, the higher only means that a portfolio is further diversified.

Figure 9 shows that the highest diversification ratio achieved was at a lookback period of 30 days, while the minimum was at 20 days. It's also possible to note the there exist a downward trend in the diversification ratio as the lookback period increases from its highest point.

4 Conclusion and Future Work

The initial aim of the project was to apply and evaluate the performance of SCE, LWE, and RIE on GMVP and MDP. To this end, the conclusion was that LWE produces the better performance for most cases, and is therefore the recommended estimator to be used when producing returns based covariance matrix estimation.

On a different note, while learning and testing the different estimators and portfolio optimization strategies we stumbled upon MMDP. A different strategy based on the original MDP work that managed to produce superior returns for every covariance matrix estimator. This leads the way to the aim for future work: further evaluation of MMDP, including under the different covariance matrix estimators, while formalizing the results for a possible publication.

References

[1] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, vol. 88, no. 2, pp. 365–411, 2004.

[2] J. Bun and J. P. Bouchaud, "Cleaning correlation matrices matrices," no. April, 2016.

[3] A. Meucci, *Risk and Asset Allocation.* Springer Finance, 2005.

[4] B. Hajek, *Random Processes for Engineers.* Cambridge University Press, 2015.

[5] J. Bun, J. P. Bouchaud, and M. Potters, "Cleaning large correlation matrices: Tools from Random Matrix Theory," *Physics Reports*, vol. 666, pp. 1–109, 2017.

[6] V. A. Marčenko and L. A. Pastur, "Distribution of Eigenvalues for Some Sets of Random Matrices," *Mathematics of the USSR-Sbornik*, vol. 1, no. 4, pp. 457–483, 1967.

[7] Y. Choueifaty and Y. Coignard, "Toward maximum diversification," *Journal of Portfolio Management*, vol. 35, no. 1, pp. 40–51, 2008.