Task 2

Abstract

In this paper we discuss the foundations of Markowitz's Modern Portfolio Theory (MPT), the relevant mathematical equations describing it and contemporary approaches to asset allocation. We build a Global Minimum Variance Portfolio (GMVP) based on MPT with a comprehensive dataset containing daily returns of 48 industries from 1926 to 2017. The computation of the variance and weights of this portfolio is discussed, considering both long and short positions. Subsequently, Ridge and Lasso regularisation schemes are applied when minimising the portfolio's variance, to better understand the effects of varying regularization parameters. Finally, GMVP's performance is analysed in an unregularized case where short selling is prohibited.

1. Introduction

Portfolio construction was formulated in a rigorous manner for the first time by Harry Markowitz in his 1952 thesis essay, "Portfolio Selection".(1) A predecessor to this work, John Burr Williams's 1938 paper, introduced the present value model but did not fully address risk and the concept of asset allocation.(2) Before this, the view was that the markets exhibited strictly stochastic behaviour (akin to placing a bet in a casino), rendering a descriptive mathematical theory of investment useless, without any quantifiable way of defining risk.

Harry Markowitz is known through his thesis essay and extensive subsequent work as the founder of Modern Portfolio Theory (MPT), for which he received the 1990 Nobel Prize in Economics. The Global Minimum Variance Portfolio (GMVP) is built upon MPT and is the portfolio with the lowest possible variance (far left side in Figure 1). GMVP considers optimising the weights (fractional amount of total invested capital) of each asset according to a mean-variance framework. The concept of diversification, which aims to add instruments to the portfolio which aren't perfectly correlated with any current holdings, stems from the findings of the GMVP. A more detailed formulation is provided below in the Methodology section.

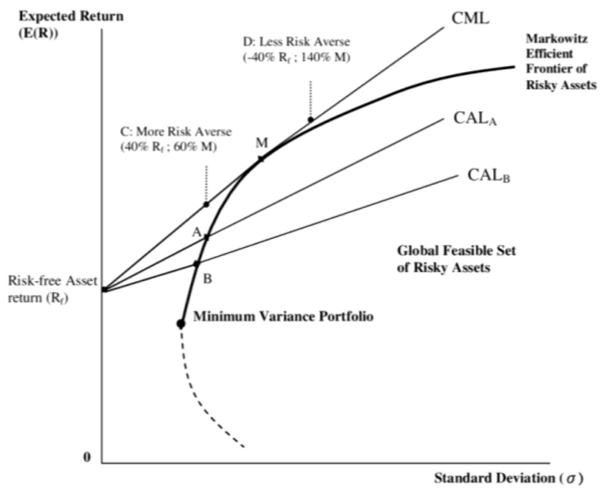

Figure 1: Markowitz Efficient Frontier (3)

The main drawdown of MPT is that it relies heavily on accurate estimations of the covariance matrix of returns as we will see further on. MPT rests on the assumption that the joint distribution of the return of all assets held in the portfolio is a multivariate normal, which was proven to be inaccurate in a real word setting. That is because, according to many academics, including Mandelbrot, the returns show a power law distribution.(4) Another famous solution to the portfolio allocation problem is the Black-Litterman model which uses a probabilistic approach in which the investor's views of the expected returns are compared with current market expectations (Implied Equilibrium Returns) to provide a new distribution. This mitigates the sensitivity observed in the mean-variance portfolio according to Best and Grauer, in which small increases in the expected return of a single asset forces significant reduction of the weights of other assets.(5)

Another issue in MPT is the need to base the analysis on historical data. More recent work in the asset allocation field involves alternative methods to quantify risk and diversification instead of looking at the expected returns. As the variance is considered the direct estimation of risk in MPT, this fails to make a difference between large possible losses and smaller, more "desirable" losses. Alternative measures, such as Value at Risk (VaR) and

