Shrinkage Covariance Estimation for Mean-Variance Portfolio Optimization

Ethan Nanavati

***Abstract*—This paper investigates the effect of a shrinkage estimate of the covariance matrix on the performance of a portfolio of stocks. Modifying the methodology of Ledoit and Wolf's paper "Honey I Shrunk the Sample Covariance Matrix" [4], It computes a convex combination of the sample covariance matrix and a highly structured constant-correlation matrix. This combination is used as a shrinkage estimator of the underlying covariance matrix and applied in a flipped Mean-Variance portfolio optimization problem. The problem is analyzed and evaluated using convex optimization techniques and an efficient algorithm is presented. Experimental results indicate that the shrinkage estimator more effectively controls the variance of a Mean-Variance portfolio than the sample covariance matrix alone. Moreover, this can be done with minimal reduction in the portfolio's overall returns. The results suggest that the shrinkage estimator is a promising way of estimating the underlying covariance matrix and can be a powerful tool for a manager seeking control over and customization of their portfolio.**

I. Introduction

When managing assets, an important question to answer is how to divide your assets up among investments to ensure that your money grows "as well as possible." This criteria can mean a number of different things to different people, from your portfolio most resembling the performance of a benchmark, to achieving a high ESG score [6]. One of the most common interpretations is to maximize the expected growth of your money subject to restrictions on variance (representative of the amount of risk that you take on).

This problem (mean-variance investing) is solvable for arbitrary covariances between assets [7]. Under certain conditions, estimates for the sample mean and covariance will lead to an accurate solution to the mean-variance investing problem [2]. Unfortunately, in many cases solutions are highly dependent upon estimated values for the mean and covariances of different assets. Extreme values can be computed using maximum likelihood estimators on noisy historical data, resulting in the maximization mean-variance problem biasing towards estimation error rather than any meaningful signal (eg that there is a stock that truly has high mean and low variance) [3]. Thus, it is essential that the mean and covariance parameters are computed carefully. Of special importance is choosing estimation methods that reduce extreme values.

To this end, I modify the methodology of Ledoit and Wolf's paper "Honey I Shrunk the Sample Covariance Matrix." Ledoit and Wolf use a method called shrinkage in which by taking a convex combination of the sample covariance matrix and a structured shrinking matrix, one can reduce the intensity of extreme values and improve expected performance of a mean-variance investor. I decide on an appropriate shrinkage matrix and determine appropriate weights for the convex combination. Then, I use the shrunk covariance matrix (along with sample means) to solve the mean-variance constrained optimization problem for an optimal asset allocation on historical data.

I compare the solutions of mean-variance optimization with sample covariance to solutions with shrunk covariance. Performance is evaluated by tracking the portfolio gains over different periods of time for different datasets.

Notably, I formulate a flipped version of Ledoit and Wolf's portfolio optimization objective wherin expected return is maximized subject to a tolerance level for variance (see (9)). As this formulation is similar to Ledoit and Wolf's original paper (minimize variance subject to an acceptable amount of return)[4], I hypothesized that the shrinkage technique will similarly reduce estimation error and improve performance.

II. Formulation

A. Optimization Problem

The mean-variance problem can be described as maximizing the expected growth of money subject to a restriction on risk. Defining variables:

- x_i is the fraction of the portfolio spent on shares of stock i
- s_i^t is the price of stock i at time t
- $r_i = s_i^{t+1} - s_i^t$ the difference in price of stock i between times t and t+1
- σ_{max} is the maximum amount of risk you want to take on
- n is the number of stocks in consideration (length of x)

The optimization problem becomes:

$$\begin{aligned} \max_x \quad & E[\sum_{i=1}^{n}(x_i r_i)] \\ \text{s.t.} \quad & x \in \Delta^n \\ & Var[\sum_{i=1}^{n}(x_i r_i)] \leq \sigma_{max}^2 \end{aligned} \tag{1}$$

which is equivalent to the vectorized form

$$\begin{aligned} \max_x \quad & E[x^T r] \\ \text{s.t.} \quad & x \succeq 0, \qquad x^T \mathbf{1} = 1 \\ & Var[x^T r] \leq \sigma_{max}^2 \end{aligned} \tag{2}$$

We can make this problem even simpler by considering the $Var[x^T r] \leq \sigma_{max}^2$ term.

