Portfolio Optimization, Data Estimation, and Shrinkage

Gérard Cornuéjols and Nicolas Naing

September 2020

1   Introduction

Markowitz [5] introduced the following idea for portfolio construction. Given n assets, construct a portfolio of these assets that balances the expected portfolio return at the end of a planning horizon against risk as measured by the variance of the portfolio return. If we know the n-vector μ of the expected asset returns at the end of the planning horizon, and the covariance matrix Σ of the asset returns, the model to solve is max $\mu^T x - \gamma x^T \Sigma x$ subject to $\sum_{i=1}^{n} x_i = 1$, where γ is a penalty for risk. The variables to optimize are the fractions x_i of the portfolio invested in asset i. If short positions are not allowed, we also impose the constraint $x \geq 0$.

In practice, μ and Σ are estimated from data. For example, we might have observed asset returns $\mu^1, \ldots, \mu^T$ from the unknown distribution of asset returns. How should we use these data in the Markowitz model? Should we use the sample average $\bar{\mu} = \frac{1}{T} \sum_{i=1}^{T} \mu^i$ and the sample covariance matrix $\hat{\Sigma}$ in place of the unknown μ and Σ? Practitioners have argued that the above optimization model maximizes the estimation errors. Following the pioneering work of Stein [6] on shrinkage, Jorion [3] showed that better portfolios can be achieved by shrinking $\bar{\mu}$ in the Markowitz model, and Ledoit and Wolf [4] suggested to shrink $\hat{\Sigma}$. See also Ceria and Stubbs [1] and Davarnia and Cornuéjols [2]. In this note we investigate the potential benefits of shrinking the sample covariance matrix $\hat{\Sigma}$.

2   Illustration of the Error Maximization Phenomenon

We illustrate the solution of the Markowitz model on an instance with 3 assets. Their returns come from the distributions $N(3, 8)$, $N(1.8, 1)$, and $N(2, 3)$ respectively. We assume that these asset returns are independent.

Choosing $\gamma = 0.2$ as our risk aversion factor, we find that the optimum portfolio is $x_1 = 0.38$, $x_2 = 0.43$, $x_3 = 0.19$ with an expected objective value $\mu^T x - \gamma x^T \Sigma x = 1.961$.

Now let us assume that we do not know μ and Σ. We observe 100 returns for these 3 assets.

We compute estimators for the 1×3 mean matrix $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}$ and the 3×3

covariance matrix $\Sigma = \begin{bmatrix} \sigma_{11}\sigma_{12}\sigma_{13} \\ \sigma_{21}\sigma_{22}\sigma_{23} \\ \sigma_{31}\sigma_{32}\sigma_{33} \end{bmatrix}$ for these 3 random variables.

Each entry of our estimator for the mean vector μ was computed by averaging the 100 given sample vectors. The estimator of the covariance matrix was computed using $Cov(X, Y) = E(XY) - E(X)E(Y)$ for random variables X and Y, where the expected values on the RHS were computed by averaging all possibilities from the given data (10000 for $E(XY)$, 100 for $E(X)$ and $E(Y)$.) Let $\bar{\mu}$ and $\hat{\Sigma}$ be the resulting estimators.

Using this estimator setup, we construct our portfolio by maximizing $\bar{\mu}^\top x - \gamma x^\top \hat{\Sigma} x$ over our constraint set $\sum_{i=1}^{n} x_i = 1$, $x \geq 0$. A typical solution might look like $x_1 = 0.46$, $x_2 = 0.5$, $x_3 = 0.04$ with an estimated objective value $\bar{\mu}^\top x - \gamma x^\top \hat{\Sigma} x = 2.093$. Note that this value happens to be greater than the optimal value of 1.961 found earlier. Plugging in the actual μ and Σ, we find the actual objective value $\mu^T x - \gamma x^T \Sigma x = 1.928$. Repeating this 20 times, we find that, on average, the estimated objective value is 2.008, whereas the actual objective value is 1.927. Not surprisingly, this value is less than the value 1.961 that we get from the optimal portfolio.

3   Shrinkage

A second question explored how multiplying the off diagonal terms of the Σ estimator by a constant affected the difference between the optimal and adjusted value. The code is similar to the code used for the first experiment, but uses the shrunk matrix $\tilde{\Sigma} = \begin{bmatrix} \hat{\sigma_{11}} & r\hat{\sigma_{12}} & r\hat{\sigma_{13}} \\ r\hat{\sigma_{21}} & \hat{\sigma_{22}} & r\hat{\sigma_{23}} \\ r\hat{\sigma_{31}} & r\hat{\sigma_{32}} & \hat{\sigma_{33}} \end{bmatrix}$ with shrinking factor $0 \leq r \leq 1$.

We tried $r = .5, .6, .7, .8, .9$ for the same 3 random variables used in Section 2. Out of these, .5 led to the best difference between optimal and adjusted.

For a second experiment, we tested shrinkage values r from 0 to 1.00 instead of just .5, .6, .7, .8, and .9. The random variables used were $Unif(0, 3)$, $N(2, 2)$ and $N(3, 3)$, which are different from those in Section 2. We used multiples of .01 as shrinkage factors from 0 to .99. Each adjusted return value associated with a shrinkage factor was the average of 200 data generation trials, where each trial was given 30 data points for each random variable. The construction of estimators for μ and Σ was the same as in the previous method.

In general, adjusted return value decreased with increased shrinkage factor. The number with the highest value was at .04, the range with the highest values was 0 to .10, and the range with the lowest values was .90 to .99. One reason for this trend would be the fact that all 3 random variables are independent,

which would mean that the best estimator of the true covariance matrix would have all 0s in entries outside the main diagonal.

We repeated the experiment with a second set of random variables. The first rv was $Unif(0,3)$, the second was $N(2,2)$, and the third was $Unif(0,2)$ if the first rv was greater than 2 and $Unif(1,3)$ otherwise. Using this setup, the first and third variables are correlated (correlation coefficient was calculated with casework and taking integrals of pdfs).

The best shrink value r for this experiment was .65. Using the same consecutive average variable, the averages rose from 0 to .65, stayed level from .65 to .75, and then decreased from .75 to 1.00. This illustrates the possiblee value of shrinkage in constructing better portfolios, even when the asset returns are correlated.

References

[1] Ceria S and Stubbs R A, Incorporating estimation errors into portfolio selection: Robust portfolio construction, In Asset management, Springer (2006) 270-294.

[2] Davarnia D, Cornuéjols G, From estimation to optimization via shrinkage, *Operations Research Letters 45* (2017) 642-646.

[3] Jorion P, Bayes-Stein estimation for portfolio analysis, *The Journal of Financial and Quantitative Analysis 21* (1986) 279-292.

[4] Ledoit O and Wolf M, Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection, *Journal of Empirical Finance 10* (2003) 602-621.

[5] Markowitz H, Portfolio Selection, *Journal of Finance 7* (1952) 77-91.

[6] Stein C, Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, 1956, pages 197-206.