Applied Mathematics
&
Statistics

Technical Report - Fall 2022

Mean-Variance Portfolio Optimization using Elastic Net Penalty

Jason Bohne and Jarryd Scully

inspired by Weighted Elastic Net Penalized Mean-Variance Portfolio Design and Computation in [Ho et al., 2015]

Contents

1 Introduction

In 1952, Harry Markowitz introduced a mathematical framework for optimal stock selection that revolutionized the field of finance [Markowitz, 1952]. Markowitz's criterion was constructed to balance a portfolio's risk and return, while operating under a handful of key assumptions. Most importantly, investors were assumed to be rational and would always prefer a portfolio with minimal risk for a given expected return. With this assumption, the formulated selection criterion is stated to minimize a portfolio's risk subject to a minimum expected value of the return.

Markowitz's theoretical approach may be simply defined, but faces justified criticisms in practice. Namely, a participant is required to specify the expected return of each asset in the portfolio and its covariances. A fund manager will not know the future means and covariances of assets and is therefore exposed to estimation risk. An incorrect parameterization of Markowitz's model may result in a portfolio that will perform sub-optimally.

The naive approach for parameterization is estimating the unknown mean and covariance matrix using historical data. While intuitive, historical data selection will force the fund manager to make some strong assumptions on the behavior of the assets. Historical data is limited and the practitioner will be forced to determine which data is relevant. There are many different methods to aid the parameterization of Markowitz's model, this paper will focus on Bayesian shrinkage methods.

To further improve out-of-sample performance, this paper will attempt to regularize the portfolio selection criterion using a weighted elastic net approach. A weighted elastic net penalty is a linear combination of a portfolio's weighted $\ell 1$ norm and the weighted $\ell 2$ norm. This approach will be justified by reformulating the mean-variance criterion and evaluating the out-of-sample performance. The inclusion of the elastic net penalty adds computational complexity during optimization. This will be addressed by utilizing the Split-Bregman [Goldstein and Osher, 2009] algorithm that exploits the sparse nature of the penalized solutions.

2 Portfolio Construction

This section will discuss the portfolio selection criterion proposed by Markowitz and introduce the altered objective function under the elastic net penalty.

2.1 Mean-Variance Criterion

Markowitz's original ideas introduced an optimal method for constructing a portfolio off of single-period daily returns. For a portfolio of risky assets with a given mean μ and covariance matrix Γ, the optimal distribution of assets w can be solved by the following optimization problem.

$$\min_w \gamma w^T \Gamma w - \mu^T w$$

In this framework, γ represents a risk aversion coefficient. Other papers have addressed the selection and effects of γ on the resulting portfolio. This paper will not attempt to address risk aversion selection methods and will perform the optimization assuming $\gamma = 1$.

As previously stated, the practitioner's contention with Markowitz's original ideas is the assumption of an expected mean return μ. Under these ideas, the optimal weight allocation w^* satisfies

$$\Gamma w^* = \mu$$

This paper will build off this idea and define the optimal mean-variance objective as the solution with a maximum return and a minimum variance. Under these ideas, the new mathematical model seeks to find some return $\mu_a \leq \mu_i \leq \mu_b$ and the resulting optimal set of weights. The updated optimization problem is the following.

$$\begin{gathered}
\min\left(\sqrt{\textstyle\sum_i \sum_j W_i W_j \sigma_{ij}} \mu \right) \\
\text{subject to} \\
\left(\textstyle\sum_{i=1}^{N} W_i \mu_i \right) \leq R \\
\textstyle\sum_{i=1}^{N} W_i = 1 \\
0 \leq W_i \leq 0.10
\end{gathered}$$

The approach will constrain a manager to only take long positions in assets, each capped at a maximum weight of some upper bound of the total portfolio holdings. This may interfere with the ability to produce a sparse set of weights, but more accurately follows practical approaches in finance. A portfolio that is overexposed to a select few positions may face unobservable systematic risks that cannot be accounted for using the mean-variance optimization approach.

