WEIGHT SHRINKAGE FOR PORTFOLIO OPTIMIZATION

Ilya Pollak

Purdue University
School of Electrical and Computer Engineering
West Lafayette, IN 47907, USA

ABSTRACT

The paper starts by reviewing the basics of the modern portfolio theory and its very well known drawbacks. After a brief overview of the existing literature that attempts to address these drawbacks, a novel portfolio mixing method is proposed. The method is then illustrated using US stock market data, and is shown to outperform both portfolios that it combines in a statistically significant way. Several avenues of further research are summarized to conclude the paper.

Index Terms—Portfolios, Markowitz, stock, market, finance, covariance, diversification, shrinkage

1. INTRODUCTION

1.1. Returns and Portfolios

The return of a security between trading day \( t_1 \) and trading day \( t_2 \) is defined as the change in the closing price over this time period, divided by the closing price on day \( t_1 \). For example, the daily (i.e., one-day) return on trading day \( t \) is defined as \( (p(t) - p(t-1))/p(t-1) \) where \( p(t) \) is the closing price on day \( t \) and \( p(t-1) \) is the closing price on the previous trading day. Note that if \( t \) is a Monday or the day after a holiday, the previous trading day will not be the same as the previous calendar day.

Suppose an investment is made into \( N \) assets whose return vector is \( \mathbf{R} \), modeled as a random vector with expected return \( \mathbf{\mu} = E[\mathbf{R}] \) and covariance matrix \( \Lambda = E[(\mathbf{R} - \mathbf{\mu})(\mathbf{R} - \mathbf{\mu})^T] \). In other words, \( \mathbf{R} = (R^{(1)}, \ldots, R^{(N)})^T \) where \( R^{(n)} \) is the return of the \( n \)-th asset. It is assumed throughout the paper that the covariance matrix \( \Lambda \) is invertible.

Out of these \( N \) assets, a portfolio is formed with allocation weights \( \mathbf{w} = (w^{(1)}, \ldots, w^{(N)})^T \). The \( n \)-th weight is defined as the amount invested into the \( n \)-th asset, as a fraction of the overall investment into the portfolio: if the overall investment into the portfolio is \( SD \), and \( SD(n) \) is invested into the \( n \)-th asset, then \( w^{(n)} = D^{(n)}/D \). Therefore, by definition, the weights sum to one:

\[
\mathbf{w}^T \mathbf{1} = 1, \tag{1}
\]

where \( \mathbf{1} \) is an \( N \)-vector of ones. Note that some of the weights may be negative, signifying short positions obtained by selling borrowed units of an asset. It is assumed throughout the paper that short selling can be done freely for all \( N \) assets.

It is easily shown then that the total portfolio return is \( \mathbf{w}^T \mathbf{R} \). The expected portfolio return is therefore \( \mathbf{w}^T \mathbf{\mu} \), and the variance of the return is \( \mathbf{w}^T \Lambda \mathbf{w} \).

1.2. Classical Markowitz Portfolio Optimization

The classical Markowitz portfolio framework [34] defines portfolio risk as the variance of the portfolio return, and seeks a portfolio weight vector \( \mathbf{w} \) which minimizes the portfolio risk subject to a target expected return \( \mu_{tgt} \):

\[
\begin{align*}
\text{Find } \mathbf{w}^* \text{ to minimize } & \mathbf{w}^T \Lambda \mathbf{w} & \text{(2)} \\
\text{subject to } & \mathbf{w}^T \mathbf{\mu} = \mu_{tgt} & \text{(3)}
\end{align*}
\]

Using Lagrange multipliers to perform minimization (2) subject to the constraints (1) and (3) yields [34, 38]:

\[
\mathbf{w}^* = \Lambda^{-1} \mathbf{\mu} - \Lambda^{-1} \mathbf{c}, \tag{4}
\]

where \( \mathbf{m} = (\mathbf{\mu}, 1) \), \( A = \mathbf{m}^T \Lambda^{-1} \mathbf{m} \), and \( \mathbf{c} = (\mu_{tgt}, 1)^T \).

