

Cleaning Correlation Matrices for Portfolio Optimization and its Applications



Guangyi (Mark) Chen

Supervised by Prof. Jim Gatheral

Baruch MFE

A dissertation submitted in partial fulfilment of the requirements for the degree of

Master of Science in Financial Engineering

December 2020

Abstract

Traditionally in portfolio optimization, we usually used the uncleaned correlation matrices to calculate the weight of numerous assets, such as the famous Markowitz' optimal portfolio theory. In this project, we explain and show that there are 3 cleaning recipes including widely-used linear shrinkage, recently-developed random matrix theory and rotationally invariant shrinkage. Then, we apply these recipes to stocks data, bond data and world indices data to check the performance and also conduct the backtest analysis to show the historical performance of the portfolio optimization with the cleaning matrices. Among them, we find that the rotationally invariant optimal shrinkage is more stable and always show better result.

Keywords: *portfolio optimization, correlation matrices, random matrix theory, stocks, bond, backtest*

Acknowledgements

I would like to express my sincere gratitude to Professor Jim Gatheral, my report's supervisor, for his time, patience and insightful guidance. Also, he taught me two courses, the MTH9879 Market Microstructure and MTH9875 Volatility Surface and I really enjoy his courses and learned a lot!

Besides, thanks so much to Prof. Stefanica and Mrs. Miriam for the beautiful Baruch MFE family they built.

Contents

1	Introduction	1
1.1	History of Portfolio Optimization	1
1.2	Overview of Objectives and Structure	2
2	Three Cleaning Recipes	3
2.1	Basic Linear Shrinkage	4
2.2	Eigenvalues Clipping (Random Matrix Theory)	4
2.3	Rotationally Invariant, Optimal Shrinkage	5
3	Application to Datasets	6
3.1	Introduction of Datasets	6
3.2	Measurement	8
3.3	Results	9
4	Backtest Analysis	10
4.1	Markowitz Optimization and Backtest Scheme	10
4.2	Backtest Performance	11
5	Conclusion	12
	References	13

Chapter 1

Introduction

1.1 History of Portfolio Optimization

Portfolio Optimization consists of determining a set of assets, and their respective portfolio participation weights, which satisfy the investor concerning the combination of risk-return binomial. Given this problem, Markowitz (1952) proposed the Mean-Variance (MV) model. In this model, the expected return is given by the average of the historical data of the stock's return, and the risk is calculated by the variance of these returns. The main idea of the Mean-Variance Model is to deal with the returns of individual assets as random variables and use the expected return and variance to quantify the return and risk, respectively. Specifically, the problem is as follows:

Suppose there are $n \geq 2$ assets in a portfolio. Let $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$ be the mean vector and $\Sigma = (\text{Cov}\{R_i, R_j\})$ the variance-covariance matrix of the asset returns $R_i, i = 1, \dots, n$. The portfolio return

$$R = \sum_{i=1}^n w_i R_i = (w_1, \dots, w_n) \begin{bmatrix} R_1 \\ \vdots \\ R_n \end{bmatrix} = \boldsymbol{w}' \boldsymbol{R}$$

has mean and variance

$$\mu = \boldsymbol{w}' \boldsymbol{\mu}, \sigma^2 = \boldsymbol{w}' \Sigma \boldsymbol{w}$$

respectively, where $\boldsymbol{w} = (w_1, \dots, w_n)'$ is the vector of weights.

For this problem, in practice our target is usually to minimize the variance σ^2 given the expected return μ or vice versa, otherwise we can regard the problem as maximizing the Sharpe Ratio that is the ratio of the expected return and the standard deviation given some constraints.

In 1971, Merton (1971) approaches the portfolio optimization problem from the perspective of how to systematically build and analyze optimal dynamic models of continuous-time under uncertainty. According to the author, the main advantage of the continuous-time model comparing to models that consider the discrete-time is the fact that it contemplates only two types of stochastic processes: functions of Brownian movements and Poisson processes.