Conditional VaR (CVaR) have become instrumental in asset allocation and risk management recently in the industry, to manage these short comings of MPT. A more modern portfolio construction approach is the risk parity model which focuses on allocating risk rather than capital and aims to provide an equal risk contribution from each asset held in the portfolio.(6)

Nevertheless, MPT is an incredible advancement and a stepping stone on which all modern asset allocation methods have been based on. In this paper we focus solely on showcasing the building of GMVP through the mean-variance framework considering broad historical data of 48 industries. The GMVP is constructed across different time frames and an analysis of possible regularisation methods is applied to assess the portfolio's performance in testing. Finally, the GMVP is constrained to a long-only strategy and the differences between the initial unregularized portfolio and this new implementation are discussed.

2. Methodology

2.1. GVMP

To provide a mathematical formulation of GMVP, let us define $\vec{R}$ to be a multivariate random variable (vector), signifying the returns of n assets, and $\vec{w}$ the vector associated with their weights (fractional amount of total invested capital):

$$\begin{cases} \vec{R} = (R_1, R_2, \dots, R_n)^T \\ \vec{w} = (w_1, w_2, \dots, w_n)^T \end{cases}$$

Then we can infer that:

$$E[R] = \sum_{i=1}^{n} w_i E[R_i]$$

By denoting the covariance matrix of the n assets returns as:

$$\Sigma = \begin{bmatrix} K_{11} & \cdots & K_{1n} \\ \vdots & \ddots & \vdots \\ K_{n1} & \cdots & K_{nn} \end{bmatrix}$$

Then the variance of the portfolio written in vector format is:

$$\sigma_P^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j K_{ij} = \vec{w}^T \Sigma \, \vec{w}$$

Minimising the variance of the portfolio is the task at hand for computing the GMVP, subject to the constraint that all weights sum to 1 (total fractional capital of the portfolio):

$$\min_{\mathrm{w}} \{\vec{w}^T \Sigma \, \vec{w}\} \; s.t. \sum_{i=1}^{n} w_i = 1$$

The weights are also considered to be $w_i \in (-1, 1)$ or $w_i \in (0, 1) \; \forall i = \overline{1, n}$ for a ban on short selling as no leverage is considered besides the initial capital. In this unregularized case, the following Lagrangian multiplier can be employed to minimise the variance:

$$\mathcal{L}(\mathrm{w}, \lambda) = \vec{w}^T \Sigma \, \vec{w} - \lambda \left(\sum_{i=1}^{n} w_i - 1 \right)$$

Which provides the solution:

$$\boldsymbol{w}^* = \frac{\Sigma^{-1} \, \vec{1}}{\vec{1}^T \, \Sigma^{-1} \, 1}$$

In this paper, considering both the regularised and unregularised cases and different constraints, we will use the minimise function provided by the SciPy library in Python.(7) The Sequential Least SQuares Programming optimizer (SLSQP) is employed in this minimisation problem with constrains.

2.2. Data Preparation

From the provided file, 48_Industry_Portfolios_daily.csv, some of the industries don't have associated returns until later years, as most, such as Soda and Health, didn't exist in the first half of the 20th century. For this reason, data was selected for analysis starting from the 1st of July 1969, to provide an equal view of all 48 industries. The dataset was checked for NaN values, and none were found after this date.

The equally weighted dataset was selected in favour of the market cap weighted one to enable minimum bias. MPT doesn't factor in the market cap of respective industries, therefore market cap weighted returns would provide higher weights for more developed industries, which also do change in time (for example the development of the health sector).

2.3. Analysis

Firstly, a function which returns the variance ($\vec{w}^T \Sigma \, \vec{w}$) was minimised using the $\sum_{i=1}^{n} w_i = 1$ constraint for both the long-short portfolio and long only variant. The weights were first computed on a period from 2013 to 2015, to showcase the differences between the 2 portfolios which is highlighted in the Results and Discussion section.

