Stockholms universitet

Lasso optimal portfolio: Construction and properties

Danie Viitanen Hellström

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2024:6
Matematisk statistik
Maj 2024

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Bachelor Thesis **2024:6**
http://www.math.su.se

Lasso optimal portfolio: Construction and properties

Danie Viitanen Hellström*

May 2024

Abstract

This thesis investigates the construction of an optimal portfolio using Lasso, Ridge and Elastic net regularization methods. Contrary to expections, portfolios constructed with Ridge regression underperforms across all evaluation metrics when compared to the market, despite their complexity involving the whole sample of assets while also including negative asset weights. In contrast, portfolios constructed using Lasso and Elastic net regularization show a different pattern. These portfolios are characterized by their sparsity, sometimes containing only a single asset. Notably, there exists a clear inverse relationship between the number of assets and the evaluation metrics, with larger portfolios yielding inferior results. While the allure of these sparse portfolios is evident, they also come with a significantly higher risk, both in the terms of volatility, measured by standard deviation, and the lack of diversification. These findings shed light on the nuanced trade-offs inherent in portfolio construction and underscore the importance of risk management in investment strategies.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: danie.hellstrom@gmail.com. Supervisor: Taras Bodnas Jan Olov Persson.

Acknowledgments

First and foremost, I would like to thank my partner, Malin Ekström, for her unwavering support and understanding throughout the process that has been this thesis. Without it, this thesis would not have been possible. I would also like to thank my supervisor, Taras Bodnar, for his invaluable inputs and encouragements. Lastly I would like to thank ChatGPT for helping me with the grammar, LaTeX-code and recommending relevent literature.

Contents

1 Introduction

Constructing an optimal portfolio has, and still is, the persuit of investors seeking to maximizing portfolio returns while managing the risk. The endeavor, however, is full of challenges. Traditional approaches such as the modern portfolio thoery (MPT), pioneered by Harry Markowitz [1], provides valuable insight into portfolio construction by focusing on mean-variance optimization and the concept of the efficient frontier. Markowitz laid the foundation for finding the portfolio with the highest expected return for a given level of risk and introduced the notion of the tangency portfolio, the portfolio on the efficient frontier with the highest Sharpe ratio.

The Sharpe ratio developed by William F. Sharpe [2] serves as a fundamental way to measure the risk-adjusted return, helping investors in assessing their portfolios relative to the risk suffered. The financial markets evolve and complexities arise however, and traditional methods often face limitations in capturing the intricate dynamics of asset returns.

To address these challenges Britten-Jones introduced a regression-based framework for portfolio construction [7]. By utilizing the ordinary least squares (OLS) regression with a constant vector of ones as the response variable, Britten-Jones showed a link between the regression estimation of the coefficients and the tangency portfolio, offering a practical way of constructing optimal portfolios.

Building on Britten-Jones foundation, this thesis explores the application of regularization methods to build optimal portfolios, namely the Least absolute shrinkage and selection operator (Lasso), Ridge and Elastic net regression. Regularization methods, originating from the field of statistical learning [8], offer a systematic approach to address overfitting, deal with highly correlated parameters and increase model interpretability by adding penalties to the regression coefficients.

The primary focus of this thesis is to investigate the application of Lasso regression in the context of portfolio optimization, but expanded to include both Ridge and Elastic net regression as well. By adding regularization to Britten-Jones framework the aim is to enhance the robustness and efficiency of the constructed portfolios.

Evaluation metrics play an important role in assessing the performance of constructed portfolios. In addition to the Sharpe ratio, this thesis also considers key metrics such as the Alpha, Beta and Return of Investment (ROI) of portfolios, as well as the expected return and the risk to provide a holistic evaluation of portfolio strategies. The constructed portfolios will be compared to the Standard and Poor's 500 market index.

This thesis begins with a quick rundown of the data used in chapter 2 alongside the methods of acquiring it. This is followed by chapter 3 where some background to the portfolio optimization problem is given. In chapter 3 we also cover the evaluation metrics we will use to test the constructed portfolios. In chapter 4 we present the Britten-Jones regression setup aswell as regularization methods and cross-validation. In chapter 5 we step by step go through the implementation of the methods used followed by the results in chapter 6. We end this thesis with a discussion about the results in chapter 7.

2 Data

The data used in this thesis consist of stock prices over the period from January 2014 to March 2024. Instead of using the daily stock price at closing or opening time, the adjusted stock prices are used. The reason for this is that the adjusted stock price take dividends, stock splits and new stock offerings into account, so the adjusted price give a more accurate price of the stock. The adjusted stock prices have been aggregated into simple weekly return denoted R_w and is calculated as

$$R_w = \frac{P_w - P_{w-1}}{P_{w-1}}$$

where P_w is the adjusted stock price at the end of week w and P_{w-1} is the the adjusted stock price at the end of week $w - 1$.