Thus, the number of parameters in the problem is reduced, which allows taking full advantage of the huge amount of written literature about these processes.

Uryasev (2000) proposed the Conditional-Value-at-Risk (CVaR) method, which is used as a risk measure of the portfolio, and therefore optimization would be on top of that metric. The CVaR is derived from the weighted average of “extreme” losses in the tail of the distribution of possible returns. One of the advantages of CVaR is that it can be minimized efficiently using LP (Linear Programming) techniques.

Another approach to the portfolio optimization problem is robust optimization that is proposed by Fabozzi et al. (2007), which is the act of incorporating into the model, the uncertainty generated by the errors of estimation of the parameters used in optimization. Also, robust optimization offers as an advantage the fact that modifications in the mathematical model do not change the characteristics of the problem, which remains a problem of quadratic programming.

Furthermore, DeMiguel et al. (2009), Pflug et al. (2012) and Behr et al. (2013) research on the uniform investment strategy (Naive diversification or $1/N$) is rational the greater the degree of uncertainty about the risk or return distribution is, and this fact is supported by several experimental studies which tell about that this strategy tends to be better in the environment of extreme uncertainty, and is usually a good strategy for risk-averse investors.

Bun et al. (2016) mainly focus on the correlation matrices that is used by portfolio optimization problem, especially the Markowitz solution. It summarizes five cleaning recipes for the correlation matrices and gives more reliable estimators of covariance or correlation matrices. Our study mainly focus on this paper, to implement those five cleaning recipes and also do something innovative to check the performance of them in the Markowitz portfolio.

1.2 Overview of Objectives and Structure

In this report, we focus on the three main recipes provided in Bun et al. (2016): Basic Linear Shrinkage, Eigenvalues Clipping and Rotationally Invariant Optimal Shrinkage. The reason that we focus on the covariance matrix of Markowitz portfolio theory is that we know that there are two key parts in this method: the expected return and the covariance matrix. In practice, we have seen that there are numerous hedge funds and quantitative researchers dive into the prediction of the expected returns by applying some machine learning models or time series models on the historical data and then forecasting. However, few of them realize the importance the covariance matrix that measures the risk of the portfolio. When we look at the Sharpe Ratio which is the ratio of the expected return and the risk, sometimes although someone gets good prediction of the future returns, their optimization of portfolio is bad. This phenomenon always occurs when they used the bad covariance matrix. In my view, therefore, researching and implementing good estimators of the covariance matrix is quite essential for the portfolio optimization problem.

The structure of this report is as follows: In Chapter 2, we introduce the three popular recipes in detail; In Chapter 3, we apply these three methods to three different financial datasets: S&P500, bond return and 22 world indices; In Chapter 3, we do something innovative: we implement the three cleaning recipes into the Markowitz portfolio with the same expected returns and do some backtest analysis to check the performance.

Chapter 2

Three Cleaning Recipes

We first set the stage. Assume that there are N different financial assets that we observe at the daily frequency, defining a vector of returns $\mathbf{r}_t = (r_{1t}, r_{2t}, \dots, r_{Nt})$ for each day $t = 1, \dots, T$. First of all, we need to standardize these returns as follows: (i) remove the sample mean of each asset; (ii) normalize each return by an estimate $\hat{\sigma}_{it}$ of its daily volatility: $\tilde{r}_{it} := r_{it}/\hat{\sigma}_{it}$. There are many choices of the estimators of $\hat{\sigma}_{it}$, and we try two of them: (i) the cross-sectional daily volatility, i.e. $\hat{\sigma}_{it} = \sqrt{\sum_{j=1}^N r_{jt}^2}$; (ii) the Parkinson's

Volatility, i.e. $\hat{\sigma}_{it} = \sqrt{\frac{\sum_{t=1}^n \frac{1}{4 \ln 2} * (x_t^{HL})^2}{n}}$, where $x_t^{HL} = \ln \frac{S_{Ht}}{S_{Lt}}$, S_{Ht} is high stock's price in t day. S_{Lt} is low stock's price in t day. So we can remove a substantial amount of non-stationarity in the volatilities. The final standardized return matrix $\mathbf{X} = (X_{it}) \in \mathbb{R}^{N \times T}$ is then given by $X_{it} := \tilde{r}_{it}/\sigma_i$ where σ_i is the sample estimator of the \tilde{r}_i which is now, to a first approximation, stationary.

