Exploring Hyperparameter Spaces: A Comprehensive Study of Ridge Regularization in Mean-Variance Optimization

Fernando Urbano

May 9, 2024

1 Introduction

Markovitz's construction of the Efficient Frontier in 1952 is among the most significant improvements in quantitative finance and set the start of the Modern Portfolio Theory.

The idea behind the Efficient Frontier provided an algorithm to choose optimal weights for the assets in the portfolios of hedge funds, banks, and other financial institutions.

It is part of the mean-variance optimization framework and served as the foundation for the CAPM theory of Willian Sharpe.

The original algorithm was highly praised for its simplicity accompanied, nonetheless, by a powerful conclusion: portfolio optimization depends on expected returns and risk (measured by variance) and investors aim to maximize expected returns given their levels of risk aversion. With mean-variance optimization, investors can transform their views into investment bets.

With time, the original mean-variance optimization started facing criticism due to estimation error and extremely aggressive allocation as a consequence of mathematical instability. Nowadays, improvements in the field of Machine Learning found paths to mitigate the problem through the use of Regularization (L1 and L2) and Resampling (Monte Carlo and Bootstrapping).

While the original algorithm provides a concise set of hyperparameters, modern applications with regularization techniques require extensive hyperparameter tuning and a gigantic number of possible combinations on how to address the problem. Practitioners often choose one possible set among the vast poll of possibilities, given limited time to train and analyze results.

The goal of our paper is to dive deeper into the tuning of hyperparameters for RIDGE regularization, tirelessly testing possible combinations of different parameters to arrive at general guidelines on how to approach the problem and which sets generate more favorable results.

2 Methodology

2.1 Mean-Variance Optimization

The optimization process is based on finding the weights for assets in a given portfolio that maximize the expected return of the portfolio given a level of risk or minimize the risk of a portfolio given a level of expected return:

$$\max_{w} \quad \mu^T w \qquad \text{s.t.} \quad w^T \Sigma w \le \sigma^2, \quad \sum_{i=1}^{n} w_i = 1$$

In the equation, w is the vector of weights, μ is the vector of expected returns, Σ is the covariance matrix of the returns, and σ^2 is the tolerated level of risk. The first constraint ensures that the risk of the portfolio is below a certain level, and the second constraint ensures that the investor uses 100% (not more nor less) of their capital in the allocation.

The Efficient Frontier gives the best allocation for every given risk (variance) level. The curved shape is a consequence of diversification: less than perfect correlation between assets allows for a reduction in the overall risk of the portfolio.

In the picture we highlight two points worth of special attention: the Global Minimum Variance Portfolio (GMV), the leftmost point, and the Tangency Portfolio.

2.2 Global Minimum Variance Portfolio

The Global Minimum Variance Portfolio (GMV) is the portfolio with the lowest possible variance.

Given the convex nature of the optimization problem, it is possible to find the $w_G MV$ with a closed form solution:

$$w_{\text{GMV}} = \arg\min_{w} \quad w^T \Sigma w \qquad \text{s.t.} \sum_{i=1}^{n} w_i = 1$$

$$w_{\text{GMV}} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}^T \Sigma^{-1}\mathbf{1}}$$

In the equation, $\mathbf{1}$ is a vector of ones and Σ^{-1} is the inverse of the covariance matrix.

2.3 Tangency Portfolio

The Tangency Portfolio is the portfolio that maximizes the Sharpe Ratio, a measure of risk-adjusted return, defined as the ratio of the excess return of the portfolio over the risk-free rate to the standard deviation of the portfolio (square root of the variance):

$$\text{Sharpe Ratio (SR)} = \frac{\tilde{\mu}^T w}{\sqrt{w^T \Sigma w}}$$

Where $\tilde{\mu} = \mu - r_f \mathbf{1}$ is the vector of excess returns, with r_f being the risk-free rate.

Again, given the convex nature of the optimization problem, it is possible to find the w_{TAN} with a closed form solution:

$$w_{\text{TAN}} = \arg\max_{w} \quad \frac{\mu^T w - r_f}{\sqrt{w^T \Sigma w}} \qquad \text{s.t.} \sum_{i=1}^{n} w_i = 1$$

$$w_{\text{TAN}} = \frac{\Sigma^{-1}\tilde{\mu}}{\mathbf{1}^T \Sigma^{-1}\tilde{\mu}}$$

2.4 Instability and Unpredictability

The two formulas give us the optimal weights for the given levels of risk. Any other point in the efficient frontier can be obtained by a linear combination of the GMV and the Tangency Portfolio.

$$w_{\text{optimal}} = \alpha w_{\text{GMV}} + (1 - \alpha) w_{\text{TAN}} \qquad \alpha > 0$$

However, the expected returns and the covariance matrix are estimated from historical data, and the estimates are subject to error. Returns present small or none autocorrelation, leading to a high degree of unpredictability. The estimates of the covariance matrix are slightly more predictable, but, given the high dimensionality of the problem, small errors in the estimates lead to large errors in optimization and final allocation.

