



# Evaluating Various Portfolio Optimization Strategies using High-Dimensional Covariance Matrix Estimators

Ruchir Sharma

**Abstract:** We compare the performance of multiple covariance matrix estimators for the purpose of portfolio optimization. This evaluation studies the ability of estimators like Sample Based Estimator (SCE), Ledoit-Wolf Estimator (LWE), and Rotationally Invariant Estimators (RIE) to estimate covariance matrix and their competency in fulfilling the objectives of various portfolio allocation strategies. In this paper, we have captured the effectiveness of strategies such as Global Minimum Variance (GMVP) and Most-Diversified Portfolio (MDP) to produce optimal portfolios. Additionally, we also propose a new strategy inspired from MDP: Most-Diversified Portfolio (MMDP), that enables diversification upon minimizing risk. Empirical evaluations show that by and large, MMDP furnishes the maximum returns. LWE are relatively more robust than SCE and RIE but RIE performs better under certain conditions.

**Keywords:** Covariance Matrix Estimators, Sample Based Estimator (SCE), Ledoit-Wolf Estimator (LWE), Rotationally Invariant Estimators (RIE), Global Minimum Variance (GMVP), Most-Diversified Portfolio (MDP), Most-Diversified Portfolio (MMDP)

## I. INTRODUCTION

The covariance matrix is, arguably, the second most important object in all of statistics. [1] It can be used to analyze the movement of random variables like two stocks or compute how the return of two assets move together. Apart from that, it facilitates the application of Principal Component Analysis for robust pattern recognition, which further helps with Exploratory Data Analysis (EDA). When one deals with very large random matrices (such as covariance matrices), one expects the spectral measure of the matrix under scrutiny to exhibit some universal properties, which are independent of the specific realization of the matrix itself. This property is at the core of Random Matrix Theory (RMT), which provides a very precise description of the convergence of the spectral measure for a very large class of random matrices. [2] Estimating covariance matrices is an important part of portfolio selection, risk management, and asset pricing. [3] This paper explores the use cases of covariance matrix estimation in Portfolio Optimization (PO). Here, we consider the input data matrix  $\mathbf{M}$  consisting  $N$  observations and  $V$  variables. It is composed of columns comprising returns, wherein every vector (for  $i \in [1, \dots, N]$ )

represents an asset, and each row represents an observed daily return (vector length  $| \cdot | = 0$ ). Our objective is to come forward with a viable allocation strategy defined by weight vector  $\mathbf{g}$ , where  $\mathbf{g} = (g_1, \dots, g_V)^N$ ,  $g_k \in [0,1]$  and  $\sum_{k=1}^V \{g_k\} = 1$ , to ascertain the measure of capital allotment to a particular asset. On the grounds of the aforementioned, we evaluate three covariance matrix estimators: Sample Covariance Estimator (SCE), Ledoit-Wolf Estimator (LWE), and Rotationally Invariant Estimators (RIEs). The sole concern of the research is to assess the ability of these matrix estimators to fulfil the goals of portfolio optimization strategies.

## II. HIGH-DIMENSIONAL COVARIANCE MATRIX ESTIMATORS

This section emphasizes the various covariance matrices that we consider in this project to fulfil the requirements of various portfolio allocation strategies. SCE happens to be the most commonly used estimator, however, we also go over LWE and RIEs in this section. Here, we apply these three covariance matrix estimators to the following problem definitions:

**Definition 2.0.1.** A data matrix  $\mathbf{M} = [r_1, \dots, r_v]$  where  $\mathbf{M} \in \mathbb{R}^{N \times V}$ ,  $V$  is the number of assets,  $N$  is the number of observations to be used for portfolio optimization,  $r_v = [r_{1v}, \dots, r_{Nv}]$ , is the vector containing observation  $v$  for each of the  $V$  assets.

**Definition 2.0.2.** Based on the above definition,  $N$  and  $V$  relate to each other as follows:  $N \approx V$ . So,  $q = \frac{V}{N}$  is not sufficiently small, so  $N = O(V)$ .