The stocks chosen for the dataset consist of the stocks that make up the Standard & Poor's 500 (SP500) index. Some stocks have been excluded since they did not have data during the previously mentioned period. The adjusted stock price data were acquired from Yahoo Finance[1] via the tidyquant[2] package in R.

By choosing weekly returns over daily returns we remove some of the noise of the stock market and instead aim to better capture trends in the market over daily fluctuations.

One could then argue monthly data would be even better, the reason for weekly over monthly is simply that using monthly data would require us to stretch out the testing period, which would lead to the exclusion of more stocks due to missing values.

3 Background

3.1 Modern portfolio theory

Modern portfolio theory (MPT) is also known as mean-variance optimization or Markowitz optimization and is the foundation for modern investment strategies. The theory was developt my Markowitz, H (1952)[1] and it focuses on optimizing the trade-off between the risk and return. In his

[1] https://finance.yahoo.com/
[2] https://cran.r-project.org/web/packages/tidyquant/index.html

paper Markowitz makes the assumption that an investor is rational and risk-averse, meaning the investor seeks high return while keeping the risk low. Markowitz also stresses the importance of diversification by spreading investments across assets with various risk-return characteristics.

3.1.1 Mean-variance optimization

The mean-variance optimization, or Markowitz optimization aims to find the optimal set of weights for a portfolio so that it maximizes the expected return for a chosen level of risk, or minimize the risk for a chosen level of return. The optimal weights for a portfolio is found by estimating the weight vector w so that

$$\hat{w} = \text{argmin}_w\{w^T C w\}$$

where C is the covariance matrix and w^T the transposed weight vector. The portfolios return can then be calculated as $\mu_p = w^T\mu$ where μ is the asset return vector. This optimization assumes that $\sum_i w_i = 1$, meaning that an investor is assumed to invest all of his/her wealth. There is also the restriction that $w_i \geq 0$ restricting an investor from short-selling assets. Repeating this process for a range of different values of μ_p will lead to a set of optimal portfolios that represent the efficient frontier.

3.1.2 Optimal portfolios, the efficient frontier and the tangency portfolio

A set of portfolios with a certain level of expected return may have varying risk, the optimal portfolio is defined as the portfolio with that level of expected return that has the lowest risk. Repeating this processes for different levels of expected return results in a set of optimal portfolios, these portfolios create the efficient frontier. In the efficient frontier the tangency portfolio stands out as the portfolio with the highest risk-adjusted return calculated by the Sharpe ratio.

3.1.3 Sharpe ratio

The Sharpe ratio was introduced by William Sharpe (1966)[2] and in Sharpe (1994)[3] he officially accepted the name Sharpe ratio for his ratio.

The Sharpe ratio is a way to measure the risk-adjused return for a portfolio, calculating the Sharpe ratio for a portfolio is done as

$$SR = \frac{E[R_p] - R_f}{\sqrt{Var(R_p)}}$$

where SR is the Sharpe ratio for the portfolio, R_p is the return for portfolio and R_f is the risk-free rate.

The expected return of the portfolio, $E[R_p]$ is estimated as the mean of the historical returns, so that

$$\bar{R}_p = \frac{1}{N}\sum_t R_t$$

where R_t is the return of the portfolio at time t and N is the total number of assets. The variance $Var(R_p)$ of the portfolio is estimated as

$$\sum_t \frac{1}{N-1}(R_t - \bar{R}_p)^2.$$

3.1.4 Evaluation metrics

To be able to effectively evaluate the performance of a constructed portfolio we need some evaluation metrics. In this thesis we will evaluate a portfolio by the portfolios *Alpha* and *Beta*, its Sharpe ratio and the return of investment (ROI). We will also look at the portfolios expected return and risk, which we define as the mean and standard deviation of the portfolio return.

The *Alpha* of the portfolio compares the portfolios performance against a market index Elton, Gruber, Brown & Goetzmann (2014)[5], it is essentially the portfolios ability to beat the market. When calculating a portfolios α, both the market return, the risk-free rate and the portfolios β needs to be considered as follows:

$$\alpha = R_p - (R_f + \beta(R_m - R_f))$$

where R_p is the portoflio return and R_m is the market index, R_f is the risk-free rate and β is the beta of the portfolio defined below.