Afterwards, the in-sample risk (training variance) and out-of-sample risk (testing variance) were analysed on different lengths of time series. This was done considering:

$$\frac{p}{N} = \frac{number\ of\ features}{number\ of\ samples}$$

The covariance matrix is highly sensitive to the estimates of the returns as was discussed in the Introduction. Since we have 48 assets in our portfolio, it is important to test different lengths of the return time series, with daily inputs, to understand how the variance progresses when computing the optimal weights on a training dataset of a varying time series. For example, a $\frac{p}{N} = 0.5$ ratio would require $N = \frac{48}{0.5} = 96\ days$. The testing time series was taken to be the same length as the training one every time to ensure a consistent measure of out-of-sample risk. Various ratios ranging from 0.1 to 1 were tested, as we will see in the following plots, to better understand the generalisation of the model with the covariance matrix computed from limited data (ill-conditioned). It is important to note that although covariance matrices calculated from large amounts of time series data might provide a more general model, if a shorter window of investment is considered, more biased models could provide better outcomes.

Cross-validation was applied both in a singular format (one training set followed by one testing set) and in a time-series specific approach, by sequentially advancing the training set/testing set window over time and averaging the computed variances.

2.4. Regularisation

When the covariance matrix is estimated on a large $\frac{p}{N}$ ratio dataset, we will see that the model becomes prone to overfitting in a sense. The weights estimated in training do not generalise well in testing and produce increasingly larger variances as the ratio becomes closer to 1. Consequently, classic regularisation techniques, which aim to minimise the weights and offer a wider diversification of the model can be employed.

2.4.1. Ridge Regularisation

One common regularisation technique is by adding a Ridge penalising term to the variance which offers the following optimisation problem:

$$\min_{\mathrm{w}} \left\{ \vec{w}^T \Sigma \, \vec{w} + \lambda \sum_{i=1}^{n} w_i^2 \right\} \ s.t. \sum_{i=1}^{n} w_i = 1$$

This approach is derived from the Ridge regression, which aims to solve overfitting in a regression format. The Ridge parameter λ provides a more uniform decrease of all weights, being efficient in lowering estimation errors (variances). This makes the portfolio less sensitive to the length of the time series and offers a solution for the ill-conditioned covariance matrix. The solution to the equation above is:

$$w_i^* = \frac{\sum_{j=1}^{n} \left(\Sigma + \lambda \vec{1} \right)^{-1}}{\sum_{k=1}^{n} \sum_{j=1}^{n} \left(\Sigma + \lambda \vec{1} \right)^{-1}}$$

2.4.2. Lasso Regularisation

Lasso regularisation adds a penalty term which represents the sum of the absolute values of the weights:

$$\min_{\mathrm{w}} \left\{ \vec{w}^T \Sigma \, \vec{w} + \lambda \sum_{i=1}^{n} |w_i| \right\} \ s.t. \sum_{i=1}^{n} w_i = 1$$

This penalty term induces a more "brutal" reduction of the weights. Instead of uniformly shrinking the weight distribution like the Ridge term, Lasso cuts large weights to zero directly. Lasso regression is therefore beneficial when it comes to feature selection, in this case reducing the amount of assets held directly to ensure minimum variance.

A combination of both Lasso and Ridge terms is called Elastic Net and aims to combine both penalising factors in a convex manner. Consequently, elastic net preserves the feature selection characteristic of Lasso, while offering a smoother transition of the general weights to zero by tuning λ.

3. Results and Discussion

Pictured below are the computed variances of the unregularised GMVP in training and testing for varying lengths of the training time series ($\frac{p}{N}$ ratio). As mentioned before the training and testing time series follow each other sequentially:

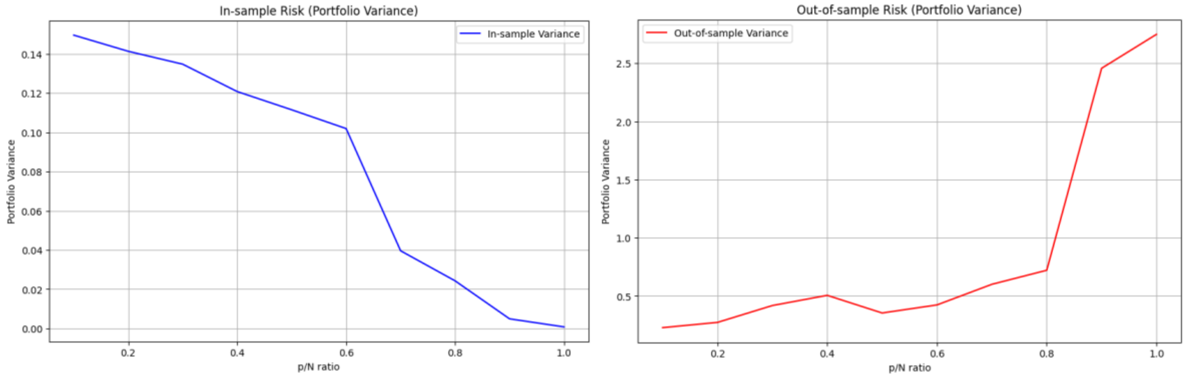

Figure 2: Training (left) and Testing (right) Variance for unregularised GMVP

We can see this is similar to the overfitting case of a regression. The training portfolio goes to zero with a decreasing number of elements in the returns time series, while the opposite is happening in testing, reaching 2.5 in the most ill-conditioned case.

Applying Ridge and Lasso regularisation parameters to the optimisation problem, changes the variance in the testing set widely. Below we have a view of how different λs, $\lambda \in \{0, 0.1, 2\}$, affect the testing variance for both Ridge (left) and Lasso (right) regularisations:

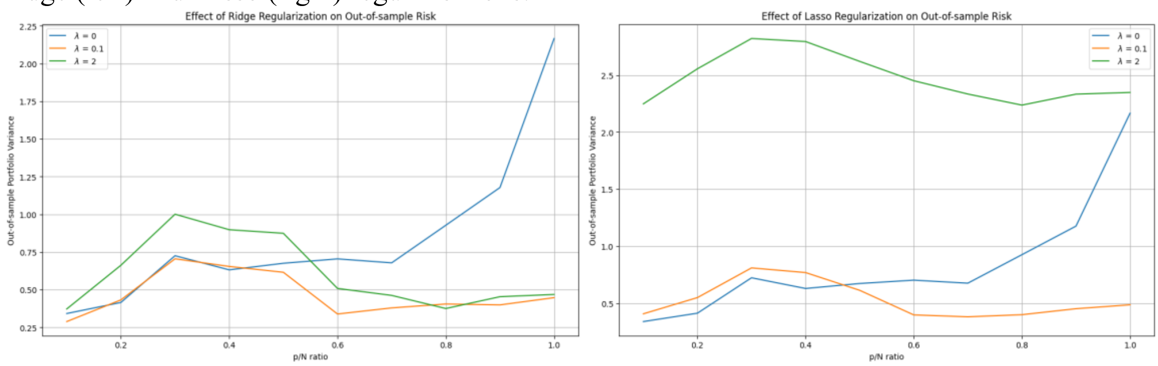

Figure 3: Ridge Regularisation (left) and Lasso Regularisation (right) for different λ values

Ridge parameters of 2 and 0.1 both help in decreasing the variance in the testing set, while a Lasso parameter of 2 produces very high variance values, showcasing its more aggressive weight reduction behaviour. For both Lasso and Ridge we can see effective reduction in variances. Cross-validation, as described in the Methodology section, helps compute the optimum λ for both types of regularisations:

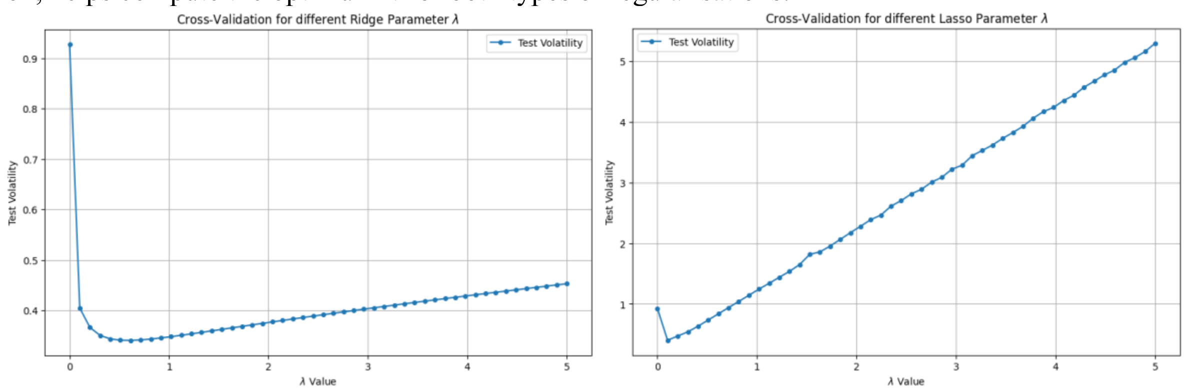

Figure 4: Test Variance at different λ values for Ridge (left) and Lasso (right)

	Ridge and Lasso optimisation with data from 2012-2015	
	Single training and testing set	Average λ across multiple sequential datasets
Ridge λ	0.61	0.09
Lasso λ	0.10	0.03
Ridge Variance	0.34	0.38
Lasso Variance	0.40	0.38

Figure 5: Table with λ and variances for Lasso and Ridge optimisation

As we can see from the non-zero parameters, lower variances are obtained from regularised portfolios. As a reminder, the averaged values represent a range of variances of different datasets recorded for the same λ. An average value of the variance is computed from these multiple datasets and the λ with the lowest associated average variance is selected. Both Ridge and Lasso show lower λs for the averaged set because they indicate a preference for simpler models. In a sense, the average over multiple datasets has the same effect as selecting a long time series of returns on which to compute the covariance matrix. It provides the model with greater generalisation and consequently less need for regularisation. With multiple datasets, the lower λ proves that the model becomes less sensitive to fluctuations, the same as with a longer time series. The variance also seems to converge to around 0.38 for both Ridge and Lasso for multiple datasets in the analysed period.

Finally, the weights of the unregularised GMVP are compared for the long-short and long-only $\left(w_i \in (0,1)\ \forall i = \overline{1,n}\right)$:

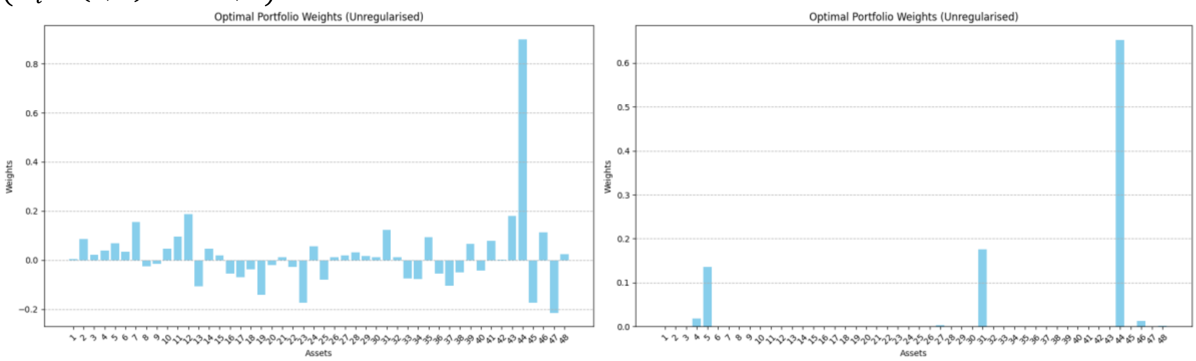

Figure 6: Optimal GMVP weights for long-short (left) and long-only (right) for 2013-2015

Both portfolios place a very high weight on the Banking sector (column 44), which indeed has always shown high returns due to its financial nature. Besides this, it becomes apparent that diversification seen in the portfolio which has both short and long positions is lost in the one constrained by a long only strategy. Consequently, banning short selling promotes few higher returning industries (Banking, Utilities and Smoking) to take up all the investment capital. This high concentration on a few sectors is given by the inability to hedge through short positions which help diversify the risk. The minimum variance of the short-selling-banned portfolio for the same time period is also almost double that of the unconstrained one, 0.35 compared to 0.18. This is indeed a direct consequence of less diversification.