Furthermore, true variance between assets  $v$  and  $z$  can be defined as follows:

$$\text{Cov}(v, z) = \mathbb{E}[v^T z] - \mathbb{E}[v]^T \mathbb{E}[z], \dots \quad (1)$$

and the true variance matrix of data matrix  $\mathbf{M}$  as:

$$\Sigma = \mathbb{E}[\mathbf{M}^T \mathbf{M}] - \mathbb{E}[\mathbf{M}]^T \mathbb{E}[\mathbf{M}], \quad (2)$$

Where  $\Sigma \in \mathbb{R}^{V \times V}$ , symmetric, and positive semi-definite.  $\mathbb{E}[\cdot]$  can be defined as the expectation of argument. Also,  $\mathbf{M}$  is composed of the entities  $z$  and  $v$ .

**Definition 2.0.3.** The above definition presents us with the problem of finding an estimator that renders and estimate  $\Sigma^*$  of the true covariance matrix  $\Sigma$ , having minimal expected mean-squared-error. In other words:

$$\text{minimize } \mathbb{E} \{ \|\Sigma^* - \Sigma\|_F^2 \} \quad (3)$$

To delve deeper into the above problem definition, we explore the performance of SCE, LWE and RIE in the following sections.

### A. Sample Covariance Matrix Estimator

Revised Manuscript Received on July 10, 2020.

\* Correspondence Author

Ruchir Sharma\*, Department of Computer Science Engineering, Thapar Institute of Engineering and Technology, Patiala, Punjab, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

When the dimension of the covariance matrix is large, the estimation problem is generally challenging. It is well-known that the sample covariance based on the observed data is singular when the dimension is larger than the sample size. [4] One recurrent problem is that the traditional estimator (that is, the sample covariance matrix) is ill-conditioned and performs poorly when the number of variables is not small compared to the sample size. [5] A sample covariance matrix  $S$  is defined as follows:

$$S = \frac{1}{N} \cdot M^T M - \bar{r} \cdot \bar{r}^T, \quad (4)$$

where  $\bar{r} = \frac{1}{V} \sum_{v=1}^V r_v$ ,  $\bar{r} \in \mathbb{R}^V$ , and  $r_v$  is the row vector at row  $v$  in  $M$

### B. Ledoit-Wolf Large-Dimensional Covariance Estimator

LWE builds on the idea that it is possible to minimize the difference between a true variance matrix  $\Sigma$  and an estimate  $\Sigma^*$ . The logic behind LWE to minimize the Mean Squared Error (MSE) can be described as follows:

$$\text{minimize } \mathbb{E} \{ \|\Sigma^* - \Sigma\|_F^2 \} \quad (5)$$

$\zeta I, \zeta 2$

$$\text{s.v. } \Sigma^* = \zeta I I + \zeta 2 S \quad (6)$$

where  $I$  is the identity matrix,  $S$  is the sample matrix, and  $\zeta I$  and  $\zeta 2$  are constants.

The prime objective now is to find an optimal linear combination of the SCM and the identity matrix that can offer an optimum estimation for the covariance matrix. The constants  $\alpha$ ,  $\beta$ , and  $\gamma$  can be defined as follows:

$$\alpha^2 + \beta^2 = \gamma^2$$

where,  $\alpha^2 = \|\Sigma - \mu I\|_F^2$

$$\mu = \text{tr}(\Sigma)/V,$$

$$\beta^2 = \|S - \Sigma\|_F^2$$

$$\gamma^2 = \|S - \mu I\|_F^2$$

The MSE can be described as follows:

$$\Sigma^* = \frac{\beta^2}{\gamma^2} \mu I + \frac{\alpha^2}{\gamma^2} \mu I \quad (7)$$

$$\mathbb{E} \{ \|\Sigma^* - \Sigma\|_F^2 \} = \frac{\alpha^2 \beta^2}{\gamma^2} \quad (8)$$