The β of the portfolio instead compares the portfolios volatility against a market index. We define the β of the portfolio as

$$\beta = \frac{\text{Cov}(R_p, R_m)}{\text{Var}(R_m)}$$

where R_p is the return of the portfolio and R_m is the return of the market index. For the covariance, $\text{Cov}(R_p, R_m)$, of the portfolio and the market we simply calculate

$$\frac{1}{N-1}\sum_t \Big((R_{p,t} - \bar{R}_p) \cdot (R_{m,t} - \bar{R}_m)\Big).$$

If a portfolio have a β value larger than 1 it is to be interpreted that the portfolio is more volatile than the market, while a value lower means less volatile. The market index of choice for both α and β is naturally the SP500.

The return of investment or ROI of a portfolio speaks for itself, it is the percentage return of an investment. For a portfolio this will be calculated as

$$\text{ROI} = \frac{V_N - V_0}{V_0} \cdot 100$$

where V_N is the ending value of the portfolio and V_0 is the initial value of the portfolio. Multiplying by a factor of 100 leads to the ROI being expressed in percentages.

4 Methods

4.1 Britten-Jones regression

This thesis takes inspiration of the regression setup from Britten-Jones(1999)[7] where he sets up a regression with a constant vector of ones as response variable and does not include an intercept, so that

$$\mathbf{1} = X\theta + u$$

where X is the design matrix and contain the asset returns, θ the coefficient vector and u the error term. He goes on to show that if the coefficient vector θ is estimated with ordinary least squares so that

$$\hat{\theta} = (X^T X)^{-1} X^T 1,$$

it is the unscaled tangency portfolio. Britten-Jones shows this by using the updating formula[3] for inverse matrices and showing that

$$\frac{\hat{\theta}}{\mathbf{1}^T \hat{\theta}} = \frac{\bar{\Sigma}^{-1}\bar{x}}{\mathbf{1}^T \bar{\Sigma}^{-1}\bar{x}}$$

where the right side is the estimated tangency portfolio with $\bar{\Sigma}$ being the sample covariance matrix and $\bar{x}$ is the sample mean vector.

4.2 Regularization methods

In this thesis three regularization methods will be introduced and tested. The ℓ_2 regularization called Ridge regularization, the ℓ_1 regularization called Lasso regularization and a combination between the two called Elastic net. The regularization methods modify the objective function in the optimization problem so that they include a penalty term that encourages a simpler model. In the context of linear regression where the objective is to estimate the coefficient vector θ so that it minimizes the sum of squared residuals, represented as

$$||y - X\theta||_2^2.$$

[3] For the full proof, see Britten-Jones (1999)[7] pp. 658.

Here y is the response variable, X the design matrix and θ the coefficient vector to be estimated. The regularization methods add a penalty term, $\lambda P(\theta)$, to the objective function. The λ is a hyperparameter that determines the strength of the regularization and is often estimated with cross-validation, as seen in section 4.3, and the function $P(\theta)$ is determined by the different regularization methods. The goal of regularization regression is hence to estimate

$$\hat{\theta} = \text{argmin}_{\theta}\left\{||y - X\theta||_2^2 + \lambda P(\theta)\right\}.$$

4.2.1 Least absolute shrinkage and selection operator, the Lasso

The least absolute shrinkage and selection operator, or the Lasso for short, was introduced by Robert Tibshirani[9] 1996 in his paper Regression Shrinkage and Selection via the Lasso. The Lasso is a ℓ_1 regularization method that adds the the penalizing function $\sum_i |\theta_i|$ to the objective function. The Lasso seeks to estimate the coefficient vector so that

$$\hat{\theta} = \text{argmin}_{\theta}\left\{||y - X\theta||_2^2 + \lambda||\theta||_1\right\}.$$

The penalty term added to the objective function penalizes the absolute value of the coefficients. This often leads to sparse models since penalizing term drives the coefficients of the less relevent parameters to zero, effectively performing parameter selection Hastie, Tibshirani & Wainwright (2015)[11].

4.2.2 Ridge regularization

Ridge regularization, or Tikhonov regularization is an ℓ_2 regularization. it is a method to prevent overfitting in statistical models James, Witten, Hastie & Tibshirani (2013)[12], especially in regression analysis and portfolio optimization. With the use of an added penalty term in form of the ℓ_2-norm Ridge will shrink coefficients toward zero, but not to zero as Lasso does with its ℓ_1-norm. With the ℓ_2-norm added as penalty, the Ridge regularization seeks estimate the coefficients so that:

$$\hat{\theta} = \text{argmin}_{\theta}\left\{||y - X\theta||_2^2 + \lambda||\theta||_2^2\right\}.$$

4.2.3 Elastic net

The Elastic net is a regularization method that combines both Ridge- and Lasso regularization Zou & Hastie (2005)[4]. The Lasso regularization tends to only select a small subset of variables, particualary in cases of multicollinearity. Consequently, the Elastic Net was developed to address this limitation and other drawback of the Lasso regularization. The Elastic net seeks to estimate the coefficients so that:

$$\hat{\theta} = \text{argmin}_{\theta}\Bigg\{||y - X\theta||_2^2 + \lambda_1||\theta||_2^2 + \lambda_2||\theta||_1\Bigg\}.$$

Just like the Lasso the Elastic net performs variable selection by setting some coefficients to zero. Due to its combination between the ℓ_1 and ℓ_2 regularization it is also more flexible when handling highly correlated predictor variables. With the combined penalty the Elastic net have better predictive performance when compared to either Lasso or Ridge alone, especially when the predictor variables are highly correlated Friedman, Hastie, Tibshirani (2010)[13].

