Application of Shrinkage in Portfolio Optimization

Fa ZHANG
Department of Mathematics
HKUST
fzhangat@connect.ust.hk

Ruizhe XIA
Department of Mathematics
HKUST
rxiaac@connect.ust.hk

Abstract

Mean-variance optimization portfolio has a poor performance due to the noisy covariance matrix estimation. Many shrinkage approaches have been applied to covariance matrix estimation and precision matrix. We test several shrinkage approaches on S&P 500 and compare with some benchmarks. We show the shrinkage of precision matrix does mitigate the noisy and unstable estimation and lead to a better out-of-sample portfolio performance.

1 Introduction

Portfolio optimization is selecting a portfolio of assets that maximizes the investor's utility. Assuming the returns of the assets follow an independent and identically distributed (i.i.d) multivariate Gaussian distribution, diversification, rather than putting all the money in a few assets, will diversify the risk and lead to a larger Sharpe ratio. The most well-known framework is the mean-variance optimization proposed by Markowitz [1]. It assumes the investors are risk averse, which means they will trade off between the expected return and risk.

Without considering some realistic constraints, it generally has three equivalent forms:

$$\begin{aligned} &\arg\max_{w} E\left(w'r\right) = w'\mu \\ &\text{s.t. } \operatorname{Var}\left(w'r\right) = w'\Sigma w \leq \sigma^2 \end{aligned} \tag{1}$$

The first form can be interpreted as maximizing the expected return for a given level of risk σ.

$$\begin{aligned} &\operatorname*{argmin}_{w} w'\Sigma w \\ &\text{s.t. } w'\mu \geq r^* := \sigma\sqrt{\theta} \end{aligned} \tag{2}$$

Where $\theta = \mu'\Sigma^{-1}\mu$ is the square of the maximum Sharpe ratio of the optimal portfolio. The second form is minimizing the risk for a given level of expected return.

$$\operatorname*{argmin}_{w} w'\Sigma w - \frac{2\sigma}{\sqrt{\theta}} w'\mu \tag{3}$$

While the last form is a tradeoff between expected return and risk.

The three optimization problems share an optimal solution $\boldsymbol{w}^* = \frac{\sigma}{\sqrt{\theta}}\Sigma^{-1}\mu$, a function of the expected returns and the inverse covariance matrix. The true values are unknown in practice and are usually estimated using historical data.

The traditional estimation is the Maximum Likelihood (ML) estimator:

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} R_t, \hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}\left(R_t - \hat{\mu}\right)\left(R_t - \hat{\mu}\right)', \widehat{w}^{\text{mv}} = \frac{1}{\gamma}\hat{\Sigma}^{-1}\hat{\mu} \tag{4}$$

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

Of course, $\hat{\mu}$ and $\hat{\Sigma}$ are consistent estimators of μ and Σ. And if fixing the asset number N, let $T \to \infty$, we have $\widehat{w}^{\mathrm{mv}} \to w^*$.

How the mean-variance portfolio has very poor out-of-sample performance in practice. A striking phenomenon is shown by DeMiguel et al [2], the simple 1/N portfolio (equal weights on all assets) outperforms the mean–variance portfolio and most of its extensions. The mean-variance framework with a solid theoretical foundation can't beat the naive approach.

There are many potential reasons, one important reason is the high-dimensional nature of the optimization problem [3][4][5]. In practice, we usually have hundreds or thousands of assets, due to the regime switch of the market, we usually use the past several years' data to perform estimation, which only contains no more than 3000 observations for daily return data and much less for monthly return data. In other words, the asset number N is comparable to the observations T. However, for the covariance matrix, we have $\frac{N(N+1)}{2}$ entries to estimate, which is far beyond observations T. Then the estimation error will lead to performance deterioration from in-sample to out-of-sample.

2 Background and Related Work

Many statistical techniques have been proposed to reduce estimation errors and improve the out-of-sample portfolio performance [6]. One efficient method is the shrinkage approach [7], inspired by the JS estimator [8]. Shrinkage is applied to the Maximum Likelihood (ML) estimator by optimally exploiting the trade-off between bias and variance:

$$\widehat{\Sigma}^{\text{Shrink}} = (1-\rho)\hat{\Sigma} + \rho\Lambda \tag{5}$$

We shrink $\widehat{\Sigma}$ towards a given matrix Λ (identity matrix, covariance matrix estimated by one factor, etc). The idea behind shrinkage is very similar to the JS estimator, which is improving the stability and accuracy of the sample covariance matrix while retaining some original information by adding some bias.