The most common estimator of the "true" underlying correlation matrix (that we denote as \mathbf{C} henceforth) is to use the sample estimator

$$\mathbf{E} := \frac{1}{T} \mathbf{X} \mathbf{X}^T$$

and the equivalent notation:

$$\mathbf{E} = \sum_{k=1}^N \lambda_k \mathbf{u}_k \mathbf{u}_k^T$$

for the eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \lambda_N \geq 0$ and the associated eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \dots \mathbf{u}_N$ of \mathbf{E} . When $q = N/T \rightarrow 0$, i.e. when the data set is very long, one expects that $\mathbf{E} \rightarrow \mathbf{C}$. In fact, since the work of Marchenko and Pastur (1967) decades ago, one can show that the spectrum of \mathbf{E} is a broadened version of that of \mathbf{C} , with an explicit q -dependent formula relating the two. In particular, small λ_i 's are too small and large λ_i 's are too large compared to the true eigenvalues \mathbf{C} . Thus, recalling that Markowitz's optimal strategy overweight low variance modes (see Bouchaud and Potters (2009)), we understand why using the sample estimator \mathbf{E} can lead to disastrous result, and why some cleaning procedure should be applied to \mathbf{E} .

2.1 Basic Linear Shrinkage

Linear combination between the sample estimate and the identity matrix

$$\Xi^{\text{bas.}} := \alpha \mathbf{E} + (1 - \alpha) I_N$$

This is probably the oldest method proposed in the literature Haff (1980). In a financial context, this method can be seen as a heuristic way to control the diversification of the optimal Markowitz portfolio.

Note that there is a hyper parameter α we need to determine. In this report, we use greedy search to optimize α using the historical data. And also we update α in a rolling base so each time we need to re-calculate the estimator of the correlation matrix.

2.2 Eigenvalues Clipping (Random Matrix Theory)

As shown in Gatheral (2008), we firstly introduce the random correlation matrices, that is:

Suppose we have M stock return series with T elements each. The elements of the $M \times M$ empirical correlation matrix E are given by

$$E_{ij} = \frac{1}{T} \sum_t x_{it} x_{jt}$$

where x_{it} denotes the t th return of stock i , normalized by standard deviation so that $\text{Var}[x_{it}] = 1$. In matrix form, this may be written as

$$\mathbf{E} = \mathbf{H}\mathbf{H}^T$$

where \mathbf{H} is the $M \times T$ matrix.

Then eigenvalues spectrum of the random correlation matrix is as follows:

Suppose the entries of \mathbf{H} are random with variance σ^2 . Then, in the limit $T, M \rightarrow \infty$ keeping the ratio $Q := T/M \geq 1$ constant, the density of eigenvalues of \mathbf{E} is given by

$$\rho(\lambda) = \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}$$

where $\rho(\lambda)$ is the Marčenko-Pastur density, the maximum/minimum eigenvalues are:

$$\lambda_{\pm} = \sigma \left(1 \pm \sqrt{\frac{1}{Q}} \right)^2$$

Finally, the result recipe is as follows:

- Fit the Marčenko-Pastur distribution to the empirical density to determine Q and σ .
- All eigenvalues above some number λ^* are considered informative; otherwise eigenvalues relate to noise.
- Replace all noise-related eigenvalues λ_i below λ^* with a constant and renormalize so that $\sum_{i=1}^M \lambda_i = M$

- Undo the diagonalization of the sample correlation matrix \mathbf{C} to obtain the denoised estimate \mathbf{C}' .