More specifically, the errors are magnified by the Σ^{-1} term in the formulas. The inverse of the covariance matrix is highly sensitive to small changes in the estimates, given its high condition number (ratio of the largest to the smallest eigenvalue of the matrix), which serves as a measure of how much the output of the matrix changes due to small changes in the input, and give often highly extreme optimal weights, with large allocations in few assets.

In simpler words, a small difference between the historical and future excess returns ($\tilde{\mu}$) or the covariance matrix (Σ) lead to a large difference in the optimal weights for the training period and the future period.
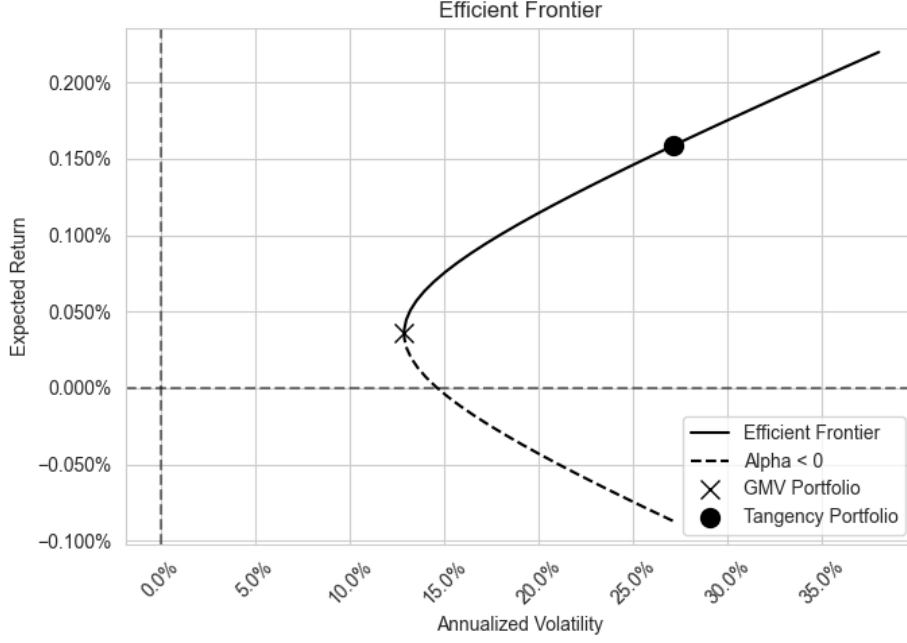
Figure 1: Efficient Frontier

2.5 Ridge Regularization

Ridge Regularization is a technique used to mitigate the problem of instability in the optimization process. It adds a penalty term to the optimization problem. The point of minimum variance with the new term is called the Regularized Minimum Variance Portfolio (RMVP).

$$w_{\mathrm{RMVP}} = \underset{w}{\arg\min} \quad w^T \Sigma w + \lambda \|w\|^2 \qquad \text{s.t.} \sum_{i=1}^{n} w_i = 1$$

The λ term is the hyperparameter that controls the trade-off between the variance of the portfolio and the penalty term. As in regression problems, Ridge has the advantage of providing a closed-form solution:

$$w_{\mathrm{RMVP}} = \underset{w}{\arg\min} \quad w^T \Sigma w + \lambda w^T w$$
$$w^T \Sigma w + w^T \lambda I w$$
$$w^T (\Sigma + \lambda I) w$$

The problem is now again a quadratic form, and the solution is:

$$w_{\mathrm{RMVP}} = \frac{(\Sigma + \lambda I)^{-1} \mathbf{1}}{\mathbf{1}^T (\Sigma + \lambda I)^{-1} \mathbf{1}}$$

The shrinks the optimal weights towards the center of the efficient frontier, reducing the extreme allocations in few assets and decreasing the sensitivity of the optimization process to small changes in the estimates.

In practice, the solution vector w degree of penalty is defined by how much the individual weights differ from the equal weight allocation of of $1/n$, where n is the number of assets in the portfolio.