### C. Rotationally Invariant Estimators (RIEs)

RIE can be classified among those estimators that are optimal from a decision theoretic point of view and can deal with the presence of outliers. More specifically, we shall only consider RIEs that depend on the data we have which is represented through the sample covariance matrix  $S$ . That being said, these estimators own the following properties: it shares the same eigenvectors than  $S$  and its eigenvalues are functions of the sample ones. [6]

The de-biased estimation of the true correlation matrix  $M$  can be described as follows:

$$\Xi^{RIE} := \sum_{k=1}^V \xi_k u_k u_k^* \quad (9)$$

where estimate  $\Xi$  is close to true correlation matrix,  $u_k$  are eigenvectors and

$$\xi_k^{RIE} = \frac{\lambda_k}{|1 - q + q z_k s_k(z_k)|^2}, \quad (10)$$

where,  $q = V/N$ ,  $z_k = \lambda_k - \frac{i}{\sqrt{V}}$ ,  $i^2 = -1$ ,

and  $k \in [1, T]$

It is important to point out that in order to reach an estimation  $\Xi^{RIE}$  which produces good results, we should perform extensive preprocessing of data  $M$ .

## III. METHODOLOGY AND RESULTS

### A. Data Retrieval

We retrieved data from Yahoo Finance which consisted of assets related to American and European exchanges. We selected the assets on a random basis and matched the date values such that every asset  $i$  has a value for the same day  $t$ . The following table is a representation of the original data matrix  $M_{raw}$ .

Each column represents a particular asset and the rows are composed of the respective closing values. We used the above values to convert them into daily return values using the following logic:

$$M_{raw} = \frac{M_{raw}[t+1,i] - M_{raw}[t,i]}{M_{raw}[t,i]} \quad (11)$$

$M_{raw}$  is composed of 4409 observations and 207 assets, which implies:

$$M_{raw} \in \mathbb{R}^{4409 \times 207}$$

We proceed further by varying estimators and dimensions like number of assets and data points considered while re-computing the covariance matrix to assess the resultant output of the strategies being used. We study the observations pertinent to the past month containing 22 trading days  $[t-22, t]$  or last three months  $[t-66, t]$ . In the end, we also take the effects of the various assets into account.

### B. Global Minimum Variance Portfolio

The global minimum variance portfolio computed using the sample covariance matrix is known to be negatively affected by parameter uncertainty, an important component of model risk. [7] We present the performance of various estimators in fulfilling the goals of the Global Minimum Variance Portfolio [GMVP] strategy in this section. The objective of

	F	L	MSFT	S	NVDA	ANSS	VZ	WMT	MRK	EQV1V_H	EQV2V_H
2000-11-02 00:00:00+02:00	15.048508	11.346896	22.8218	17.588865	10.2844	2.9075	21.8678	34.0712	43.3611	0.963153	
2000-11-03 00:00:00+02:00	14.615457	10.939939	22.1531	18.663744	10.9564	2.9075	21.5838	33.3256	42.7529	1.016682	
2000-11-06 00:00:00+02:00	15.265035	11.083570	22.5589	18.321737	11.8021	2.9375	21.0385	34.4650	43.7843	1.177187	
2000-11-07 00:00:00+02:00	14.435014	11.330938	22.8835	19.298891	11.0040	2.8125	20.8038	34.4228	42.2665	1.230696	
2000-11-08 00:00:00+02:00	14.326763	11.474570	22.5394	19.885189	10.6780	2.8750	20.8492	34.2892	44.1784	1.123679	

this strategy is to optimize returns by minimizing the risk. The logic behind this strategy is as follows:

$$\text{minimize } g^T \Sigma g$$

$$g \in \mathbb{R}^N$$

$$\text{s.t. } 1 \cdot g = 1$$

$$(12)$$

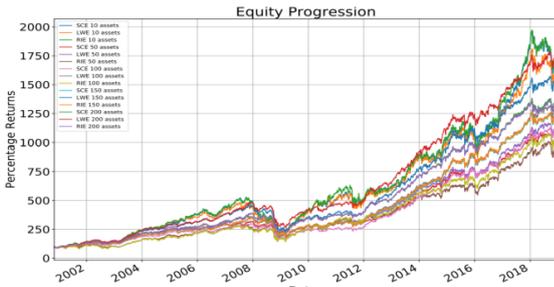
$$g \geq 0.$$

$\Sigma$  denotes the covariance matrix in the above equation for a group of asset returns, and  $w$  is the measure of capital allotted to a group of assets. For an unconstrained case, the direct solution is as follows:



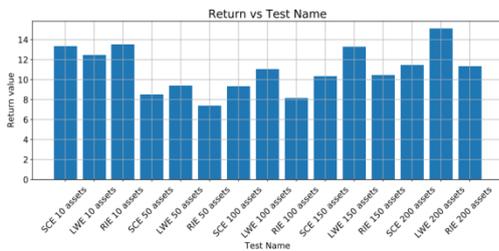
$$g = \frac{\Sigma^{-1}1}{1^T \Sigma^{-1}1} \tag{13}$$

A visual depiction of the overall performance of the estimators is presented through the images below.



The image above shows the total capital return of the GMVP strategy, rebalanced monthly. Meaning that, the portfolio was analyzed every 22 days on the grounds of a 6-month lookback period and a covariance matrix was computed which was finally reweighted based on the newly achieved weights. It demonstrates how total asset returns change with varying covariance matrix estimators and number of assets  $V$  with observations  $N$ , and a given optimization strategy.

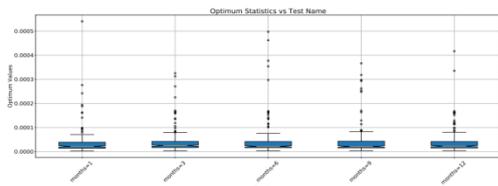
From the image, we can conclude that the estimators fulfil the goals of the strategy to produce an estimation that renders positive returns.



The figure mentioned above is composed of the frequency of return on equity for a GMVP strategy and a 6 months look-back window on the y axis. We have chosen assets on a random basis from matrix  $M$ .

We calculate the returns by considering the capital accumulated at the end of investment period with the initial capital. LWE fetched the largest return with 200 assets and the trends evidence that the number of assets is directly proportional to returns.

These findings help in investigating which lookback period can produce the minimum risk.



The above boxplot depicts how a change in the number of lookback months produced variation in risk.

We choose 150 assets, LWE, and GMVP strategy in this case. The reason behind selecting LWE and 150 assets is that it worked well for the previous strategy on a random set of assets. Most risk values achieved occupy the 25-75

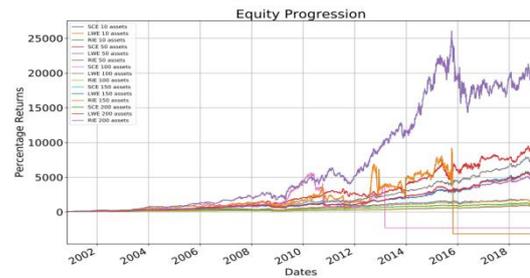
percentile margin and we have a smaller outlying risk for the three months look-back.

### C. Most Diversified Portfolio

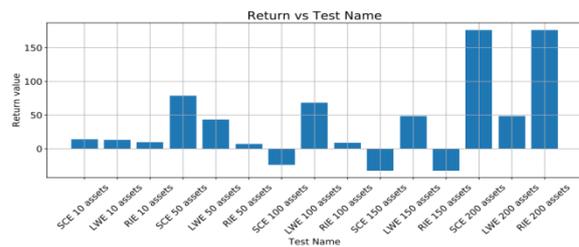
Most Diversified Portfolio (MDP) is based on the premise that volatility vector and covariance matrix description can signify how diversified a portfolio is. This strategy helps in producing maximum diversification in a portfolio. The diversification ratio  $D(g)$  for a portfolio  $g$  is stated below:

$$D(g) = \frac{g^N \text{diag}(S)}{\sqrt{g^N S g}} \tag{14}$$

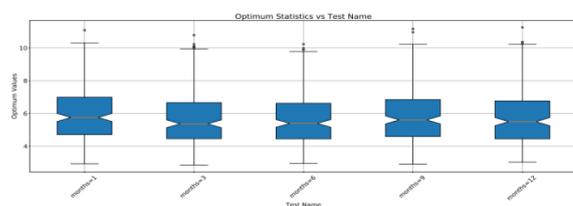
Given that the diversification ratio cannot be computed as a result of the constraints pertinent to GMVP strategy, therefore this strategy attempts to maximize  $D(g)$  without such constraints. The images which follow represent the performance of covariance matrix estimators to fulfil the objects of the MDP strategy.



The above Equity Progression plot having a 6 month look-back period results from applying the MDP strategy. At certain points, MDP outperforms GMVP but the returns turn negative for various tests including SCE 150 assets and RIE 150 assets. Also, there are no constraints to the MDP strategy and the assets are selected on a random basis which implies that there is an additional risk associated with the strategy upon combining with estimators. The results that follow described the returns achieved in more details:



The Return vs Test Name graph shows that for 150 assets, SCE and RIE render negative returns upon applying to the MDP strategy. On the contrary, RIE provides returns 150 times more than the initial investment. Thus, MDP outperforms GMVP in this case.



This boxplot demonstrates that a majority of the diversification ratios achieved occupy the same 25-75 percentile range.

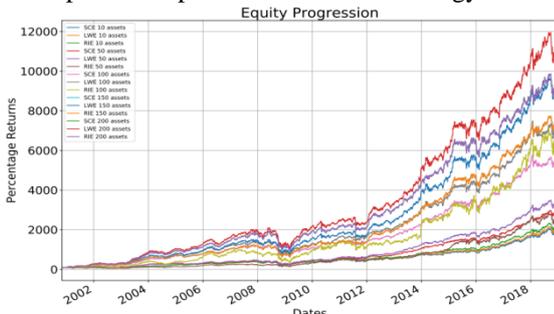
In comparison, a different look-back window brings a change in the ratio as outliers are higher for 12 months. Thus, we conclude that the diversification ratio varies with a change in the number of look-back months in the context of MDP strategy.

**D. Modified Maximum Diversification Portfolio Strategy**

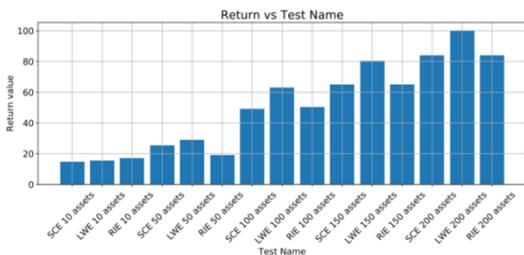
In this research, we propose a new strategy called Modified Maximum Diversification Portfolio strategy (MMDP) to evaluate covariance estimation methods, which is inspired from the MDP strategy. MMDP follows MDP in terms of maximizing the diversification ratio, however, with a relatively lesser risk constraint. The aforementioned can be described as follows:

$$\begin{aligned}
 & \text{maximize } g^T \text{diag}(S) \\
 & g \in \mathbb{R}^N \\
 \text{s. t. } & 1^T \cdot g = 1 \\
 & v \geq 0 \\
 & g^T S g \leq \frac{1}{1+c}, \forall c > -1.
 \end{aligned}
 \tag{15}$$