2.2 Regularization

The elastic net penalty is a regularized regression method that takes a linear combination of the penalties presented in the LASSO and Ridge approaches. The weighted elastic net penalty can be defined using the following expressions.

$$||w||_{\lambda_1,\ell_1} = \lambda_1 \sum_{k=1}^{N} |w_k|$$
$$||w||^2_{\lambda_2,\ell_1} = \lambda_2 \sum_{k=1}^{N} |w_k|^2$$

Considering the mean-variance optimization criterion, the penalized objected function is updated using the elastic net penalties.

$$\min_w w^T \hat{\Gamma} w - \hat{\mu}^T w + ||w||_{\lambda_1,\ell_1} + ||w||^2_{\lambda_2,\ell_1}$$

Several justifications for using ℓ_1 and squared ℓ_2 norms as penalties have been well documented in academic literature. In the context of portfolio optimization, the use of a uniformly weighted ℓ_1 penalty can result in a sparse asset allocation. The use of an ℓ_2 penalty can stabilize the inverse covariance matrix, which can often be ill-conditioned. Following the previously laid out methodology, the out-of-sample performance of the penalized objective function is tested using both the unbiased in-sample historical and biased shrinkage estimators.

3 Parameter Estimation

3.1 Regularized Covariance Matrix

Due to the potentially large amount of assets a fund manager can hold at a particular time, this paper will attempt to improve out-of-sample performance by using shrinkage on the mean and covariance. The portfolio's covariance matrix is redefined as a linear combination of the in-sample historical covariance matrix and the identity matrix.

$$\hat{\Gamma} = \rho_1 \Gamma_S + \rho_2 I$$

Both ρ_1 and ρ_2 are constants defined using the approach outlined in [Ledoit and Wolf, 2004]. Constraining $\rho_1, \rho_2 > 0$ will ensure that $\hat{\Gamma}$ will always be positive definite.

$$\rho_1 = \frac{E[||S - \Sigma||^2]}{E[||S - \mu I||^2]}$$
$$\text{with } \rho_1 + \rho_2 = 1$$

With $S = XX^t/n$, Σ as the true covariance matrix, and $\mu = \Sigma \cdot I$. The intuition and proofs behind the methods employed by Ledoit and Wolf can be further studied in their cited paper.

3.2 Biased James-Stein Estimation

With the estimated covariance matrix defined, the mean can be estimated using the James-Stein approach [Jorion, 1986]. Stein has shown that the classical sample mean is inadmissible and should be outperformed by a Bayes estimator. In the application of portfolio optimization, the Bayes-Stein estimators should provide gains during portfolio selection. The estimated mean is computed using the following equation.

$$\hat{\mu} = (1 - \rho)\mu_s + \rho\eta 1$$

Where μ_s represents the in-sample mean returns. Shown here, the Stein estimated mean is a linear combination of the scaled sample returns and the mean return across the portfolio assets. This computation will bias the mean return of each asset towards the total mean return of the portfolio, the more central mean. Both η and μ are defined using

$$\eta = \min\left(\tfrac{1}{N}\sum_{i=1} N\mu_{S,i}, 0.0004\right)$$
$$\rho = \min\left(1, \tfrac{N-2}{T_{train}(\mu_S = \eta 1)^T\hat{\Gamma}^{-1}(\mu_S - \eta 1)}\right)$$

With T_{train} representing the number of training days and N being the number of assets in the portfolio. For larger sets of training data, the correction due to estimation risk may disappear. In other words, the biased estimators $\hat{\mu}$ and $\hat{\Sigma}$ will tend to the naive values μ and Σ. The coefficients for the biased estimator can be redefined as the following.

$$\rho = \min\left(1, \tfrac{N+2}{(N+2)+T_{train}(\mu_S = \eta 1)^T\hat{\Gamma}^{-1}(\mu_S - \eta 1)}\right)$$

The performance of the shrinkage estimators is tested using the out-of-sample portfolio performance and compared to the results of the naive approach.

4 Statistical Inference

4.1 Split Bregman Algorithm

The algorithm proposed within [Ho et al., 2015] is the Split Bregman algorithm originally proposed to solve L1 regularization problems such as absolute shrinkage and total variation within [Goldstein and Osher, 2009]. The general idea of the algorithm is to reformulate the penalized objective into distinct optimization problems that are then solved iteratively. The first problem is an unconstrained quadratic problem that incorporates the L2 penalty. This can be solved either in closed form or numerically with a minimization routine. The second problem is a constrained linear problem incorporating the L1 penalty which can be solved via soft thresholding. An original implementation for such can be found in our implementation.

As the hyperparameters λ_1 and λ_2 must be specified on the initialization of the algorithm; it is common practice to tune these by minimizing MSE across a grid with cross-validation. Functionality for this can again be found within our implementation.

