Question 2

Regularization Techniques for the Global Minimum Variance Portfolio

In this question we examine the global minimum variance portfolio using two different regularization methods: Ridge (ℓ_2) and LASSO[1] (ℓ_1). We use cross-validation methods to ensure robustness of the results and to identify the optimal parameters associated with each penalty term. We then compare the resulting variances and portfolio composition for each of the regularization techniques. In addition, we consider the variance of an evenly weighted portfolio for each of the test sets. Finally, we discuss what this means in a financial context.

This report is structured as follows: section one provides a brief introduction; section two covers the methodology and approach; section three presents the results; section four provides a discussion; and section five closes with concluding remarks and opportunities for further work.

1. Introduction

In modern day finance risk plays a critical role. A clear understanding of the exposure relating to an individual asset or portfolio allows organisations to make informed decisions. There are a numerous metrics that can be calculated to provide insight, each with their own strengths and weaknesses. One of the most fundamental metrics is standard deviation, used synonymously with volatility. As a result, its square, or variance, is also a measure of risk and is the focus of this question.

The Markowitz portfolio optimisation problem was first introduced in the early-1950's [4]. It concerns finding the optimal weights of a portfolio of assets that minimises its variance. We focused on the simplified case, only considering the budget constraint,

$$\min_{\boldsymbol{w}}\{\boldsymbol{w}^T \boldsymbol{C}\, \boldsymbol{w}\} \quad s.t. \quad \sum_{i=1}^{p} w_i = 1.$$

Where $\boldsymbol{w}$ is a vector of portfolio weights, $\boldsymbol{C}$ the covariance matrix, and p the number of assets in the portfolio. In general, the true covariance matrix is not known, hence, we were required to use the empirical covariance matrix computed from the dataset.

2. Methodology

2.1 Data

The dataset contained returns for fifty (unknown) stocks from the FTSE100 index. There were 500 rows of data corresponding to a time series of daily returns. It is well known that a trading year comprises of 252 observations, hence, for the purposes of this report we treated the dataset as essentially two years' worth of returns for the specified assets. This allowed us to split the data into natural financial groups such as months (consisting of ~21 days) and quarters (consisting of ~63 days).

To improve the robustness of the outputs, traditional machine learning techniques utilise cross-validation. As opposed to simply splitting the data into training and test sets once, cross-validation performs this separation multiple times and generates multiple models using each of these pairs. One popular approach is k-fold cross-validation which performs the train/test split a fixed number, k, times with each iteration using a different subset to train and test the model.

Unfortunately, however, this technique does not work with time-series data. Randomly selecting different subsets of the data has little meaning if we used next year's data to estimate last year's portfolio variance. For our financial time-series, we were forced to take an alternative approach to validation. In this report we used a sequential approach, that is iteratively moving the testing set forward and increasing the training set as the amount of 'historical' data increases.

As mentioned, our dataset contains approximately two years of data. Considering the number of assets, p, is relatively large, we did not want our initial training set to be too small. Similarly, we did not want each test set to be too small. As a result, we chose to split the dataset into blocks that represented quarters, with the first training window composed of the first two quarters, making the first half-year. This is summarised in Table 1.

Training Set		Test Set	
1	Q1 – Q2	1	Q3
2	Q1 – Q3	2	Q4
3	Q1 – Q4	3	Q5
4	Q1 – Q5	4	Q6
5	Q1 – Q6	5	Q7
6	Q1 – Q7	6	Q8

Table 1: The sequential train/test split used. It is described by the individual quarters over the two-year time horizon.

[1] Least Absolute Shrinkage and Selection Operator

This approach had several benefits. Firstly, it increased the robustness of the outputs by performing an appropriate amount of validation without becoming burdensome or time-consuming, whilst each set contained enough datapoints not to introduce a conflict with the number of assets being considered. In addition, it made intuitive sense and could easily be interpreted. For each regularization technique, we can imagine a scenario in which our portfolio required quarterly rebalancing to the global minimum variance portfolio. Naturally, one would want to use the historical data available to make that decision, and that is precisely what we achieved.

2.2 Regularization

As with most models, finding the right balance between over and underfitting was difficult. Overfitting may lead to excellent performance in the training dataset, we may even be able to obtain a riskless or substantially low level of portfolio variance, but this is likely to be followed by high out-of-sample variance when the model is applied to the test set. Whereas underfitting may lead to overly simple models that do not capture some of the context specific, and important, information.