The global minimum-variance portfolio (GMVP) is obtained by dropping the mean constraint (3) and instead minimizing portfolio risk (2) subject only to the weight normalization constraint (1). This yields the following weight vector:

\[
\mathbf{w}_{gmvp} = \frac{\Lambda^{-1} \mathbf{1}}{\mathbf{1}^T \Lambda^{-1} \mathbf{1}}. \tag{5}
\]

1.3. Practical Difficulties with the Framework and Prior Literature

In practice, the expected returns and the covariance matrix of the returns are not known and are therefore estimated from historical data [13, 22, 23, 24, 29, 37, 7, 30, 2, 31, 32, 9, 44]. This introduces two well-known problems:

- The optimal portfolio weights are very sensitive to the estimated means and covariances [22, 23, 24, 35, 3, 25, 8, 4]. In other words, a small change of the estimates may lead to a drastic change of portfolio weights.

- The optimal portfolio tends to amplify large estimation errors in certain directions [22, 35]. This stems from the fact that if the variance of an asset (or a sub-portfolio of assets) is significantly underestimated and thus appears to be small, the optimal portfolio will assign a large weight to it. Similarly, a large weight will be assigned if the mean return of an asset or a sub-portfolio appears to be large as a result of being significantly overestimated. As a result, the risk of the estimated optimal portfolio is typically underpredicted and its return is overpredicted [27, 28].

These two issues introduce a significant gap between the theory and practical portfolio management applications. In order to narrow this gap, a number of approaches have been proposed which, roughly speaking, can be grouped into four broad categories:

1. Robust Portfolios: Replacing the original optimization problem (1,2,3) with various robustified versions [17, 43, 6, 16].
2. **Resampled Portfolios**: Randomizing the portfolio selection procedure and computing the average weights from many randomized simulations [22, 25, 15, 39, 36].

3. **Norm-Constrained Portfolios**: Adding a constraint on the norm of the portfolio weights to the optimization problem (1,2,3) [21, 20, 11, 5, 19].

4. **Portfolio Mixtures**: Combining the estimated optimal portfolio with either a fixed portfolio or a portfolio which depends on a small number of estimated parameters [18, 26, 9, 10, 42].

Each of the four groups of approaches has substantial literature behind it; however, all of them come with their own sets of drawbacks which so far have prevented any of them from being widely adopted by the industry practitioners. The main such issue with the last category, portfolio mixtures, is the necessity to choose or estimate the mixing weights. This paper proposes a novel portfolio mixing strategy whose mixing weight selection naturally incorporates the uncertainty of the optimal weight estimates.

### 1.4. A Novel Portfolio Weight Shrinkage Method

It is assumed that investors strive to hold optimal portfolios in the sense of Markowitz—i.e., if the true mean return vector \( \mu \) and covariance matrix \( \Lambda \) were known, then investors would always select the Markowitz optimal portfolio. In reality, neither \( \mu \) nor \( \Lambda \) are known. These quantities are estimated, to produce an estimate of the weight vector for the optimal portfolio. The errors in estimating \( \mu \) and \( \Lambda \) cause errors in estimating the optimal portfolio weights. A commonly used strategy is to construct the portfolio from the estimated optimal weights, despite the fact that they have errors. This strategy is used as a benchmark in the experiments described below.

The strategy I am proposing uses several common-sense considerations that portfolio managers might use in practice. First, suppose that the estimates of \( \mu \) and \( \Lambda \) are either extremely noisy or simply unavailable. In this case, a natural thing to do is to use equal weights for all the assets, \( w^{(n)} = 1/N \) for all \( n \). This is what various authors have called naive [20] or Talmudic diversification [42]. In order to deviate from this allocation, the portfolio manager must possess solid evidence of its suboptimality. I propose to make this judgement on the basis of whether or not the 1/N portfolio is statistically distinguishable from the estimated Markowitz optimal portfolio. The result is a mixture of the 1/N portfolio and the estimated Markowitz optimal portfolio, with the mixing coefficient determined by the statistical distance between the two portfolios. At one extreme, the two portfolios are statistically indistinguishable, and the resulting portfolio mixture is identical to the 1/N portfolio; at the other extreme, the two portfolios are distinct with certainty, and then the resulting mixture is equal to the estimated Markowitz optimal portfolio. The details of this construction are given in the next section, and its empirical evaluation is performed in Section 3.