$$Var[x^T r] \tag{3}$$

$$= \quad Var\left[\sum_{i=1}^{n}(x_i r_i)\right] \tag{4}$$

$$= \sum_{i=1}^{n} x_i^2 Var(r_i) + 2\sum_{i<j} x_i x_j Cov(r_i, r_j) \tag{5}$$

$$= \sum_{i=1}^{n} Var(x_i r_i) + 2\sum_{i<j} Cov(x_i r_i, x_j r_j) \tag{6}$$

$$= \sum_{i=1}^{n} x_i^2 Cov(r_i, r_i) + 2\sum_{i<j} x_i x_j Cov(r_i, r_j) \tag{7}$$

$$= \quad x^T \Sigma x \tag{8}$$

Where Σ is the covariance matrix of r such that $\Sigma_{i,j} = Cov(r_i, r_j)$. Thus, the overall problem can be rewritten in the simple form(letting $\mu_i = E[r_i]$).

$$\begin{aligned} \max_x \quad & x^T \mu \\ \text{s.t.} \quad & x \succeq 0, \qquad x^T \mathbf{1} = 1 \\ & x^T \Sigma x \leq \sigma^2_{max} \end{aligned} \tag{9}$$

It is well known that the covariance matrix is positive semi-definite, so (9) is a convex optimization problem (indeed it is a quadratically-constrained linear program) and can be solved by established methods (for example, interior point methods).

B. Approximating Covariance Matrix

I will use the methodology of Ledoit and Wolf to compute the shrinkage matrix. That is, using the sample covariance matrix S and a structured matrix F, compute a convex combination

$$S_{shrink} = \hat{\delta^*} F + (1 - \hat{\delta^*}) S \tag{10}$$

that reduces estimation error to better approximate the actual underlying covariance matrix.

The structured matrix I used in this project (F) relies only upon the less noisy variance estimations and infers covariances from them. It assumes that pairwise correlations are the same and multiplies the off diagonal elements by a constant related to the average pairwise correlations. This allows for some additional information from the problem to trickle through, while also heavily smoothing out the data.

$$\begin{aligned} f_{ii} &= s_{ii} \\ f_{ij} &= \bar{r} \cdot \sqrt{s_{ii} s_{jj}} \end{aligned} \tag{11}$$

There exists an estimator for the weight applied to the structured matrix $\hat{\delta^*}$ and is unbiased as T (the number of timesteps provided) goes to infinity [4]. However, the values I obtained in implementing it seemed erroneous, so I instead opted for the heuristically good $\hat{\delta^*} = .6$. This generally results in decreased variance with minimal cost to the expected returns relative to other values.

III. Analysis

(9) is both an equality and inequality constrained convex problem and thus cannot be solved via unconstrained options like Gradient Descent.

Fortunately, the problem satisfies Slater's condition for strategic choices of σ^2_{max}. Note first, that

$$x \succeq 0, x^T 1 = 1 \tag{12}$$

are both affine conditions so it suffices for Slater's that there exists a feasible point. Moreover, the objective function is also affine. We now consider the non-affine condition. Let $y^T \Sigma y$ have minimum value m for feasible y. If $\sigma^2_{max} = m + \epsilon$ for some $\epsilon > 0$, then the inequality holds tightly so Slater's condition is satisfied.

Practically, what this means is that if we have a feasible solution to (9) $x \in R^N$, we can set $\sigma^2_{max} = x^T \Sigma x + \epsilon$ for whatever $\epsilon > 0$ we choose. In particular, we can let x be the unit vector with 0s everywhere except at the entry corresponding to the minimum diagonal entry of Σ. This is certainly feasible, and is an easy way to identify a reasonably low value for σ^2_{max} that admits solutions. For ease, and since in most cases this would be a reasonable lower bound, I enforce this quantity as the minimum allowed value for σ^2_{max}.

Since the problem is convex, and Slater's condition holds, (9) exhibits strong duality. Unfortunately, while a dual problem exists and has the same solution as the primal, the form is exceedingly messy and does not offer much useful information about the problem. Fortunately, since (9) is also twice continuously differentiable in all parts, the KKT conditions determine its optimal value. This is important because it indicates that a wealth of methods can be employed to solving (9), notably interior-point methods.