4.3 K-fold cross-validation

Cross-validation is a widely used validation method for estimating performance of a model. K-fold cross-validation splits the data into K approximately equally sized sets. The model is then trained on $K - 1$ of the sets and evaluated on the one remaining set. The $K - 1$ sets that the model is trained upon are called the training sets and the remaining set that the model is evaluated upon is called the validation set. This process is repeated K times, so that every set is left out once. Some evaluation metric, such as the mean square error, is calculated for each validation set.

Cross-validation can be used to select a value for the hyperparameter λ in the previously mentioned regularization methods. During this process a range of λ values are tested to select the best suited one. The range is determined by λ_{max} which is chosen large enough so that all parameters are set to zero, resulting in the null model Tibshirani et. al. (2010)[13]. The λ_{min} is chosen based on the λ_{max}, so that $\lambda_{min} = \epsilon \cdot \lambda_{max}$ where ϵ is some small value Tibshirani, et. al. (2010) [13].

For every λ in the range, cross-validation is performed and the averaged mean square error is calculated for the λ. The value for λ deemed most suitable for the regularization method is the λ value that resulted in the lowest average mean square error.

5 Implementation

The data consist of the design matrix X, which consists of the explanatory variables. As explanatory variables, the weekly return for a set of stocks have been chosen. The stocks chosen make up the Standard and Poor's 500 index, but only the stocks that have data during the full testing period have been selected. The testing period begins at January 2014 and stretches to March 2024, these dates are arbitrarliy chosen. This results in 468 stocks which had data for the full period, and the period stretches 532 weeks.

The response vector y is inspired from Britten-Jones (1999)[7] so that it is a constant vector of ones. The length of the vector is 532 so that it matches the design matrix.

The goal is not to explain the stock prices but rather to predict them, hence we want to see how it performs out of sample. With this in mind we split the data into a training set and a evaluation set. We have chosen to keep 80% of the data in the training set and 20% of the data in the test set. The model will be trained on the period 2014-2022 and evaluated upon the period 2022-2024.

Both the Ridge- and Lasso regularization models will be trained and tested on the data set. The Elastic net will also be trained and tested, but since we will implement the Elastic net as

$$\hat{\theta} = \text{argmin}_\theta \left\{ ||y - X\theta||_2^2 + (1-\alpha)\lambda||\theta||_2^2 + \alpha\lambda||\theta||_1 \right\} \tag{1}$$

rather than

$$\hat{\theta} = \text{argmin}_\theta \left\{ ||y - X\theta||_2^2 + \lambda_1||\theta||_2^2 + \lambda_2||\theta||_1 \right\}$$

we will train the model with different values for α.

The implementation is done in R with the help of the glmnet[4]-package. Two functions from the package are used, the function cv.glmnet that will run the cross-validation algorithm and estimate the best values for the hyperparameter λ, and the function glmnet that will fit a model to the data for the chosen λ value. The cross-validation algortihm uses ten-fold cross-validation and will create a range of one hundred different λ values to iterate over. While the range of values for λ can be created manually, the function creates it as explained in chapter 4.3. In this thesis we use the range of λ values created by cv.glmnet. The cross-validation algorithm will calculate the average mean square error for each λ in the range and will save the one with the lowest average mean square error to later be used when fitting the model.

While the cross-validation algortihm itself is deterministic Tibshirani, et. al. (2010)[13], it contains some randomness when splitting the data. For this reason the cross-validation algorithm is arbitrarily called upon 1000 times.

When it is time to fit a model to the data, the glmnet function is called upon to estimate the coefficient vector. The glmnet function uses the same formula for Lasso, Ridge and Elastic net, namely formula (1). This means that if α is set to zero or one, a Ridge- or Lasso model is fitted respectively. Any other value for α will fit an Elastic net model. Since an α value in the range $(0, 0.5)$ will give the ℓ_2-norm stronger influence and an α in the range

[4]

$(0.5, 1)$ will give the ℓ_1-norm stronger influence we will try multiple values for α inbetween zero and one.

Due to the fact that an intercept is not included in the model, the coefficients need not sum to one, but as shown by Britten-Jones (1999) [7], scaling the coefficient vector so that it sums to one leads to the tangency portfolio.