Algorithm 1 Split Bregman Algorithm

Require: $(\hat{R}, \hat{\mu}, \lambda), k = 1, b^k = 0, w^k = 0, d^k = 0$
  while $||w^{k+1} - w^k||_2 > \epsilon$ **do**
    $w^{k+1} =_w w^T \hat{R} w - w^T \hat{\mu} + \frac{\lambda}{2}||d^k - \beta w^k - b^k||_2^2$
    $d^{k+1} =_d \frac{\lambda}{2}||d - \beta w^{k+1} - b^k||_2^2 + ||d||_1$
    $b^{k+1} = b^k + \beta w^{k+1} - d^{k+1}$
    $k+ = 1$
  end while

5 Numerical Results

5.1 Data Preprocessing

U.S. equity OHLCV data was fetched using Yahoo Finance by [Aroussi, 2015] between 2010 and 2020 for underlying components of the S&P500. Assets were removed if they were missing more than 1% of observations across the time period, adjustable in the feature preprocessing scripts. Log returns were generated between daily close prices with the full data set being partitioned into a training and testing set using an equal allocation split.

5.2 Problem Specification

This analysis includes the comparison of five methods for portfolio construction; respectively the Minimum Variance Criteria, Mean-Variance Criteria, Biased Mean Variance Criteria, Mean-Variance Criteria with Elastic Net Penalty, and lastly Biased Mean-Variance Criteria with Elastic Net Penalty. The first three problems can be solved using the closed-form solution or numerically. The last two are solved with the original implementation of the Split-Bregman. All portfolios are long-only with an upper bound on the portfolio weight of 10%. A parameter grid across regularization parameters λ_1 and λ_2 were geometrically equally spaced between $10e^{-6}$ and 1.

5.3 Unbiased Mean-Variance Optimization

As in the original Markowitz framework, the unbiased mean-variance objective is solved by maximizing the risk-reward payoff within the potential set. Here the potential set encompasses all target returns $\mu_i \in [\mu_a, \mu_b]$ such that μ_a is the return of the minimum variance portfolio and μ_b is the return of the maximum return portfolio. To return a single portfolio from the optimization, a risk aversion constant must be chosen to result in the target return. In other words, a practitioner must set their risk preference a priori to balance out the mean-variance relation. In the analysis, we compute the optimization across a grid of 10 potential target returns; generating the sample efficient frontier below. The estimate for the true efficient frontier for the unbiased estimates is determined via smoothing splines under monotonicity constraints using [Servén D., 2018] and is marked in red.

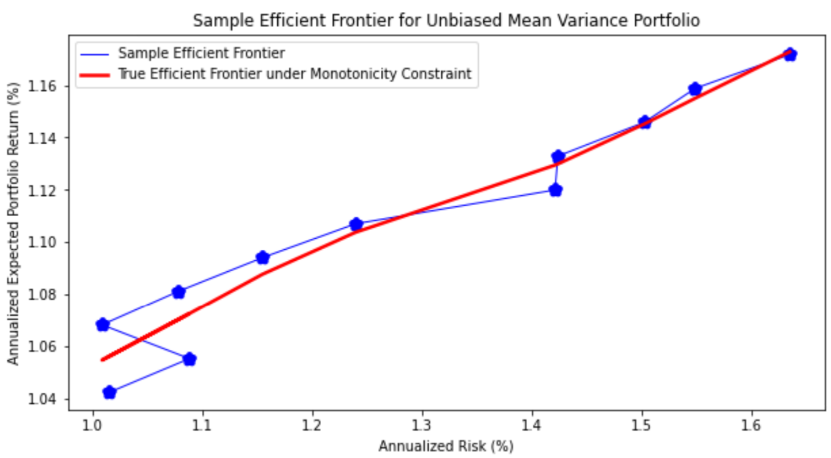

Figure 1: Unbiased Mean-Variance Optimization Efficient Frontier

The risk-reward metric used for all results will be the Sharpe ratio defined by

$$S_r = \frac{\mu_i - r_f}{\sigma}$$

To compute the Sharpe Ratio, the risk-free rate is set to the yearly average of the "on the run" 13-Week US Treasury rates. Although they are not truly risk-free in practice, the Treasury market is sufficiently deep and liquid to serve as a benchmark. Examining the Sharpe values across the unbiased mean-variance sample efficient frontier.