### 2. DETAILS OF THE PROPOSED METHOD

Let \( \hat{w} \) be the weights for the estimated global minimum-variance portfolio, and let \( w_{eq} \) be the weights for the equally-weighted portfolio. In the experiments below, \( \hat{w} \) is obtained by substituting the sample covariance matrix into Eq. (5); however, any estimator can be used.

The proposed strategy for combining these two portfolios involves four ingredients:

1. Define a matrix \( S \) which measures the amount of information in the estimate \( \hat{w} \). For example, \( S \) may be an estimate of the inverse covariance matrix of \( \hat{w} \); however, a simpler alternative is used in the experiments below. Specifically, \( S \) is chosen to be a diagonal matrix whose only nonzero diagonal entry is the reciprocal of the maximum sample variance of all the components of \( \hat{w} \).

2. Compute the statistical distance

\[
D(\hat{w},w_{eq}) \equiv \sqrt{(\hat{w} - w_{eq})^T S(\hat{w} - w_{eq})}
\]

between the estimated global minimum-variance portfolio and the equally-weighted portfolio. Without the matrix \( S \), this would be a regular Euclidean distance between the two portfolio weight vectors. However, the intention here is to also measure the confidence that the GMVP is distinct from the equally-weighted portfolio. Matrix \( S \) serves this purpose: it is designed to reduce the distance if there is significant uncertainty in the GMVP estimate \( \hat{w} \), and to increase the distance if the uncertainty is small.

3. Define shrinkage intensity \( g(D(\hat{w},w_{eq})) \) that satisfies the following three properties:

- \( g(0) = 1 \)—i.e., if the estimated global minimum variance portfolio \( \hat{w} \) and the shrinkage target \( w_{eq} \) are statistically indistinguishable, then the combined portfolio will consist entirely of the shrinkage target;
- \( g(\infty) = 0 \)—i.e., if ones is fully confident that \( \hat{w} \) is distinct from \( w_{eq} \), then the weight of the shrinkage target in the mixture portfolio will be zero;

The specific form of \( g \) used below is:

\[
g(x) = \frac{1}{1 + 0.5x}.
\]

This particular function \( g \) has the additional property that if \( D(\hat{w},w_{eq}) = 2 \)—which essentially means that \( w_{eq} \) is two standard deviations away from \( \hat{w} \)—then the corresponding shrinkage intensity is 0.5.

4. Define the combined portfolio weights \( w_c \) as follows:

\[
w_c = w_{eq}g(D(\hat{w},w_{eq})) + \hat{w}(1 - g(D(\hat{w},w_{eq}))).
\]

### 3. EXPERIMENTS WITH MARKET DATA

The experiment described in this section uses the daily closing prices for nine ETFs downloaded from Yahoo! Finance on July 5, 2011: XLB, XLE, XLF, XLI, XLK, XLP, XLU, XLV, and XLY. These ETFs correspond to the following sectors, respectively: materials, energy, financial, industrial, technology, consumer staples, utilities, health care, and consumer discretionary. Together, these ETFs cover most of the large US stocks. Therefore, investing in these ETFs is equivalent to the problem of investing in the US stock market and trying to decide how to diversify among the major sectors of the US economy.

All these nine ETFs have existed since December 22, 1998. Therefore, daily return data exist starting on December 23, 1998. All the daily returns are formed from the closing prices on consecutive trading days. The last day for which the data was available at the time of the download was July 1, 2011. For each day, all sample means, sample standard deviations, and sample covariances of returns are computed using 60 previous trading days. The computation of \( \hat{w} \) can therefore start on the 61-st day of the daily return data. In order to compute the distance \( D \), the sample standard deviation of
and defined to be the largest sample standard deviation among the nine; and \( \hat{w}^* \) is defined as the corresponding component of \( \hat{w} \). The distance \( D \) is then defined as \( |\hat{w}^* - 1/9|/s \). This is fed into the function \( g \) defined above that specifies the shrinkage intensity and is used in constructing the combined portfolio \( \hat{w}_c \). Since a history of 60 days of \( \hat{w} \) is needed in order to start calculating \( \hat{w}_c \), the test results commence on the 121-st day of the daily return data, which is June 17, 1999. The 120 daily returns over the period 12/23/1998–6/16/1999 are therefore only used for training.