In short, we keep the $\lceil N\alpha \rceil$ top eigenvalues and shrink the others to a constant γ that preserves the trace, $\text{Tr}(\Xi^{\text{clip.}}) = \text{Tr}(\mathbf{E}) = N$:

$$\Xi^{\text{clip.}} := \sum_{k=1}^N \xi_k^{\text{clip.}} \mathbf{u}_k \mathbf{u}_k^T, \quad \xi_k^{\text{clip.}} := \begin{cases} \lambda_k & \text{if } k \leq \lceil N\alpha \rceil \\ \gamma & \text{otherwise} \end{cases}.$$

In this report, we choose α as the the Marčenko-Pastur edge $(1 + \sqrt{q})^2$. However, this cleaning overlooks the fact that the large empirical eigenvalues are over-estimated.

2.3 Rotationally Invariant, Optimal Shrinkage

This method was first proposed by Ledoit and P  ch   (2011) who realized that one can actually compute the "overlap" between true and sample eigenvectors, precisely the ingredient missing from Eigenvalues Clipping method. Then it was extended to discrete eigenvalues in Bun and Knowles (2016) which also provided a practical implementation. Finally, the method was further studied in Bun et al. (2017) where, in particular, an ad-hoc method was proposed to correct for a systematic bias for small eigenvalues. The algorithm is as follows:

Given the eigenvalues $[\lambda_k]_{k=1}^N$ of the sample correlation matrix \mathbf{E} and $q = N/T$, we define the complex variable $z_k = \lambda_k - i/\sqrt{N}$ for any $k \in [1, N]$. Then, the cleaning scheme reads:

- Compute for any $k \in [1, N]$

$$\xi_k^{\text{rie}} = \frac{\lambda_k}{|1 - q + qz_k s_k(z_k)|^2}$$

with

$$s_k(z_k) = \frac{1}{N} \sum_{\substack{j=1 \\ j \neq k}}^N \frac{1}{z_k - \lambda_j}$$

- Evaluate at the same time for any $k \in [1, N]$

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

where $g_{mp}(z)$ is the Stieltjes transform of the (rescaled) Marčenko-Pastur density given by

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

with

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

where λ_N is the smallest empirical eigenvalue.

- The debiased optimal RIE for any $k \in [1, N]$ reads

$$\hat{\xi}_k = \begin{cases} \Gamma_k \xi_k^{\text{rie}} & \text{if } \Gamma_k > 1 \\ \xi_k^{\text{rie}} & \text{otherwise} \end{cases}$$

Chapter 3

Application to Datasets

3.1 Introduction of Datasets

In this study, we implement three recipes that are introduced in Chapter 2 to three financial datasets: S&P500 equities, Bond and world indices to check and compare the performance of each recipe.

- Equity: S&P500 equities

Source: WRDS(CRSP) daily stock file

Data range: 10 Years, 1999-12-31 ~ 2019-12-31

We multiply price with a accumulated factor to get the adjusted price and also deal with some data issues such as missing values, outliers, etc. The snapshot of the data after processing is as follows:

	PERMNO	date	TICKER	BIDLO	ASKHI	PRC	CFACPR	ADJPRC
0	10104	1999-12-31	ORCL	110.8125	113.3750	112.0625	4.0	448.25
1	10104	2000-01-03	ORCL	111.6250	125.1875	118.1250	4.0	472.50
2	10104	2000-01-04	ORCL	105.0000	118.6250	107.6875	4.0	430.75
3	10104	2000-01-05	ORCL	96.0000	106.3750	102.0000	4.0	408.00
4	10104	2000-01-06	ORCL	94.6875	105.0000	96.0000	4.0	384.00
...
2449808	81774	2019-12-24	FCX	12.9400	13.1400	12.9900	1.0	12.99
2449809	81774	2019-12-26	FCX	13.0200	13.1800	13.1700	1.0	13.17
2449810	81774	2019-12-27	FCX	13.0100	13.2500	13.0300	1.0	13.03
2449811	81774	2019-12-30	FCX	12.9900	13.1400	13.0100	1.0	13.01
2449812	81774	2019-12-31	FCX	12.9200	13.1500	13.1200	1.0	13.12

Figure 3.1: Snapshot of S&P500 equities dataset.