Many literatures consider the shrinkage of the covariance matrix, which does lead to better estimation. Then we invert it to obtain the precision matrix, then we obtain the optimal portfolio weights. However, due to the complexity of matrix inversion, the inverse covariance matrix may still be unstable and lead to noisy portfolio weight.

To overcome this difficulty, Kourtis et al [9] suggest directly putting shrinkage on the precision matrix instead of the covariance matrix, since the optimal weights are directly related to the precision matrix:

$$\widehat{S}^{\text{Shrink}} = c_1\hat{\Sigma}^{-1} + c_2\Lambda \tag{6}$$

The coefficients c_1 and c_2 can be obtained by cross-validation in historical data.

Senneret et al [10] derive the closed-form expression of the optimal shrinkage parameter ρ by minimizing the quadratic loss function:

$$\begin{aligned} &\min_{\rho} E\left\{\left\|\hat{\Sigma}^{\text{Shrink}} - \Sigma\right\|_{\mathrm{F}}^{2}\right\} \\ &\text{s.t. } \hat{\Sigma}^{\text{Shrink}} = (1-\rho)\hat{\Sigma} + \rho\Lambda \end{aligned} \tag{7}$$

$$\begin{aligned} &\min_{\rho} E\left\{\left\|\hat{S}^{\text{Shrink}} - \Sigma^{-1}\right\|_{\mathrm{F}}^{2}\right\} \\ &\text{s.t. } \hat{S}^{\text{Shrink}} = (1-\rho)\hat{\Sigma}^{-1} + \rho\Lambda \end{aligned} \tag{8}$$

This is called the Oracle approximating shrinkage (OAS) estimator.

Let:

$$S_n = \frac{1}{n-1}\sum_{t=1}^{n}\left(R_t - \hat{\mu}\right)\left(R_t - \hat{\mu}\right)', P_n = \frac{n-p-2}{n-1} \cdot S_n^{-1} \tag{9}$$

We have the following four shrinkage approaches, two for the covariance matrix, and two for the precision matrix.

2.1 Shrinkage of Covariance Matrix

For $\Lambda = I$:

$$\hat{\Sigma}^{(1)} = \hat{\rho} \cdot \frac{\operatorname{Tr} S_n}{p} \cdot I + (1-\hat{\rho}) \cdot S_n \tag{10}$$

Where:

$$\hat{\rho} = \min\left\{ \frac{\left(1-\frac{2}{p}\right)\operatorname{Tr}\left(S_n^2\right) + \left(\operatorname{Tr} S_n\right)^2}{\left(n-\frac{2}{p}\right)\cdot\left[\operatorname{Tr}\left(S_n^2\right) - \frac{\left(\operatorname{Tr} S_n\right)^2}{p}\right]}, 1 \right\}. \tag{11}$$

For $\Lambda = \operatorname{Diag}(S_n)$:

$$\hat{\Sigma}^{(2)} = \hat{\rho} \cdot \operatorname{Diag} S_n + (1-\hat{\rho}) \cdot S_n \tag{12}$$

Where:

$$\hat{\rho} = \min\left\{ \frac{\operatorname{Tr}\left(S_n^2\right) + \left(\operatorname{Tr} S_n\right)^2 - 2\operatorname{Tr}\left[\left(\operatorname{Diag} S_n\right)^2\right]}{n \cdot \left(\operatorname{Tr}\left(S_n^2\right) - \operatorname{Tr}\left[\left(\operatorname{Diag} S_n\right)^2\right]\right)}, 1 \right\} \tag{13}$$

Let $\hat{\mu}$ be the sample mean, then the optimal weight is

$$\omega = \frac{\hat{\Sigma}^{-1}\hat{\mu}}{\|\hat{\Sigma}^{-1}\hat{\mu}\|_1} \tag{14}$$

2.2 Shrinkage of Precision Matrix

For $\Lambda = I$:

$$\widehat{\Pi}^{(1)} = \hat{\rho} \cdot \frac{\operatorname{Tr} P_n}{p} \cdot I + (1-\hat{\rho}) \cdot P_n \tag{15}$$