Fig. 1. Cumulative returns during 6/17/1999–7/1/2011 of the proposed portfolio mixture (red), the global minimum-variance portfolio estimated using the sample covariance (blue), and the equally-weighted portfolio (black). Each portfolio contains nine ETFs: XLB, XLE, XLF, XLI, XLK, XLP, XLU, XLV, and XLY.

On each day starting June 17, 1999, the proposed portfolio \( \hat{w}_c \) is constructed and its daily return is recorded. The cumulative returns over the entire simulation period 6/17/1999–7/1/2011 (3031 trading days total) are shown in Fig. 1 in red. For comparison, the same figure shows the cumulative returns over the same period of the estimated global minimum-variance portfolio (in blue) and the equally-weighted portfolio (in black). The total cumulative returns for the three portfolios are 74% for the proposed mixture, 29% for the estimated GMVP, and 71% for the equally-weighted portfolio. The respective sample standard deviations of portfolio returns are (with standard errors computed using [1] in parentheses): 0.0101 (1.3 \( \times \) 10\(^{-4} \)), 0.0091 (1.2 \( \times \) 10\(^{-4} \)), and 0.0127 (1.6 \( \times \) 10\(^{-4} \)). (Both the standard deviation estimates and their standard errors are reported as daily quantities, computed over the entire simulation period.) Therefore, the proposed mixture achieves approximately the same overall out-of-sample return as the equally-weighted portfolio but with a significantly lower risk: the estimated standard deviation of the mixture portfolio is about 20% lower than that of the equally-weighted portfolio, and the difference is statistically significant. On the other hand, the return of the estimated GMVP is less than half that of the mixture portfolio.

A common metric of the tradeoff between the portfolio return and risk is the Sharpe ratio [41] which is the ratio of the mean return to the standard deviation of the return, typically reported as an annualized quantity. For the portfolio mixture, the annualized Sharpe ratio is 0.37; the annualized Sharpe ratios for the estimated GMVP and the equally-weighted portfolio are 0.22 and 0.32, respectively.

4. CONCLUSIONS AND FUTURE WORK

A novel method has been proposed for constructing mixture portfolios and illustrated it using real data from the US equities market. The proposed portfolio formation strategy performs better out of sample, in a statistically significant way, than either of the two portfolios that it combines: an estimate of the global minimum-variance portfolio and an equally-weighted portfolio.

While the proposed technique shows initial promise, several important lines of investigation remain.

- Portfolio rebalancing is associated with trading costs. Therefore, in practice it is cheaper to run portfolio strategies that require less rebalancing. Therefore, the first item for future research is modeling and estimating portfolio trading costs.
- The function \( g \) that determines the shrinkage intensity can be generalized using a shape parameter:

\[
g(x) = \frac{1}{1 + bx}.
\]

The optimal shape parameter \( b \) can be learned using historical data. For example, on each day, the value of \( b \) may be selected that would have optimized the Sharpe ratio during the period that starts at the beginning of the simulation and ends on the previous day. This is the second avenue of future research.
- It may also be interesting to investigate different ways of constructing the matrix \( X^* \) whose inverse essentially characterizes the amount of uncertainty in the weight estimate \( \hat{w} \).
- More extensive experiments must be run—both applying the proposed framework to different estimators of the GMVP and to different data sets.
- The proposed framework must be evaluated against other weight shrinkage estimators from the literature. For example, [10] propose mixing the equal-weighted portfolio with the GMVP, albeit with a different mixing strategy.

5. ACKNOWLEDGMENTS

I would like to thank Nathan Mullen and Ka Ki Ng for many helpful discussions.

6. REFERENCES