- Bond: 13 weeks, 5Y, 10Y, 30Y government bond

Source: Yahoo Finance

Data range: 10 Years, 1999-12-31 ~ 2019-12-31

	Date	Open	High	Low	Close	Adj Close	Volume	Treasury	bond_ret
0	2000-01-03	6.411	6.473	6.411	6.457	6.457	0.0	FVX	0.018133
1	2000-01-04	6.434	6.450	6.388	6.396	6.396	0.0	FVX	-0.009447
2	2000-01-05	6.427	6.489	6.396	6.489	6.489	0.0	FVX	0.014540
3	2000-01-06	6.443	6.473	6.427	6.450	6.450	0.0	FVX	-0.006010
4	2000-01-07	6.443	6.490	6.397	6.397	6.397	0.0	FVX	-0.008217
...
19971	2019-12-20	2.362	2.371	2.344	2.346	2.346	0.0	TYX	0.000853
19972	2019-12-23	2.335	2.368	2.321	2.363	2.363	0.0	TYX	0.007246
19973	2019-12-26	2.344	2.349	2.327	2.337	2.337	0.0	TYX	-0.011003
19974	2019-12-27	2.323	2.323	2.305	2.311	2.311	0.0	TYX	-0.011125
19975	2019-12-30	2.373	2.390	2.342	2.343	2.343	0.0	TYX	0.013847

Figure 3.2: Snapshot of Bond returns dataset.

- World Index, 22 in total, including: Dow 30, Nasdaq, FTSE 2000, etc

Source: Yahoo Finance

Data range: 10 Years, 1999-12-31 ~ 2019-12-31

	Date	Open	High	Low	Close	Adj Close	Volume	Indices
0	1999-12-31	3369.610107	3369.610107	3369.610107	3369.610107	3369.572510	0.0	99001.SZ
1	2000-01-03	3369.610107	3369.610107	3369.610107	3369.610107	3369.572510	0.0	99001.SZ
2	2000-01-04	3374.110107	3512.300049	3360.209961	3497.060059	3497.020996	0.0	99001.SZ
3	2000-01-05	3500.129883	3589.179932	3468.689941	3486.290039	3486.250977	0.0	99001.SZ
4	2000-01-06	3475.459961	3663.219971	3454.350098	3655.199951	3655.158936	0.0	99001.SZ
...
110119	2019-12-23	2504.129883	2538.219971	2501.919922	2536.770020	2536.770020	0.0	XAX
110120	2019-12-24	2534.179932	2542.639893	2528.560059	2541.679932	2541.679932	0.0	XAX
110121	2019-12-26	2547.620117	2565.500000	2541.679932	2552.500000	2552.500000	0.0	XAX
110122	2019-12-27	2556.090088	2556.090088	2533.110107	2533.860107	2533.860107	0.0	XAX
110123	2019-12-30	2532.030029	2544.879883	2530.330078	2543.919922	2543.919922	0.0	XAX

Figure 3.3: Snapshot of World Indices dataset.

3.2 Measurement

We divide the total length of out time series T_{tot} into n consecutive, non-overlapping samples of length T_{out} . The "training" period has length T , so n is given by:

$$n := \left\lfloor \frac{T_{tot} - T - 1}{T_{out}} \right\rfloor$$

In this study, we choose $T = 1000$ as the training period data length for each rolling window and $T_{out} = 60$ for the out-of-sample test period.