In conclusion, we have constructed a GMVP for a dataset containing 48 industries and their respective returns. The GMVP was computed for the unregularised case and the regularised case, which has implemented Ridge and Lasso penalty terms in the optimisation problem. Different lengths of the time series were analysed for the unregularised case, in which overfitting was clearly present as the $\frac{p}{N}$ ratio got closer to 1. Finally, a ban on short selling was also applied and the portfolio was compared to the unconstrained one to notice the effects of less diversification.

References:

Task 1:

1) Ellis W. Tallman and Elmus R. Wicker, Federal Reserve of Cleveland, Banking and Financial Crises in United States History: What Guidance Can History Offer Policymakers?, 2010
2) Caccioli, F., Barucca, P. and Kobayashi, T. (2017) 'Network models of Financial Systemic Risk: A Review', *Journal of Computational Social Science*, 1(1), pp. 81–114. doi:10.1007/s42001-017-0008-3.
3) Lagasio, V. *et al.* (2022) 'Assessing bank default determinants via machine learning', *Information Sciences*, 618, pp. 87–97. doi:10.1016/j.ins.2022.10.128.
4) Tsai, C.-F., Hsu, Y.-F. and Yen, D.C. (2014) 'A comparative study of classifier ensembles for bankruptcy prediction', *Applied Soft Computing*, 24, pp. 977–984. doi:10.1016/j.asoc.2014.08.047.
5) Li, H. and Wu, W. (2024) 'Loan default predictability with explainable machine learning', *Finance Research Letters*, 60, p. 104867. doi:10.1016/j.frl.2023.104867.
6) Omer Fadl Elssied, N., Ibrahim, O. and Hamza Osman, A. (2014) 'A novel feature selection based on one-way ANOVA F-test for e-mail spam classification', *Research Journal of Applied Sciences, Engineering and Technology*, 7(3), pp. 625–638. doi:10.19026/rjaset.7.299.
7) *SMOTE - Version 0.12.2*. Available at: https://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.SMOTE.html
8) Breiman, L. (2001) *Machine Learning*, 45(1), pp. 5–32. doi:10.1023/a:1010933404324.
9) James, G. *et al.*, *An introduction to statistical learning: With applications in R*.

Task 2:

1) Markowitz, H. (1952) 'Portfolio selection*', *The Journal of Finance*, 7(1), pp. 77–91. doi:10.1111/j.1540-6261.1952.tb01525.x.
2) Burr, J.W. (2014) *The theory of investment value*. Miami: BN publ.
3) Hodnett, K. and Hsieh, H.-H. (2012) 'Capital market theories: Market efficiency versus investor prospects', *International Business & Economics Research Journal (IBER)*, 11(8), p. 849. doi:10.19030/iber.v11i8.7163.
4) Mandelbrot, B.B. (2008) '"New methods of statistical economics," revisited: Short versus Long Tails and gaussian versus power-law distributions', *Complexity*, 14(3), pp. 55–65. doi:10.1002/cplx.20264.
5) Best, M.J. and Grauer, R.R. (1991) 'On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results', *Review of Financial Studies*, 4(2), pp. 315–342. doi:10.1093/rfs/4.2.315.
6) Risk Parity Portfolio, Convex Optimization, The Hong Kong University of Science and Technology (HKUST) , Fall 2020-21
7) *Scipy.optimize.minimize#* (no date) *scipy.optimize.minimize - SciPy v1.12.0 Manual*. Available at: https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html