IV. Algorithm

I chose to use the primal-dual method in solving (9) since the solution satisfies the KKT conditions (which are the basis of the algorithm) and offers good speed. It is especially fast since terms for all gradients and hessians are readily derivable.

See the appendix for a description of KKT matrix, and $r_t(x, \lambda, \nu)$.

V. Methodology

I sought to emulate a portfolio manager trading monthly.

I pulled end of day monthly closing prices of S&P 500 stocks from August 1st 2008 to December 1st 2019 using YahooFinance. These dates were chosen to omit the period around the Covid 19 pandemic while still allowing for a substantial amount of recent data. I included only stocks with complete information in this period, then chose a random sample of N=50 of them for computational ease. I subtracted row-wise to end up with a 50x137 matrix of monthly stock returns.

I batched the data into 76 windows of T=60 contiguous months and used those T months of returns data to construct the sample covariance matrix (S), and the shrunk covariance

Algorithm 1 Primal-Dual Method

Require: $\sigma^2_{max} > \min(\text{diag}(\Sigma)) + \epsilon$ ▷ For Slater's condition
Ensure: $y = x^n$

Initially, x is the unit vector with all 0 except a 1 in the location corresponding to the minimum element along diag(Σ)
$\lambda \leftarrow \mathbf{1} \in R^{N+1}$
$\nu \leftarrow 1$
$\alpha, \beta \leftarrow$ between 0 and 1
$\epsilon_{feas} = \epsilon \leftarrow 10^{-8}$
$||r_{pri}||_2, ||r_{dual}||_2, \eta > epsilon$
while $||r_{pri}||_2 > \epsilon_{feas}$ or $||r_{dual}||_2 > \epsilon_{feas}$ or $, \eta > \epsilon$ **do**
    $[\Delta x^T, \Delta\lambda^T, \Delta\nu^T]^T \leftarrow -KKT^{-1} r_t(x, \lambda, \nu)$ ▷ Find search direction
    s = 1 ▷ Perform a line search for next iterant
    while not $\lambda + s\Delta\lambda \succeq 0$ **do**
        $s \leftarrow \beta s$
    end while
    while $x + s\Delta x$ does not satisfy inequality constraints **do**
        $s \leftarrow \beta s$
    end while
    while $(1 - \alpha s)||r_t(x, \lambda, \nu)||_2 < ||r_t(x + s\Delta x, \lambda + s\Delta\lambda, \nu + s\Delta\nu)||_2$ **do**
        $s \leftarrow \beta s$
    end while
    $x, \lambda, \nu \leftarrow x + s\Delta x, \lambda + s\Delta\lambda, \nu + s\Delta\nu$
    update $||r_{pri}||_2$, $||r_{dual}||_2$, η
end while
return $[x, x^T\mu]$

matrix estimator (S_{shrink}). For simplicity, I set the expected return of each asset to its average value in the current window.

I solved the optimization problem in (9) both using the primal-dual method, and CVXPY 'solve' method for the optimal allocation of portfolio weights for a comparison in time and accuracy. The optimization was solved using both S and S_{shrink} as inputs.

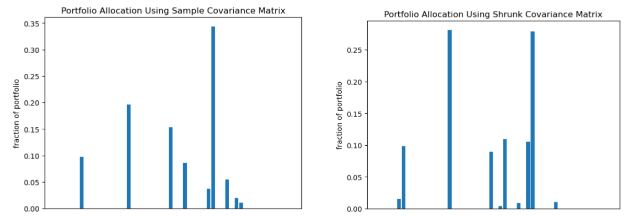
For each time window, I assumed that the portfolio manager would buy on month windowStart+T and sell at windowStart+T+1 (whose prices are unknown to the manager). They invest at the portfolio weights suggested by both solutions for (9) (corresponding to each covariance matrix estimate) and calculate profit after their trading completes at period windowStart+T+1. I determined the variance across the portfolio earnings and its overall return for each of the 76 time windows.

Maximum variance was set at 8 for simplicity and to allow for a larger and more exotic feasible space of portfolio weights.

This framework was used to determine all quantities of interest. Experiments were also run with different samples of stocks from the data to validate results.

VI. Results

Using the methodology described previously, I determined the optimal allocations for stocks in a portfolio. A sample of this from the final window is in Figure 1.