Where:

$$\hat{\rho} = \min\left\{ \frac{\frac{n-p-\frac{2}{p}(n-p-2)}{n-p-1}\cdot\operatorname{Tr}\left(P_n^2\right) + \frac{n-p-2-\frac{2}{p}}{n-p-1}\cdot\left(\operatorname{Tr} P_n\right)^2}{\left[\frac{n-p-\frac{2}{p}(n-p-2)}{n-p-1} + n - p - 4\right] \times \left[\operatorname{Tr}\left(P_n^2\right) - \frac{\left(\operatorname{Tr} P_n\right)^2}{p}\right]}, 1 \right\} \tag{16}$$

For $\Lambda = \operatorname{Diag}(P_n)$:

$$\widehat{\Pi}^{(2)} = \hat{\rho} \cdot \operatorname{Diag}(P_n) + (1-\hat{\rho}) \cdot P_n \tag{17}$$

Where:

$$\hat{\rho} = \min\left\{ \frac{2\cdot\operatorname{Tr}\left(\operatorname{Diag}(P_n)^2\right) + \frac{n-p}{n-p-1}\cdot\operatorname{Tr}\left(P_n^2\right) + \frac{n-p-2}{n-p-1}\cdot\left(\operatorname{Tr} P_n\right)^2}{\left(\frac{n-p}{n-p-1} + n - p - 4\right) \times \left[\operatorname{Tr}\left(P_n^2\right) - \operatorname{Tr}\left(\operatorname{Diag}(P_n)^2\right)\right]}, 1 \right\} \tag{18}$$

Let $\hat{\mu}$ be the sample mean, then the optimal weight is

$$\omega = \frac{\hat{\Pi}\hat{\mu}}{\|\hat{\Pi}\hat{\mu}\|_1} \tag{19}$$

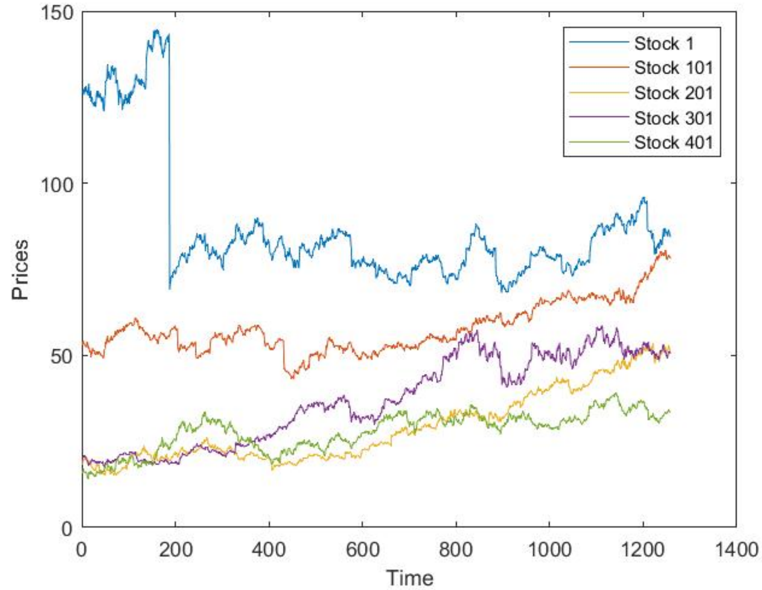
We can see as $n \to \infty$, we have $\rho \to 0$, which means the OAS estimator can adaptively adjust the degree of shrinkage. When n is small, the estimation can be noisy. To stabilize the estimation, more bias is added. When n is large, the sample estimator is already accurate, so we can rely more on it and decrease the degree of shrinkage.

3 Data

We use daily data, which contains closed prices of 452 stocks in SNP'500 for 1258 days in 4 years.

`https://yao-lab.github.io/data/snp452-data.mat`

Figure 1: Data



4 Approaches and Benchmarks

We examine the effect of shrinkage on the covariance matrix (CM1, CM2) and precision matrix (PM1, PM2) towards the identity matrix and a diagonal matrix. We compare the four shrinkage approaches with three benchmarks, equal weight portfolio, mean-variance portfolio, and global minimum variance portfolio. The global minimum variance portfolio is defined as follows:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \mathbf{w}^T\mathbf{\Sigma}\mathbf{w} \\ \text{subject to} & \mathbf{1}^T\mathbf{w} = 1. \end{array} \tag{20}$$

It does not involve estimating expected returns and is found stable and competitive with other portfolios. The closed-form solution of GMVP is $\mathbf{w}_{\mathrm{GMVP}} = \frac{1}{\mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{1}}\mathbf{\Sigma}^{-1}\mathbf{1}$.