The estimated and scaled coefficient vector are to be interpreted as portfolio weights and are such applied to the corresponding stocks. To evaluate the portfolios performance we will consider the evaluation metrics presented in chapter 3.1.4. These metrics will be compared against eachother as well as against the SP500 index.

The cross-validation algorithm will provide a total of 11000 values for λ, 1000 each for Ridge and Lasso and 9000 for Elastic net. But this will not lead to 11000 different portfolios, but rather some 200 different ones as many of the λ values will be identical.

6 Results

The constructed portfolios are to be compared to the Britten-Jones portfolio, the S&P 500 and the equal weighted portfolio. The weights for the equal weighted portfolio are simply calculated as $w_i = \frac{1}{N}$ where N is the total number of assets. The weights for the S&P 500 portfolio can be obtained using the tidyquant package. To calculate the weights for the Britten-Jones portfolio however, we need to implement the Moore-Penrose pseudoinverse [10] since the training data consists of only 426 data points while there are 468 parameters.

Another important consideration when examining the portfolio constructed using the Britten-Jones formula is that it include negative weights, indicating short selling stocks. It is noteworthy that a large amount of capital is invested in shorts. When applying the Britten-Jones formula to the training data, approximately 2900% of the portfolios value is allocated in short positions.

The constructed portfolios built by using Ridge-, Lasso and Elastic net regularization are somewhat disapointing. While some constructed portfolios have a high risk-adjusted return they include very few parameters, often suggesting that the portfolio only contain a single stock. The portfolios constructed that contains at least five or more stocks instead have a negative α value, indicating that they perform worse than the market index. In short the portfolios built by Ridge regularization tends to favor low risk and low return while the portfolios built by Elastic net and Lasso favors very sparse portfolios often resulting in few stocks and ending up with very high ROI and expected return but also high risk.

We will now go deeper into the resulting portfolios for the different models.

6.1 Ridge

The 1000 iterations of the cross-validation algorithm for Ridge regression resulted in 21 different values for λ. The values for λ ranges from between around 75 to around 190. A histogram of the distribution for λ values is shown in figure 1.

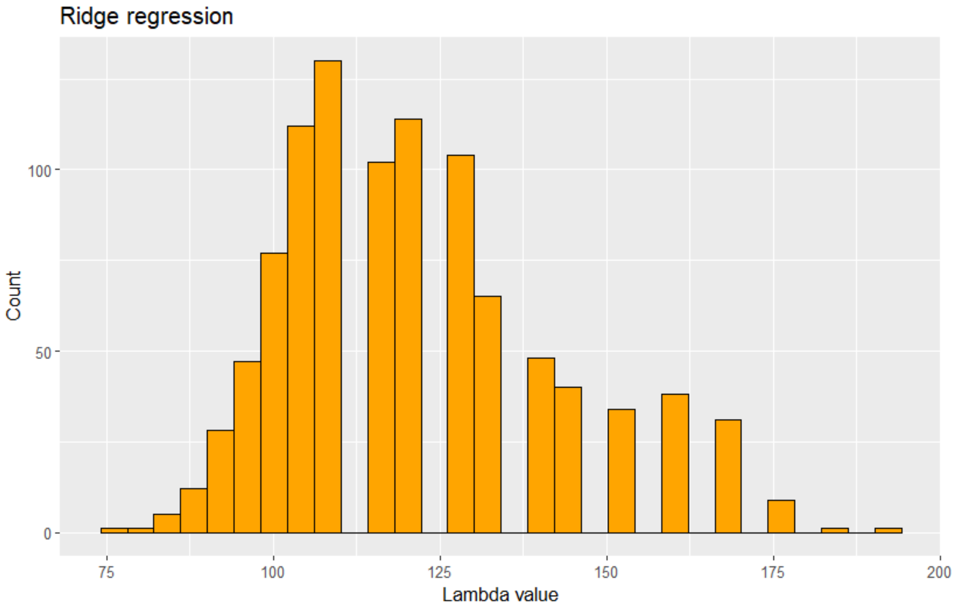

Figure 1: Histogram over λ values for Ridge regularization.

As expected, the portfolios constructed when fitting the different λ values to Ridge regularization contains every stock from the sample stocks in varying degrees, resulting in diverse porfolios. Due to the fact that Ridge does not shrink any coefficient to zero the resulting portfolios are of size 468 and are quite complex. The evaluation metrics are show in table 1.

Table 1: Ridge evaluation metrics together with SP500, Equal weights and Britten-Jones.