2.6 Methodology

In this paper, we explore the hyperparameter space of the RMVP. Using the point of minimum variance steady of the Regularized Tangency Portfolio is viewed as a better approach, given that historical
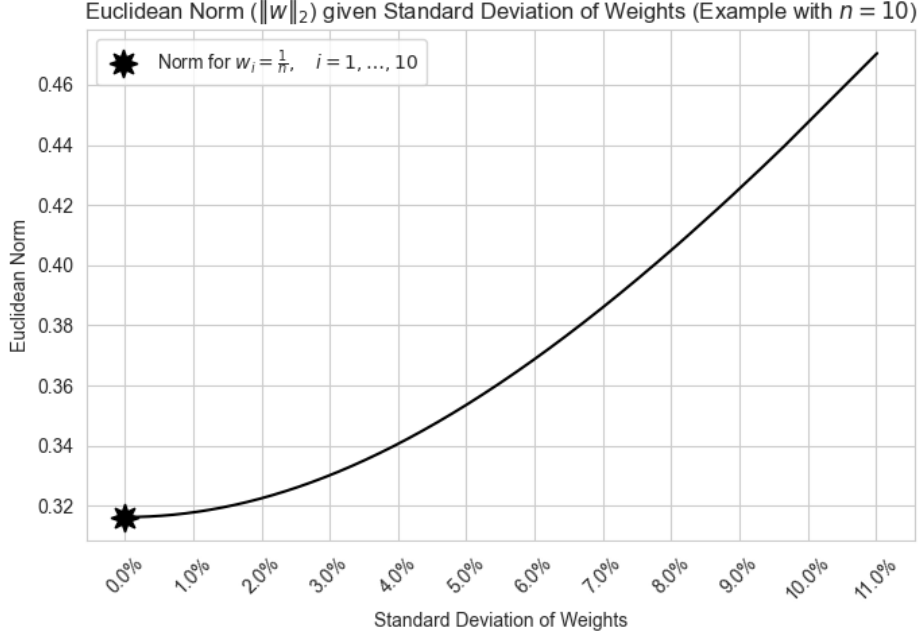
3

Figure 2: Euclidean Norm as a Function of Weights' Standard Deviation

returns are not good predictors of future returns and we intend to remain agnostic about more complicated models on how to predict, given that the focus should remain in the exploration of the hyperparameter of the optimization.

In our tests, we use the conventional format of the regularized optimization problem. Given the selected assets and training data, we use time-series cross-validation to find the optimal λ for the RMVP selecting based on the Sharpe Ratio performance. After have choosen the optimal λ, we calculate the RMVP in the training sample and see their performance in the testing sample.

The hyperparameter space tested is defined by the following parameters:

- Choice of number of assets n: 5, 10, 12, 17, 30.

- Choice of training size T in days: 63, 126, 252, 504.

- Choice of testing size t in days (how long the portfolio will run until a new optimization is calculated): 5, 10, 21, 42, 126, 252.

- Choice of number of time-series cross-validation n_{cv}: 1 to 8.

- Choice of cross-validation size t_{cv} as a percentage of the training size: 50%, 75%, 100%, 125%, 150%.

The λ's tested are always in the range of $[0, 3]$, with a step of 0.10. Before running the simulations, we tested the range in which the optimal λ's are found, and limit the search until 3 given that the optimal is always below this threshold.

Each of 11,200 combinations in the simulation runs from January 1st 2000 to December 31st 2022, totalizing 8.4 million calculations. As a result, we have 11,200 testing portfolios, each with 23 years of data.

The simulations provide answer to questions such:aa

- How does the optimal λ change with the number of assets?

- How does the optimal training size T changes as a function of number of assets and testing size t?

- What is the optimal number of time-series cross-validation (n_{cv})?

- What is the optimal size of the cross-validation (t_{cv})?

- How does the distribution of the optimal λ's change with the number of assets?

- How do moments of financial stress affect the previous questions?

The testing portfolios for each of the combinations are aggregated yearly, allowing for comparison of Sharpe Ratio, volatility and returns between the portfolios. The results are analyzed using multivariate ANOVA and visualizations.

3 Data

The data is collected from the Kenneth French Website. The website provides daily industry returns in tables dividing the industries in 5, 10, 12, 17, 30, 38, 48, 49 assets from 1926 to 2024 (updated regularly). Each of those tables is used as one of the possible n (number of assets) in the portfolio. The risk-free rate is also provided in the website, inside the "Fama/French 3 Factors" table, and subtract the returns.

4 How to add Citations and a References List

[Markowitz, 1952] or [Sharpe, 1964] or [Bruder et al., 2013]

References

[Bruder et al., 2013] Bruder, B., Gaussel, N., Richard, J.-C., and Roncalli, T. (2013). Regularization of portfolio allocation.

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

[Sharpe, 1964] Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442.