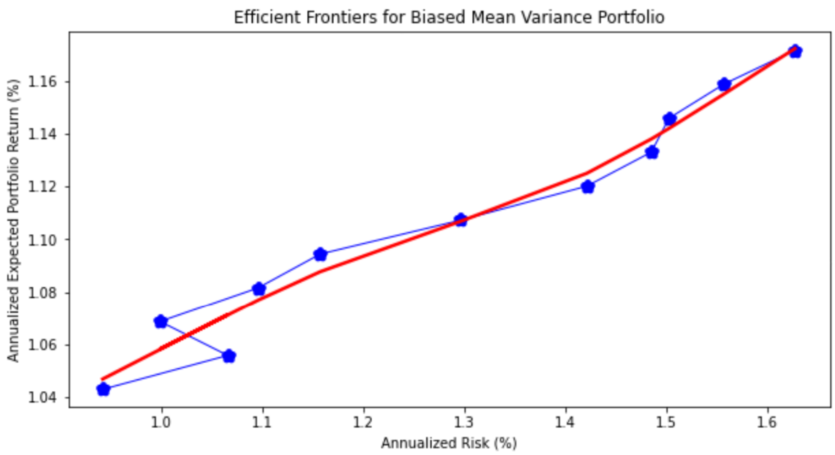

Figure 2: Biased Mean-Variance Optimization Efficient Frontier

Target Return	**Portfolio Standard Deviation**	**Sharpe**
1.06	1.01	1.05
1.06	1.09	0.97
1.07	1.01	0.94
1.08	1.08	1.00
1.09	1.15	0.94
1.11	1.24	0.89
1.12	1.42	0.78
1.13	1.42	0.79
1.15	1.50	0.76
1.16	1.55	0.74
1.17	1.64	0.71

Note the optimal portfolio has an expected return of $\mu_i = 1.06$ with a standard deviation of 1.01, resulting in a Sharpe ratio of 1.05.

5.4 Biased Mean-Variance Optimization

In addition to the naive unbiased sample estimates used in the original mean-variance framework proposed in [Markowitz, 1952], the authors also include biased estimates as discussed in 3. For the estimation of the mean asset return; the James Stein estimator has been used to bias the estimates toward the average market return of the training set. For the covariance matrix, L2-regularization is performed using a selection process performed on the naive in-sample mean and covariance estimates.

Target Return	Portfolio Standard Deviation	Sharpe
1.06	0.94	1.12
1.07	1.06	1.00
1.08	1.10	0.98
1.09	1.16	0.94
1.11	1.30	0.85
1.12	1.42	0.78
1.13	1.49	0.76
1.15	1.50	0.76
1.16	1.56	0.74
1.17	1.63	0.72

Finally, we can compare the estimates for the true efficient frontier for both the unbiased and biased estimators. Note as these are quite similar the effectiveness of using biased estimates is in question. Future work exploring the strength of bias is worth pursuing.

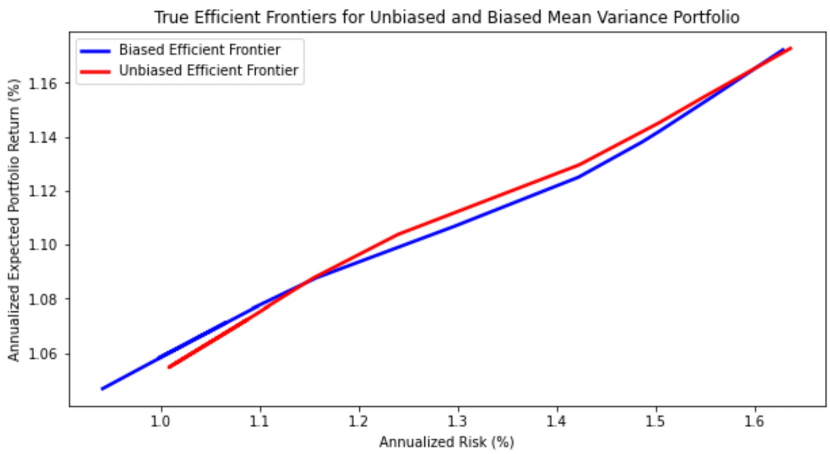

Figure 3: Estimates for True Efficient Frontier using Smoothing Splines

5.5 Mean-Variance Optimization with Regularization

As performed in [Ho et al., 2015]; L1 and L2- regularization are incorporated into the original mean-variance objective function. In order to solve the inference problem; the Split Bregman algorithm is utilized as discussed in 4.1. Performing cross-validation across the parameter grid of λ_1 and λ_2, optimal values are selected for both the unbiased and biased estimates.