The metrics we use to measure the performance of each recipe (the "true" risk associated to a portfolio) is the oracle estimator, which is computed as:

$$\xi_i^{ora.} \approx \frac{1}{n} \sum_{j=0}^{n-1} \mathcal{R}^2(t_j, \mathbf{u}_i) \quad i = 1, \dots, N$$

for $t_j = T + j \times T_{out} + 1$ and $\mathcal{R}^2(t, \mathbf{w})$ denotes the out-of-sample variance of the returns of portfolio \mathbf{w} built at time t , that is to say

$$\mathcal{R}^2(t, \mathbf{w}) := \frac{1}{T_{out}} \sum_{\tau=t+1}^{t+T_{out}} \left(\sum_{i=1}^N \mathbf{w}_i X_{i\tau} \right)^2$$

where $X_{i\tau}$ denotes normalized return matrix that we derived after normalizing the returns. This implies that $\sum_{i=1}^N \mathcal{R}^2(t, \mathbf{u}_i) = N$ for any time t .

For the weight of the portfolio, we constructed as follows:

$$\mathbf{w} := \frac{\hat{\Sigma}^{-1} \mathbf{g}}{\mathbf{g}^T \hat{\Sigma}^{-1} \mathbf{g}}$$

where \mathbf{g} is a vector of predictions and $\hat{\Sigma}$ is the cleaned covariance matrix $\hat{\Sigma}_{ij} := \sigma_i \sigma_j \hat{\Xi}_{ij}$ for any $i, j \in [1, N]$.

In order to ascertain the robustness of our results in different market situations, we consider the following four families of predictors \mathbf{g} :

- (i) The minimum variance portfolio, corresponding to $g_i = 1, \forall i \in [1, N]$
- (ii) The omniscient case, i.e. when we know exactly the realized returns on the next out-ofsample period for each stock. This is given by $g_i = \mathcal{N} \tilde{r}_{i,t}(T_{out})$ where $r_{i,t}(\tau) = (P_{i,t+\tau} - P_{i,t}) / P_{i,t}$ with $P_{i,t}$ the price of the i th asset at time t and $\tilde{r}_{it} = r_{it} / \hat{\sigma}_{it}$
- (iii) Mean-reversion on the return of the last day: $g_i = -\mathcal{N} \tilde{r}_{it} \forall i \in [1, N]$
- (iv) Random long-short predictors where $\mathbf{g} = \mathcal{N} \mathbf{v}$ where \mathbf{v} is a random vector uniformly distributed on the unit sphere.

where $\mathcal{N} = \sqrt{N}$.

Also, we include the origin recipes that take the identity matrix (total shrinkage, $\alpha = 0$ in Basic Linear Shrinkage recipe) and the uncleaned, in-sample correlation matrix ($\alpha = 1$) as control groups.

3.3 Results

The results for the three recipes on different dataset are shown in Table 3.1. Note that here, we transform the oracle estimator described above to the annualized average volatility and the standard deviations are given in bracket in the table.

Recipe	Equity	Bond	World Index
Total Shrinkage	0.2516 (0.055)	4.2813 (0.814)	1.6768 (0.215)
Uncleaned Matrix	0.1657 (0.032)	3.8339 (0.926)	1.5362 (0.173)
Basic Linear Shrinkage	0.1448 (0.029)	3.8772 (0.837)	1.5419 (0.185)
Eigenvalues Clipping	0.1410 (0.030)	/	1.5883 (0.187)
debiased RIE	0.1405 (0.029)	3.8309 (0.921)	1.5341 (0.174)

Table 3.1: Annualized average volatility of different recipes for different dataset. Standard deviations are given in bracket.