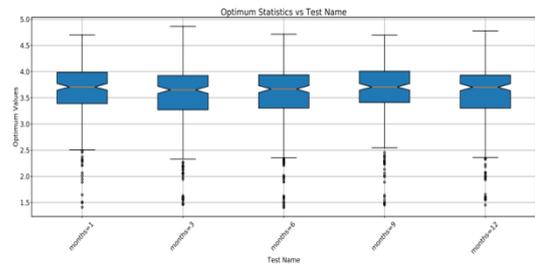
We conducted this experiment by assigning the value 0.50 to the constraint  $c$  which determines the measure of deviation from the optimum  $D(g)$  that a user is willing to undergo. If this strategy cannot find a suitable optima to the above logic, it resorts to an  $g = 1/V$  strategy. The results that follow captures the performance of this strategy:



The above Equity Progression plot demonstrates that the MMDP strategy shows a positive trend in returns for a 6 month look-back period while producing an increasing equity value.



The above graph provides deeper insights into the overall returns for MMDP strategy on a 6 month look-back window. Interestingly, LWE provides the highest returns for a 150 asset portfolio for this strategy. Apart from outperforming MDP in providing higher returns, MMDP also shows positive returns for all covariance matrix estimators.



The above boxplot demonstrates that the highest diversification ratio is achieved at a look-back period of 3 months. It shows results similar to that of the MDP strategy, by and large.

**IV. CONCLUSION AND FUTURE WORK**

The prime objective of our project was to evaluate the performance of various covariance estimators in fulfilling the goals of portfolio optimization strategies. Our evaluation suggests that LWE performs best in most cases and we recommend using the same for covariance matrix estimation. We also proposed a new strategy - MMDP, based on MDP, which produced excellent results in our experiments. Our future work on this involves further evaluation of MMDP and interpreting the returns when it is paired with various covariance matrix estimators.

**REFERENCES**

- O. Ledoit and M. Wolf, The Power of (Non-)Linear Shrinking: A Review and Guide to Covariance Matrix Estimation, Working Paper University of Zurich No. 323, 2019.
- J. Bun, R. Allez, J. Bouchaud and M. Potters, "Rotational Invariant Estimator for General Noisy Matrices," in IEEE Transactions on Information Theory, vol. 62, no. 12, pp. 7475-7490, Dec. 2016, doi: 10.1109/TIT.2016.2616132.
- Bai, Jushan & Shi, Shuzhong. (2011). Estimating High Dimensional Covariance Matrices and its Applications. Annals of Economics and Finance. 12. 199-215.
- Fan, Jianqing & Liao, Yuan & Liu, Han. (2015). An Overview on the Estimation of Large Covariance and Precision Matrices. The Econometrics Journal. 19. 10.1111/ectj.12061.
- Ledoit, Olivier; Wolf, Michael. Optimal estimation of a large-dimensional covariance matrix under Stein's loss. Bernoulli 24 (2018), no. 4B, 3791--3832. doi:10.3150/17-BEJ979.
- J. Bun. An Optimal Rotational Invariant Estimator for General Covariance Matrices: the outliers. Preprint, 2018.
- Tokpavi, Sessi & Maillet, Bertrand & Vaucher, Benoit. (2015). Global Minimum Variance Portfolio Optimisation Under some Model Risk: A Robust Regression-based Approach. European Journal of Operational Research. 244. 289-299. 10.1016/j.ejor.2015.01.010.

**AUTHOR PROFILE**



**Ruchir Sharma**, I am pursuing my undergraduate studies in Computer Science Engineering with major in Mathematics and Computing at Thapar Institute of Engineering and Technology, Patiala. I have been actively pursuing research in the field of Financial Mathematics, Portfolio Theory, Portfolio Optimization and other mathematical techniques. I am also interested in Mathematical Finance, Matrix Computation, Numerical Techniques, Stochastic Techniques and other optimization Techniques. With a knack for innovation in fields of Finance and Economics, I am also pursuing several projects pertaining to Financial Markets, Indian Economy and various other topics. Education Pursuing Bachelor of Engineering in Thapar Institute of Engineering and Technology, Patiala (graduating in year 2021)