After obtaining the optimal weight, we normalize its ℓ_1-norm to be 1 for a fair comparison.

5 Experiments and Results

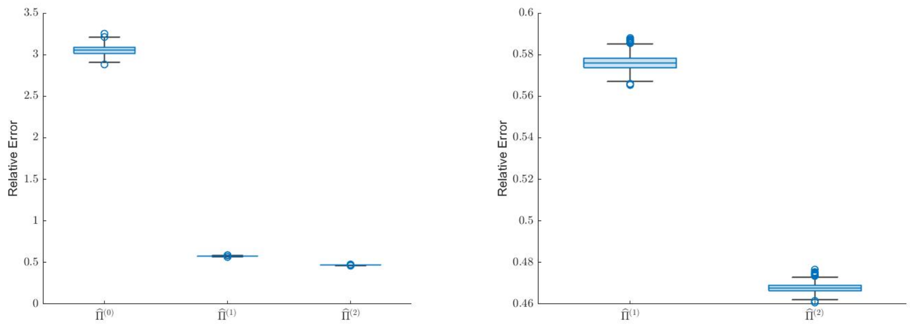
Let $\mathbf{X} \in \mathbb{R}^{m\times p}$ be the price matrix, where $X(t, k)$ denotes the prices of stock k at time t. Then the corresponding return $\mathbf{R} \in \mathbb{R}^{(m-1)\times p}$ reads: $R(t, k) = (X(t+1, k) - X(t, k))/X(t, k)$.

5.1 Synthetic Data

Let $\hat{\Sigma}$ denote the sample mean and sample covariance matrix of $\mathbf{R}$. In each trial, we generate a matrix $\mathbf{Y}_i \in \mathbb{R}^{n\times p}$ of n random vectors chosen from the multivariate normal distribution $\mathcal{N}(0, \hat{\Sigma})$. Let $\hat{\Sigma}_i$ be the sample covariance matrix of $\mathbf{Y}_i$. Then we compute three estimations for the precision matrix: one from the inverse of sample covariance matrix $\widehat{\Pi}_i^{(0)} = \hat{\Sigma}_i^{-1}$ and the other two $\widehat{\Pi}_i^{(1)}, \widehat{\Pi}_i^{(2)}$ from equations (15)-(18). Since the true precision matrix for each $\mathbf{Y}_i$ is $\Pi_i = \hat{\Sigma}^{-1}$. Then we compute the relative error of the estimation of the precision matrix with respect to the Frobenius norm:

$$e_i^{(k)} = \frac{\|\hat{\Sigma}^{-1} - \widehat{\Pi}_i^{(k)}\|_{\mathrm{F}}}{\|\hat{\Sigma}^{-1}\|_{\mathrm{F}}}, \quad k = 0, 1, 2.$$

We set $n = 700$ and conduct 500 random experiments, the results are as follow:



(a) Comparsion of $\widehat{\Pi}_i^{(0)}, \widehat{\Pi}_i^{(1)}, \widehat{\Pi}_i^{(2)}$ (b) Comparsion of $\widehat{\Pi}_i^{(1)}, \widehat{\Pi}_i^{(2)}$

Figure 2: Estimations for the precision matrix

Figure(a) suggests that the inverse of the sample covariance matrix is not a good estimation for the precision matrix. Figure(b) suggests both $\widehat{\Pi}^{(1)}$ and $\widehat{\Pi}^{(2)}$ are biased, and $\widehat{\Pi}^{(2)}$ has less bias and variance than $\widehat{\Pi}^{(1)}$ in the estimation error.

5.2 Finance Data

We estimate the portfolio weights ω_s for each approach s using the returns for the days $1, 2, \cdots, T$. Then for each day $t > T$, we calculate the corresponding portfolio returns $R_t^s = e_t^\top R\omega_s$. We let $\hat{\mu}_s$ and $\hat{\sigma}_s$ denote the sample mean and standard deviation of the time series $\{R_t^s\}_{t=T+1}^{m-1}$ and compute the following out-of-sample performance metrics:

- Annualized Return: $\hat{\mu}_s \cdot 252$
- Annualized Risk: $\hat{\sigma}_s \cdot \sqrt{252}$
- Annualized Sharpe Ratio: $\widehat{SR}_s = \dfrac{\hat{\mu}_s}{\hat{\sigma}_s} \cdot \sqrt{252}$

The out-of-sample performances are summarized in Table 1 and the portfolio returns on the test data are shown in Figure 3.