Regularization is a technique that aims to strike a balance between these trade-offs by introducing a penalty term. There are plenty of proposed penalty terms discussed in the literature [5] with varying levels of complexity and convexity – with the latter impacting the efficiency and our ability to solve. We explored two of the simpler cases: Ridge and LASSO regularization.

2.2.1 Ridge Regularization [5]

The concept of Ridge regularization is to use the ℓ_2 norm as a penalty for the objective function. We can modify the minimisation problem from Section 1 to introduce the penalty term.

$$\min_{\boldsymbol{w}} \left\{ \boldsymbol{w}^T \boldsymbol{C} \boldsymbol{w} + \lambda \sum_i w_i^2 \right\} \quad s.t. \quad \sum_{i=1}^{p} w_i = 1.$$

Where we can rewrite the objective function as,

$$\{\boldsymbol{w}^T (\boldsymbol{C} + \lambda \boldsymbol{I}) \boldsymbol{w}\}.$$

This form allowed us to find an analytical solution, just as with the traditional minimisation problem.

In this setting λ is our regularization parameter which determines the strength of the penalty. This led to a natural question: What is the most appropriate value for lambda? Whilst, this cannot be observed directly, it can be found using validation techniques. To compute the optimal value of lambda we used a grid search method over the range $[0,10]$ with step sizes $\alpha = 0.01$. We then solved the previous equation using the different values of lambda, computed the minimum variance and subsequently identified the best value of lambda.

2.2.2 LASSO Regularization [5]

The concept of LASSO regularization is analogous with that of Ridge regularization but using the ℓ_1 norm as the penalty term,

$$\min_{\boldsymbol{w}} \left\{ \boldsymbol{w}^T \boldsymbol{C} \boldsymbol{w} + \lambda \sum_i |w_i| \right\} \quad s.t. \quad \sum_{i=1}^{p} w_i = 1.$$

Unfortunately, with the introduction of the absolute value we no longer had a closed form solution of the minimum. This meant we were required to use optimisation software. To minimise our objective function, we used the *fmincon(.)* function provided by MATLAB. This is a constrained optimisation solver that allows for the input of inequality, equality, and boundary constraints. In our setting we only required one equality constraint, outlined in the introduction, which corresponds to the budget constraint.

Once again, the question arose regarding the best value for lambda. We used the same grid search method as our Ridge implementation, with the domain $[0,10]$, but with larger step sizes $\alpha = 0.1$, to reduce the computational effort required.

As we will see, the portfolio suggested by LASSO is sparse. To account for the tolerances within the optimisation software, we took the decision to set any positions strictly less than 1% to zero. This marginally impacted the total sum of the weights, which must sum to one. For all test sets the implementation of our condition produced weights equal to 1 ± 0.02, which was sufficiently close for our purposes.

2.3 Evenly weighted portfolio

To benchmark our regularization approaches we also considered the evenly weighted portfolio. That is, a portfolio equally weighted in all constituent assets. In our setting the weights were each set at *2%*. We computed the variance of this portfolio for each test set and examined it alongside our other results.

3. Results

The main results of our analysis are presented in Tables 2–3. We also include Figures 1–2 and 4–5, which show the in-sample and out-of-sample risk for varying lambdas for *one* train and test set for both Ridge and LASSO regularization as an illustration. Each training and test set has its own associated graphs (available upon request). We also include Figures 3 and 6, which give an example of the distribution of portfolio weights for both Ridge and LASSO regularization.

Test Set	λ_{opt}	Minimum Variance σ^2	Evenly weighted Variance
1	2.95	0.3187	0.4413
2	0.44	0.7636	1.2502
3	0.38	0.3322	0.7714
4	2.64	0.1811	0.4025
5	0.00	0.4231	0.7343
6	6.31	0.5468	0.6786
Avg. λ_{opt}	2.12		
Avg. Minimum Variance		0.4276	0.7131

Table 2: Empirical results for **Ridge** regularization. λ_{opt} was computed using grid search. Minimum and evenly weighted variances were computed using empirical covariance and minimised portfolio weights.

Test Set	λ_{opt}	Minimum Variance σ^2	Evenly weighted Variance
1	2.30	0.2884	0.4413
2	0.50	0.7551	1.2502
3	0.20	0.3532	0.7714
4	7.70	0.1872	0.4025
5	0.00	0.4231	0.7343
6	9.80	0.6375	0.6786
Avg. λ_{opt}	3.42		
Avg. Minimum Variance		0.4407	0.7131

Table 3: Empirical results for **LASSO** regularization. Computational methods as described in Table 2.