Given the selected hyperparameters; the inference problem can now be solved via the Split-Bregman algorithm for both the unbiased and biased estimates; specified by the optimal target mean respectively. These results are shown in 1

Estimator	λ_1	λ_2	MSE
Unbiased Estimates	$10e^{-5}$	$10e^{-4}$	0.08
Biased Estimates	$10e^{-5}$	$10e^{-4}$	0.07

Table 1: Selection of Optimal Regularization Hyperparameters

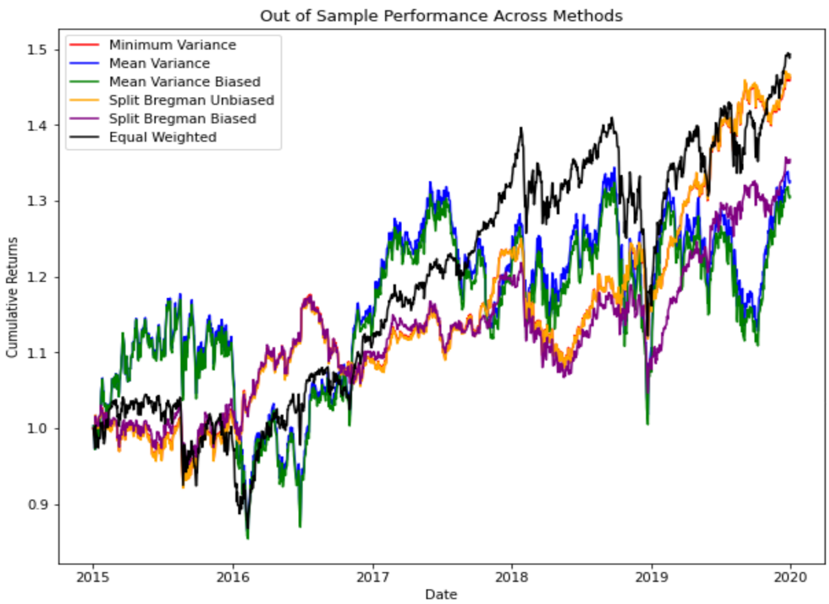

Figure 4: Out of Sample Cumulative Equity Curve

Evaluation of the approaches above was on an out-of-sample data set spanning 2015-2020. Both the cumulative equity curve for each method along with the naive equally weighted portfolio and the Out-of-Sample Sharpe Ratios in 2.

Estimator	**Sharpe Ratio**
Minimum Variance	0.93
Mean-Variance	1.33
Biased Mean Variance	1.26
Mean-Variance with Elastic Net Penalty	0.92
Biased Mean-Variance with Elastic Net Penalty	1.07
Equally-Weighted	1.03

Table 2: Sharpe Ratios across Portfolio Selection Criteria

6 Conclusion

This paper built on the original ideas proposed by Markowitz in an attempt to improve the out-of-sample performance of the mean-variance optimal portfolio. Firstly, the naive estimators for the mean and covariance were compared against biased estimators. The biased mean and covariance matrix was computed using the James-Stein approach. Most importantly, Markowitz's original objective criterion was penalized using the elastic net penalties. This penalization typically promotes sparse solutions relative to the unpenalized objective.

Performing real data analysis on U.S. equity data spanning 2010-2020; optimal portfolios were constructed under the Minimum Variance Criteria, Mean-Variance Criteria, Biased Mean Variance Criteria, Mean-Variance Criteria with Elastic Net Penalty, and Biased Mean-Variance Criteria with Elastic Net Penalty. This included both the generation of a sample efficient frontier in addition to the implementation of the Split-Bregman algorithm to solve the penalized optimization problem. Methods were evaluated on an out-of-sample test set with the computation of equity curve and Sharpe ratios. Overall while the presence of biased estimates and regularization improved performance relative to the naive minimum variance; performance was still suboptimal to the original mean-variance objective proposed by Markowitz.

7 Appendix

7.1 Code Access

The full repository of all scripts and data utilized in this project can be accessed from our **GitHub**. Note the csv files containing the raw data for our real data analysis can be found **here**

References

[Aroussi, 2015] Aroussi, R. (2015). Yahoo finance library for python.

[Goldstein and Osher, 2009] Goldstein, T. and Osher, S. (2009). The split bregman method for l1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343.

[Ho et al., 2015] Ho, M., Sun, Z., and Xin, J. (2015). Weighted elastic net penalized mean-variance portfolio design and computation. *SIAM Journal on Financial Mathematics*, 6.

[Jorion, 1986] Jorion, P. (1986). Bayes-stein estimation for portfolio analysis. *The Journal of Financial and Quantitative Analysis*, 23(1).

[Ledoit and Wolf, 2004] Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *The Journal of Multivariate Analysis*, 88(1):365–411.

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

[Servén D., 2018] Servén D., B. C. (2018). pygam library for python.