From Table 3.1, firstly we can see that it is obvious that for all of the three datasets, the debiased RIE recipe always shows the best result, i.e. the lowest annualized average volatility. Secondly, from the equity data, the three recipes we introduced above are all better than the original methods (i.e. the Total Shrinkage and Uncleaned Matrix methods), which coincides with the original paper by Bun et al. (2016). But the performance of Eigenvalues Clipping is not good for both Bond and Word Index datasets, which may be due to the number of assets for Bond and World Index are so small (4 and 22, respectively) that the number of eigenvalues of the correlation matrices for these two datasets are few. Based on the description of this method that keeps the large eigenvalues above a threshold unchanged and normalizes the others, we can see it performs poorly when the number of eigenvalues is small. Even so, from the result, we can see the robustness of debiased RIE recipe.

Recipe	Minimum Varaince	Omniscient	Mean-Reversion	Uniform
Total Shrinkage	0.2516 (0.055)	0.1557 (0.042)	0.1528 (0.082)	0.0464 (0.008)
Uncleaned Matrix	0.1657 (0.032)	0.1211 (0.027)	0.0587 (0.018)	0.0480 (0.007)
Basic Linear Shrinkage	0.1448 (0.029)	0.1109 (0.023)	0.0510 (0.019)	0.0370 (0.004)
Eigenvalues Clipping	0.1410 (0.030)	0.1078 (0.021)	0.0512(0.020)	0.0381 (0.005)
debiased RIE	0.1405 (0.029)	0.1090 (0.021)	0.0495 (0.018)	0.0370 (0.005)

Table 3.2: Annualized average volatility of different recipes for different predictors using equity dataset. Standard deviations are given in bracket.

We have seen that equity dataset is better comparing to the other two, so from now on, we will mainly use the equity dataset to do further research. Next, we test the performance of different recipes on different predictors and the result is shown in Table 3.2. It is clear that for all predictors, performances of our three recipes are better than the original methods. And again, debiased RIE recipe performs best for Minimum Variance, Mean-Reversion and Uniform predictor, except for Omniscient predictor (Eigenvalues Clipping in this case), which justifies the robustness of debiased RIE method.

Chapter 4

Backtest Analysis

4.1 Markowitz Optimization and Backtest Scheme

Then we aim to implement these three recipes on Markowitz portfolio and perform the backtest analysis. To get the new equation for the weight based on Markowitz Portfolio, let's recall the Markowitz Optimization problem:

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

$$\text{with constraints } \mathbf{w}^T \boldsymbol{\mu} = \mu^* \text{ and } \mathbf{w}^T \mathbf{1} = 1$$

where $\mathbf{1} = (1, \dots, 1)^T$ is a column vector of 1's.

Using Lagrange multipliers, it is equivalent to

$$\mathbf{w}^T \Sigma \mathbf{w} - 2\lambda (\mathbf{w}^T \boldsymbol{\mu} - \mu^*) - 2\gamma (\mathbf{w}^T \mathbf{1} - 1)$$

with respect to \mathbf{w} , λ and γ .

Then we can get the new weight:

$$\boldsymbol{\omega} = \frac{1}{D} (B\Sigma^{-1}\mathbf{1} - A\Sigma^{-1}\boldsymbol{\mu} + \mu^* (C\Sigma^{-1}\boldsymbol{\mu} - A\Sigma^{-1}\mathbf{1}))$$

where

$$A = \boldsymbol{\mu}^T \Sigma^{-1} \mathbf{1}, \quad B = \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu}, \quad C = \mathbf{1}^T \Sigma^{-1} \mathbf{1}, \quad D = BC - A^2$$

So next we utilize this new weight to allocate equities of S&P500 in a daily base and multiply with the true daily return of each stock in order to perform the backtest analysis to check the performance from 2004 to 2020. In this report, we use the weighted moving average of historical daily return as the expected return for the above formula.

4.2 Backtest Performance



Figure 4.1: Backtest performance for each recipe implementing in Markowitz portfolio.

	Uncleaned Matrix	Basic Linear Shrinkage	Eigenvalues Clipping	debiased RIE
Sharpe Ratio	0.20	0.43	0.50	0.53

Table 4.1: Sharpe ratio for each recipe in the backtest.