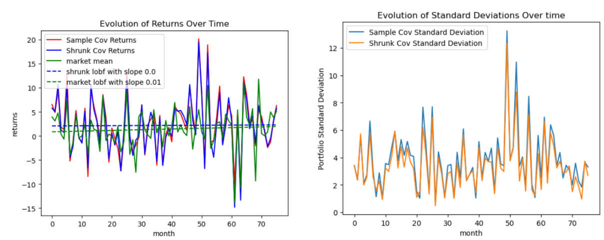


(a) (b)

Fig. 1: Optimal Portfolio Allocations

The Shrinkage estimator approximately preserved the stocks included in the sample covariance portfolio, but adjusted the weights substantially. This is consistent with expectation as certain stocks can be expected to give higher returns than others or can provide more utility to a mean-variance optimizer. Interestingly, however, there are some discrepancies between the portfolio based on S and S_{shrink}. I imagine that this is reflective of hedging by the portfolio and that as different stocks are allocated weight, new assets that are anti-correlated to them are brought into the mix.

Performing this multiple times to get many allocations yielded the results in figure 2.



(a) (b)

Fig. 2: Mean-Variance Performance Over Time

Portfolio Type	Return	Standard Deviation
S_{shrink} estimate	1.535	4.250
S estimate	1.753	4.014
Market Data	1.425	4.436

TABLE I: Performance of different portfolios

Note: Return and Standard Deviation are averages.

The shrinkage estimator almost always found an allocation with lower variance than solutions from the sample covariance matrix. This supports the hypothesis that S_{shrink} more closely resembles the true covariance matrix.

This improvement in variance comes at the expense of some return which is in tune with standard asset management theory. However, Ledoit and Wolf [4] suggest that there should be approximately no difference in the expected returns of each portfolio. Their results suggest that shrunk covariance matrices yield strictly better out-of-sample performance than using the sample covariance matrix. The results here indicate that there is still a balance that must be struck between the improved variance of S_{shrink} and S. I imagine that this difference is caused by the different formulations of our objectives. The fact that Ledoit and Wolf minimize variance subject to a minimum return might allow them to more effectively reap the benefits of a shrinkage estimator.

It is also noteworthy that the returns of both methods follow the market (the N=50 stocks the manager had to choose from) quite well. The returns tend to be appreciably higher than the market average (.78 in this experiment, and similar in others) and the difference is somewhat significant $.1 > p > .05$. However, while the average returns in the market tend to slightly increase, the manager's returns do not meaningfully improve over time. The elevated growth of the market over the manager's portfolio returns is consistent across different groups of stocks. These differences are slight, though, and might be attributable to the specific time period over which I tested. Moreover, while returns are appreciably higher for each manager's portfolios over the market, the variance– as expected– increases as well.

I also sought to determine the impact of increased shrinkage on returns and variance. To that end, I created figure 3 by running the described methodology for various values of $\hat{\delta^*}$ and averaging returns and variance over every trial. As a reminder, $\hat{\delta^*}$ is the parameter that determines how much to shrink the sample covariance matrix in the convex combination in (10).

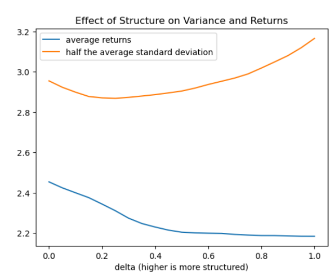


Fig. 3: Covariance Estimator Structure and MV Performance

Note: Standard deviation is scaled to make graphical comparison between returns and volatility easier

With more structure, Variance and returns both decrease. This is expected behavior if weighting the structured matrix truly lessens estimation error for the portfolio's covariance matrix. Notably, though, the variance only decreases to a point, after which it begins to increase with no corresponding increase in returns. This is likely indicative of information loss by the structured matrix. While it might be helpful to weight F into the covariance matrix estimator, it does omit information about the portfolio, so controlling it with inequality will have less of an effect on the actual variance of the portfolio.

Interestingly, the minimum value is closer to 0 than 1. This means that the variance is most effectively controlled with more of S than F. This indicates that the improvement to the covariance estimator is prevalent, but subtle. This is the desired behavior, since the method seeks to reduce estimation error for the covariance matrix and if the error was too extreme (as indicated by the weighting of F over S) it is hard to expect that the signal (true covariance) was recovered.