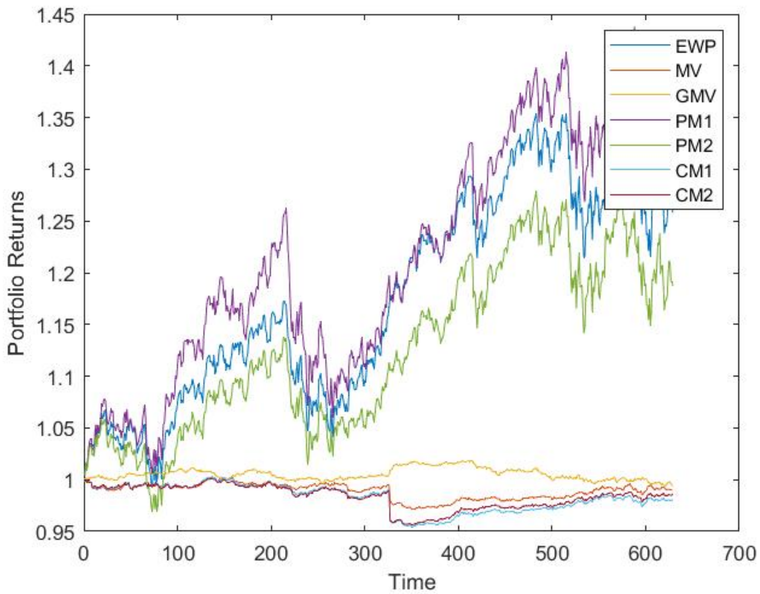
Table 1: Out-of-sample performance

Method	Annualized Return	Annualized Risk	Annualized Sharpe Ratio
EWP	0.1014	0.1369	0.7405
MV	-0.0037	0.0178	-0.2099
GMV	-0.0025	**0.0154**	-0.1619
PM1	**0.1365**	0.1527	**0.8936**
PM2	0.0782	0.1381	0.5665
CM1	-0.0074	0.0245	-0.3042
CM2	-0.0051	0.0239	-0.2147

6 Conclusions

The results coincide with previous findings. The equal-weight portfolio outperforms the mean-variance portfolio. The performances of GMVP, CM1, and CM2 are comparable to the MVP, indicating although shrinkage on the covariance matrix helps reduce the estimation error, its inverse is still very noisy and leads to poor portfolio performance. On the contrary, directly putting shrinkage on the precision matrix leads to a major improvement on the MVP. And PM1 (shrinkage towards identity matrix) outperforms EWP, which suggests shrinkage on the precision matrix is indeed an efficient way to reduce the estimation error and improve the out-of-sample performance of portfolio.

Figure 3: Portfolio returns



Contribution

Fa Zhang: literature review, report writing.

Ruizhe Xia: code implementation, results analysis.

References

[1] Markowitz, H. (1952). Portofolio selection. Journal of finance, 7, 77-91.

[2] DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?. The review of Financial studies, 22(5), 1915-1953.

[3] Jobson, J. D., & Korkie, B. (1980). Estimation for Markowitz efficient portfolios. Journal of the American Statistical Association, 75(371), 544-554.

[4] Michaud, R. O. (1989). The Markowitz optimization enigma: Is 'optimized'optimal?. Financial analysts journal, 45(1), 31-42.

[5] Kan, R., & Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. Journal of Financial and Quantitative Analysis, 42(3), 621-656.

[6] Brandt, M. W. (2010). Portfolio choice problems. In Handbook of financial econometrics: Tools and techniques (pp. 269-336). North-Holland.

[7] Ledoit, O., & Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. Journal of empirical finance, 10(5), 603-621.

[8] James, W., & Stein, C. (1992). Estimation with quadratic loss. Breakthroughs in statistics: Foundations and basic theory, 443-460.

[9] Kourtis, A., Dotsis, G., & Markellos, R. N. (2012). Parameter uncertainty in portfolio selection: Shrinking the inverse covariance matrix. Journal of Banking & Finance, 36(9), 2522-2531.

[10] Senneret, M., Malevergne, Y., Abry, P., Perrin, G., & Jaffres, L. (2016). Covariance versus precision matrix estimation for efficient asset allocation. IEEE Journal of Selected Topics in Signal Processing, 10(6), 982-993.