"Model"	"Sharpe ratio"	"Alpha"	"Beta"	"ROI"	"Expected return"	"Risk"	"Short percentage"	"Parameters"
"Equal"	0.077112	-0.205538	0.916144	20.004844	0.002001	0.025943	0	468
"SP500"	0.132795	0	1	40.172172	0.003504	0.026384	0	468
"Britten"	0.086302	-0.109499	0.093441	128.61194	0.006996	0.081069	2914.265	468
"Ridge 12"	0.077682	-0.208042	0.918282	19.858264	0.001966	0.025314	-0.1983	468
"Ridge 11"	0.077679	-0.208534	0.917613	19.842632	0.001965	0.025299	-0.1614	468
"Ridge 13"	0.077672	-0.207557	0.918986	19.870762	0.001968	0.025331	-0.1729	468
"Ridge 10"	0.077663	-0.209029	0.916977	19.824461	0.001964	0.025286	-0.1594	468
"Ridge 14"	0.077646	-0.207084	0.919726	19.879531	0.001968	0.025349	-0.2198	468
"Ridge 9"	0.077639	-0.209523	0.916373	19.804319	0.001962	0.025273	-0.2256	468
"Ridge 8"	0.077607	-0.210013	0.915797	19.782732	0.00196	0.025261	-0.2706	468
"Ridge 15"	0.077601	-0.206625	0.920506	19.884006	0.001969	0.025369	-0.3455	468
"Ridge 7"	0.077569	-0.210496	0.915248	19.760172	0.001959	0.025251	-0.2899	468
"Ridge 16"	0.077537	-0.206185	0.921325	19.883699	0.001969	0.02539	-0.6	468
"Ridge 6"	0.077527	-0.21097	0.914724	19.737051	0.001957	0.025241	-0.3043	468
"Ridge 5"	0.077482	-0.211432	0.914225	19.713719	0.001955	0.025232	-0.2485	468
"Ridge 17"	0.077452	-0.205767	0.922185	19.878249	0.001968	0.025413	-1.0407	468
"Ridge 4"	0.077436	-0.211881	0.913747	19.690468	0.001953	0.025223	-0.2702	468
"Ridge 3"	0.077388	-0.212315	0.91329	19.667534	0.001951	0.025216	-0.6832	468
"Ridge 18"	0.077344	-0.20537	0.923087	19.867485	0.001968	0.025438	-2.246	468
"Ridge 2"	0.077341	-0.212734	0.912853	19.645101	0.00195	0.025209	-1.8222	468
"Ridge 19"	0.077215	-0.204995	0.924031	19.851491	0.001966	0.025465	-4.2596	468
"Ridge 20"	0.077065	-0.204639	0.925014	19.830684	0.001965	0.025495	-7.9767	468
"Ridge 1"	0.076911	-0.213901	0.912787	19.515023	0.001939	0.025212	-1.6856	468
"Ridge 21"	0.076898	-0.204295	0.926036	19.805884	0.001963	0.025526	-8.5659	468

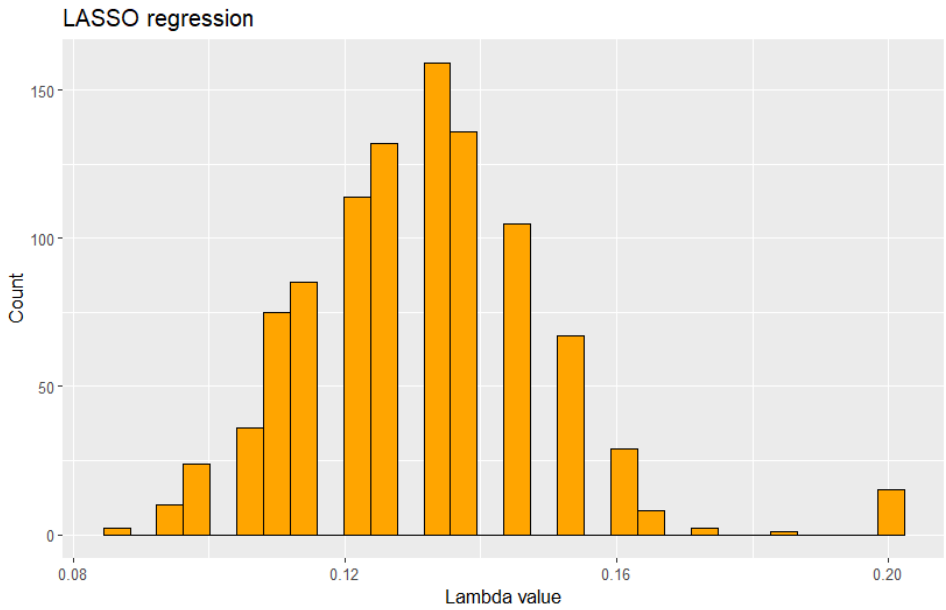

Figure 2: Histogram over λ values for Lasso regression.