Firstly, we can see that whilst the exact values of the lambdas are slightly different for the Ridge and LASSO regularization – indeed, for test sets 4 and 6 the values are noticeably higher for LASSO – the profile of the lambdas across the test sets 1–6 is almost identical.

We cannot infer much information from this in the context of the minimum variance portfolio, as each test set clearly produces different variances depending on factors present in that specific time window. However, it does provide us with some confidence that our two methods behave in similar ways and reduces the chance we incorrectly implemented either method.

Furthermore, we notice from Tables 2 and 3, that our average minimum variance across all the test sets is similar at approximately 0.43 and 0.44 respectively. Again, this provides us with confidence our results are, at least to some degree, valid.

One clear outcome from our results is that both the Ridge and LASSO regularized portfolios are far superior to the evenly weighted portfolio. This is obvious as the variance returned by the equally weighted portfolio is greater in every test set and the overall average.

Using Figures 1–2 and 4–5 we can explore the relationship between lambdas and in- and out-of-sample risk. The x-axis, lambda, indicates the level to which our portfolio construction has been constrained, with smaller values for lambda indicating less constrained. As we would expect the optimum performance in-sample is with no regularization. This follows as we can always expect to find a model that fits the training data well (if not perfectly). Consequently, in Figures 1 and 4, we see a monotonically increasing line for the in-sample risk. This, however, is not useful in practice and so we must examine Figures 2 and 5 showing the out-of-sample risk. Here we see steep descent followed by a slow ascent for increasing lambda, indicating there is a minimum at the specific lambda listed in our results.

It is worth noting that our closed form solution for Ridge regularization is what produces the smooth curve for both in- and out-of-sample risks. Whereas the jagged lines for LASSO follow because we have used optimisation software.

4. Discussion

Let us first discuss the superior results of our regularization methods over and above the evenly weighted portfolio. This may appear a straightforward benefit but does, in fact, have significant meaning.

An equally weighted portfolio (without good cause) is referred to as 'naïve diversification' [6] and is a problem that is seen quite frequently in the financial world. It is often seen in Defined Contribution Pension Plans, in which unsophisticated investors decide to place an equal amount of their capital in all funds without understanding the constituent components of each fund and how they interact – thereby diversifying, or more often, magnifying their risk exposure. Our results show that there is a significant benefit to introducing regularization into the asset allocation process.

Another intriguing outcome of our analysis is the high variance in test set 2 regardless of the method. The minimum variance achieved is, in some cases, two or three times that of other test sets. We can speculate that two interlinked elements were likely the cause of this: (1) there was a period of high volatility in Q4 of our dataset; and (2) the portfolio construction, trained using the previous three quarters, was ill-equipped to handle such volatility.

We now turn our attention to the optimum portfolio weights indicated by the optimal value of our regularization parameter, lambda. A new set of weights will be proposed for each test set, so we display the weights for test set 4 for illustrative purposes. There are two important differences between the weights proposed by Ridge regularization and those of LASSO.

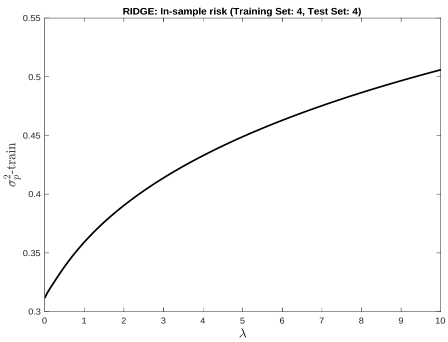

Fig 1: **Ridge** – In-sample risk for test set 4. The x-axis shows the regularization parameter lambda. Higher values of lambda indicate greater constraint.

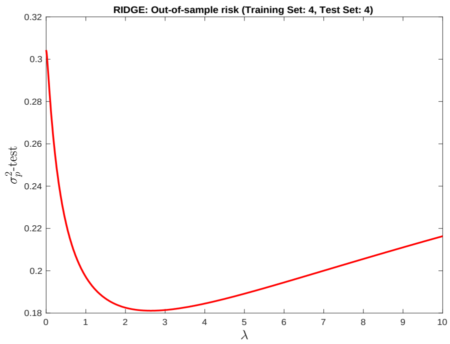

Fig 2: **Ridge** – Out-of-sample risk for test set 4. Each test set produces a similar graph. This test set is shown for illustrative purposes.