Lastly, the primal-dual method performed very well, computing consistent solutions with CVXPY CLARABEL in comparable (and sometimes better) time. The real differentiator, though is that CVXPY failed to compute the solution to (9) for some cases when the σ^2_{max} value was too low (< 8). This was a meaningful constraint as it would not allow for more market neutral strategies with stronger variance controls.

The reason for this difference is likely stability. Solvers like CLARABEL convert the feasible set using SOC constraints which reformulates the constraint $x^T \Sigma x \leq \sigma^2_{max}$ into $||\Sigma^{1/2} x||_2 \leq \sigma_{max}$. When the feasible region becomes small, this reformulation can introduce numerical instabilities with the matrix square root, resulting in an infeasible problem. The Primal-Dual method requires no such transformation, and as such is more robust to lower σ^2_{max} .

VII. Conclusion

This paper presented a method of reducing estimation error in the sample covariance matrix to improve the performance of a mean-variance optimized portfolio. I found that a portfolio's variance can be appreciably reduced by leveraging a shrinkage estimator with little (but still relevant) reduction in return. These findings suggest that the flipped MV optimization in (9) can be performed efficiently via a primal-dual algorithm. However, the benefits are not as significant as for the standard problem of minimizing variance with a lower bound acceptable return as put forth in [4].

Nonetheless, it is clear that shrinkage estimators are a powerful tool for a portfolio manager that allows for customization and improved insights.

VIII. Bibliography

References

[1] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[2] Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *The Review of Financial Studies*, 22(5):1915–1953, 12 2007.

[3] Raymond Kan and Guofu Zhou. Optimal Portfolio Choice with Parameter Uncertainty. *Journal of Financial and Quantitative Analysis*, 42(3):621–656, September 2007.

[4] Olivier Ledoit and Michael Wolf. Honey, I shrunk the sample covariance matrix. Economics Working Papers 691, Department of Economics and Business, Universitat Pompeu Fabra, June 2003.

[5] Olivier Ledoit and Michael Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10(5):603–621, 2003.

[6] Abraham Lioui and Patrice Poncet. Optimal benchmarking for active portfolio managers. *European Journal of Operational Research*, 226(2):268–276, 2013.

[7] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

[7] [3] [4] [2] [6] [5] [1]

IX. Appendix

Here, I give a description of some of the matrices and vectors that go into the Primal-Dual algorithm.

The KKT Matrix is of the form indicated in takes the form [1], but is specialized for the mean variance optimization problem in (9).

$$\begin{bmatrix} \lambda_0 \Sigma & Df(x)^T & \mathbf{1} \\ -\operatorname{diag}(\lambda) Df(x) & -\operatorname{diag}(f(x)) & \mathbf{0} \\ \mathbf{1}^T & \mathbf{0} & 0 \end{bmatrix} \in \mathbf{R^{2N+2,2N+2}} \tag{13}$$

Where

$$f(x) = \begin{bmatrix} x^T \Sigma x - \sigma_{max}^2 \\ -\mathbf{x} \end{bmatrix} \in \mathbf{R^{N+1}} \tag{14}$$

$$Df(x) = \begin{bmatrix} 2x^T \Sigma \\ -\mathbf{I_N} \end{bmatrix} \in \mathbf{R^{N+1,N}} \tag{15}$$

$r_t(x, \lambda, \nu)$ is a vector of residuals

$$r_t(x, \lambda, \nu) = \begin{bmatrix} r_{dual} \\ r_{cent} \\ r_{pri} \end{bmatrix} \in \mathbf{R^{2N+2}} \tag{16}$$

where

$$r_{dual} = Df(x)^T \lambda - \mu + \mathbf{1}\nu \in \mathbf{R^N} \tag{17}$$

$$r_{cent} = -diag(\lambda) f(x) - \frac{1}{t}\mathbf{1} \in \mathbf{R^{N+1}} \tag{18}$$

$$r_{pri} = \mathbf{1}^T x - 1 \in \mathbf{R^1} \tag{19}$$

$$t = \frac{10N + 10}{\eta}. \tag{20}$$

These objects are specified so that the update step (line 8) in the primal-dual method solves a second order approximation to the KKT conditions for (9). By repeatedly finding the solution to an approximation of the KKT conditions, the primal-dual method will converge to the optimal constrained solution of the original problem.