In table 1 it can be seen that the different portfolios constructed by Ridge regularization underperforms when compared to both the S&P 500 index, the equal weighted portfolio as well as the Britten-Jones portfolio. All of the portfolios constructed using Ridge regression are somewhat similiar, not leaving much choice to the investor. All of the portfolios are quite close the S&P 500 and the equal weighted portfolio in terms of risk while having a significan lower expected return than the S&P 500. When comparing these portfolios to the equal weighted portfolios, it is hard to spot a difference, indicating one could just as well use a naively equal-weighed portfolio as constructing one using Ridge regression.

6.2 Lasso

One of the main reasons for Lasso regularization when construction portfolios are its parameter selection property. Constructing portfolios using Lasso often leads to sparse portfolios, which we will see is also the case in this thesis.

The 1000 iterations of the cross-validation algortihm provided 16 different values for λ whose distribution is shown in figure 2. Comparing the λ_{Lasso} with λ_{Ridge} one can see that the values in λ_{Lasso} are significantly smaller than the values in λ_{Ridge}. Low values for the hyperparameter λ would indicate weak regularization compared to large values for λ that indicates strong regularization. Fitting models with these values for λ a range of

16 different portfolios can be constructed. These portfolios are quite sparse, including as few as only 1 parameter up to 10 parameters. The metrics for these portfolios are shown in table 2 together with the metrics for the S&P 500, the equal weighted portfolio and the Britten-Jones portfolio.

Table 2: Lasso evaluation metrics together with SP500, Equal weights and Britten-Jones.

”Model”	”Sharpe ratio”	”Alpha”	”Beta”	”ROI”	”Expected return”	”Risk”	”Short percentage”	”Parameters”
”Equal”	0.077112	-0.205538	0.916144	20.004844	0.002001	0.025943	0	468
”SP500”	0.132795	0	1	40.172172	0.003504	0.026384	0	468
”Britten”	0.086302	-0.109499	0.093441	128.61194	0.006996	0.081069	2914.265	468
”Lasso 1”	0.205182	1.747447	2.044611	264.570987	0.014631	0.071309	0	1
”Lasso 2”	0.205182	1.747447	2.044611	264.570987	0.014631	0.071309	0	1
”Lasso 3”	0.20136	1.632626	2.00097	240.020248	0.013776	0.068413	0	2
”Lasso 4”	0.186943	1.317438	1.881174	178.356614	0.011427	0.061124	0	2
”Lasso 5”	0.177646	1.153496	1.813502	150.581556	0.010232	0.057595	0	3
”Lasso 6”	0.171657	1.009964	1.736417	131.048561	0.009274	0.054027	0	3
”Lasso 7”	0.167288	0.92144	1.688874	119.523648	0.008684	0.051908	0	3
”Lasso 8”	0.162839	0.82608	1.630157	108.435596	0.008085	0.04965	0	4
”Lasso 9”	0.156476	0.68127	1.529484	93.625419	0.007233	0.046225	0	4
”Lasso 10”	0.151338	0.582105	1.460472	83.917811	0.00665	0.043942	0	4
”Lasso 12”	0.151154	0.252552	1.154962	62.321024	0.005092	0.033688	0	10
”Lasso 14”	0.148266	0.304048	1.212394	64.570194	0.005287	0.035661	0	8
”Lasso 11”	0.147053	0.508531	1.409351	76.936709	0.006217	0.042278	0	4
”Lasso 16”	0.14583	0.332824	1.247799	65.437706	0.00538	0.036891	0	8
”Lasso 13”	0.143468	0.452316	1.370291	71.731623	0.005886	0.041029	0	4
”Lasso 17”	0.143297	0.367988	1.289818	66.591409	0.005499	0.038376	0	6
”Lasso 15”	0.142241	0.406752	1.331289	68.439408	0.005655	0.039755	0	5

Table 2 is ordered by decreasing Sharpe ratio with the reference portfolios added at the top. The observent reader can see that a low count of parameters in the portfolio seem to correspond to a higher Sharpe ratio and portfolio α.

While some of the portfolios constructed using Lasso regularization seem to outperform the market index based on their positive α values and β values below 1 while also having a high ROI and Sharpe ratio it is important to take into consideration that they have very limited diversification. These portfolios only consist of one to four different stocks and therefor have a high concentration. While these portfolios seem to outperform the index in the short term they deviate from one of the main principles of the modern portfolio theory, diversification Markowitz (1952)[1]. Investors should approach these portfolios with the utmost care.

The relationship between the risk and return for Lasso constructed portfolios seem almost linear, as seen in figure 3. In table 2 it would seem that the evaluation metrics get worse as the number of parameters increase, so we plot the relationship between the risk-adjusted return and the number of parameters in figure 4. There seem to be a significant drop in the risk-adjusted return from one to five parameters, but after that the risk-adjusted return seem to increase when the number of parameters increase.