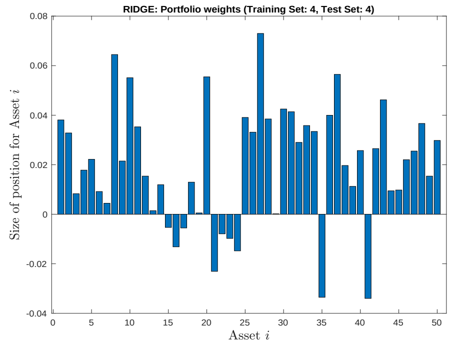

Fig 3: **Ridge** – Illustration of the optimal portfolio weights for test set 4.

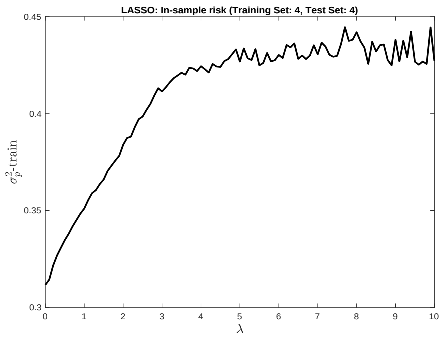

Fig 4: **LASSO** – In-sample risk for test set 4.

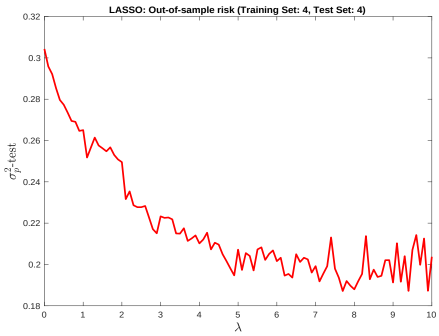

Fig 5: **LASSO** – Out-of-sample risk for test set 4.

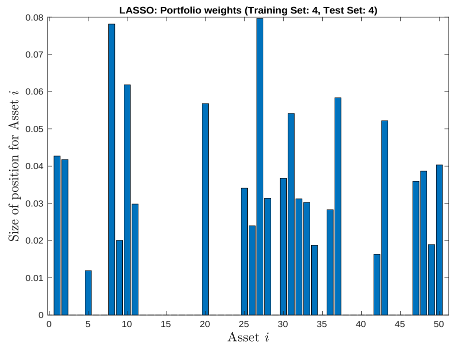

Fig 6: **LASSO** – Illustration of the optimal portfolio weights for test set 4.

Firstly, we notice the sparsity of the LASSO portfolio. One possible interpretation is the LASSO technique tends to focus on those positions which add significant levels of risk reduction to the portfolio [7]. Those positions that only produce a marginal reduction in the level of risk are discarded.

On the other hand, the Ridge portfolio take a position in almost every asset in the portfolio. This could be an explanation for the marginally higher minimum variance displayed by LASSO since we would naturally expect fewer positions to increase the level of risk, ceteris paribus. The sparsity of the portfolio could have contextual benefits though. In the real world rebalancing a portfolio is not free and an increased number of assets held could increase the amount transaction costs at each rebalancing date. Furthermore, depending on the investor, sales could trigger tax liabilities.

The second difference is the inclusion of short positions in the Ridge portfolio, whilst the LASSO portfolio in general has very few (or none for test set 4). This is simply an extension of our previous discussion as the increased number of positions suggested by Ridge must somehow be balanced to comply with the budget constraint. By taking short positions, it allows for greater investment in more assets – otherwise we would tend towards the equally weighted portfolio. Short positions also have the benefit of directly reducing portfolio variance. If two assets are perfectly correlated then taking an equal long and short position means the investor is perfectly hedged, and consequently has no risk exposure. This is clear evidence for the substantial reduction in variance over the equally weighted portfolio.

5. Conclusion

In conclusion, we have used two different regularization techniques to establish the minimum variance portfolio for our dataset. We have seen both methods produce similar values, clearly beating the equally weighted portfolio, but have starkly different portfolio constructions. In addition, the portfolio variances change considerably over time, as we include more historical data into our training set. This highlights the challenges faced in modern day finance and shows there is no perfect solution to problem.

As an extension of this work, one could generalise these results to the traditional minimum variance portfolio which includes a constraint on returns, rather than just on the budget. One could explore different penalty terms and extend the comparisons presented in this question. Alternatively, on a local level one could repeat this analysis using different optimisation software to understand whether there are improved minima.