Figure 4.1 shows the backtest performance for the three recipes plus the Uncleaned Matrix, that is the traditional Markowitz Optimization method. From the result, it is clear that the three recipes gives much higher accumulated returns than the traditional one and they have smaller drawdown, which shows the effectiveness of the three cleaning recipes.

Table 2 gives the Sharpe Ratio results, where we can get the rank of these three methods, that is debiased RIE > Eigenvalues Clipping > Basic Linear Shrinkage. So we may use these recipes in portfolio optimization problems in real financial market, especially the debiased RIE method due to its effectiveness and robustness.

Chapter 5

Conclusion

In this report, we study the three recipes introduced in Bun et al. (2016): Basic Linear Shrinkage, Eigenvalues Clipping and debiased RIE. And then implement them on three different datasets: Equity, Bond and World Index with various predictors. The results show that the three recipes are always better than the traditional Markowitz Optimization and the debiased RIE is always the most effective and robust one. In addition, we apply these three methods for Markowitz portfolio with the weighted moving average daily return as the expected return to perform a backtest analysis. Again, debiased RIE comes as the top one so we can see its great robustness.

For further research, firstly rather than using weighted moving average returns as expected return, we can use some machine learning models to predict future return as some hedge funds do, which may give us higher Sharpe Ratio. Secondly, to have better performance on bond and world index data, we need ensure that the number of assets are large enough, especially for the Eigenvalues Clipping method. Finally, comparing to the other two methods, debiased RIE is slower according to the computation speed, so we need to improve it for efficiency, such as using parallel computing.

References

- Behr, P., Guettler, A., and Miebs, F. (2013). On portfolio optimization: Imposing the right constraints. *Journal of Banking & Finance*, 37(4):1232–1242.
- Bouchaud, J.-P. and Potters, M. (2009). Financial applications of random matrix theory: a short review. *arXiv preprint arXiv:0910.1205*.
- Bun, J., Bouchaud, J.-P., and Potters, M. (2016). My beautiful laundrette: Cleaning correlation matrices for portfolio optimization.
- Bun, J., Bouchaud, J.-P., and Potters, M. (2017). Cleaning large correlation matrices: tools from random matrix theory. *Physics Reports*, 666:1–109.
- Bun, J. and Knowles, A. (2016). An optimal rotational invariant estimator for general covariance matrices.
- DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the $1/n$ portfolio strategy? *The review of Financial studies*, 22(5):1915–1953.
- Fabozzi, F. J., Kolm, P. N., Pachamanova, D. A., and Focardi, S. M. (2007). *Robust portfolio optimization and management*. John Wiley & Sons.
- Gatheral, J. (2008). Random matrix theory and covariance estimation. In *NYU Courant Institute Algorithmic Trading Conference*.
- Haff, L. (1980). Empirical bayes estimation of the multivariate normal covariance matrix. *The Annals of Statistics*, pages 586–597.
- Ledoit, O. and Péché, S. (2011). Eigenvectors of some large sample covariance matrix ensembles. *Probability Theory and Related Fields*, 151(1-2):233–264.
- Marchenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik*, 114(4):507–536.
- Markowitz, H. (1952). Portfolio analysis. *Journal of Finance*, 8:77–91.
- Merton, R. (1971). Optimum consumption and portfolio-rules in a continuous-time framework. *Journal of Economic Theory* (December 1971).
- Pflug, G. C., Pichler, A., and Wozabal, D. (2012). The $1/n$ investment strategy is optimal under high model ambiguity. *Journal of Banking & Finance*, 36(2):410–417.
- Uryasev, S. (2000). Conditional value-at-risk: Optimization algorithms and applications. In *Proceedings of the IEEE/IAFE/INFORMS 2000 Conference on Computational Intelligence for Financial Engineering (CIFEr)*(Cat. No. 00TH8520), pages 49–57. IEEE.