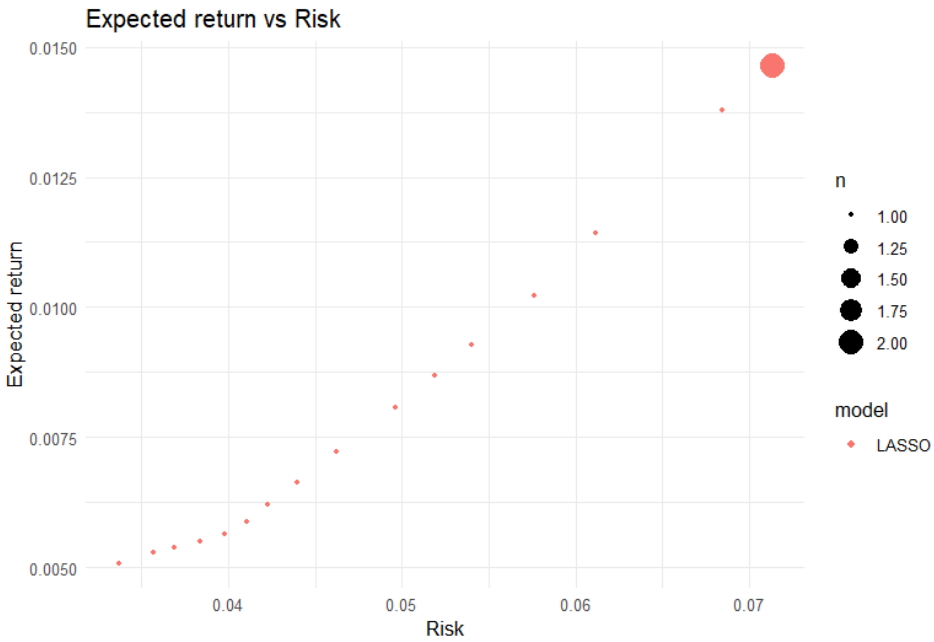

Figure 3: Expected return vs Risk for Lasso portfolios.

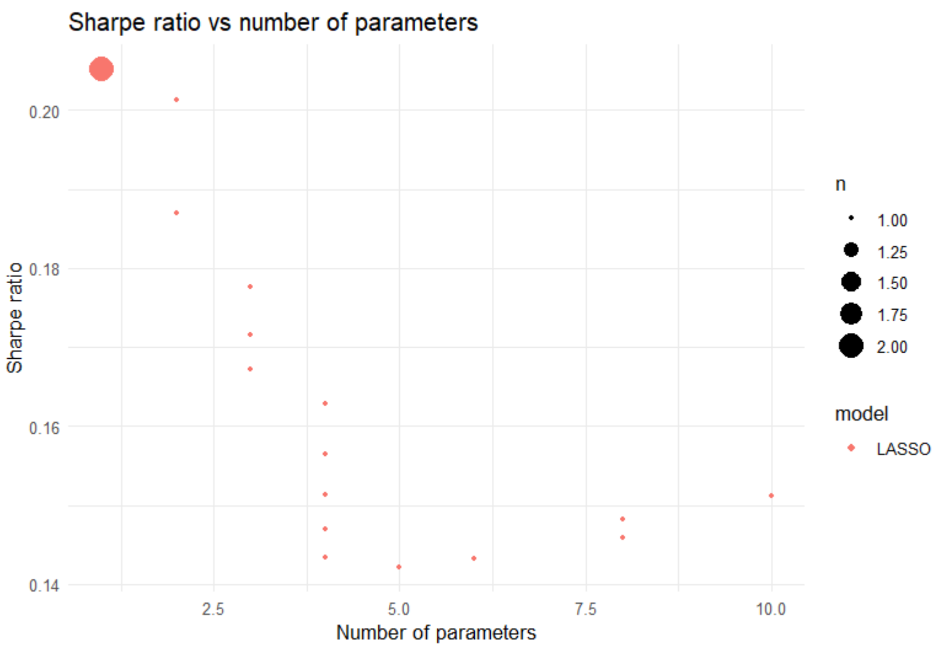

Figure 4: Sharpe ratio vs the number of parameters for Lasso portfolios.

6.3 Elastic net

Repeating the cross-validation 1000 times for each α in $[0.1, 0.2, \ldots, 0.9]$ the Elastic net regularization gets between 16 and 19 different values for λ depending on α. These values are shown in figure 5. Much like the values provided for the Lasso regularization the values for λ are quite small when compared to the ones provided for Ridge regularization, implying a somewhat weaker regularization.

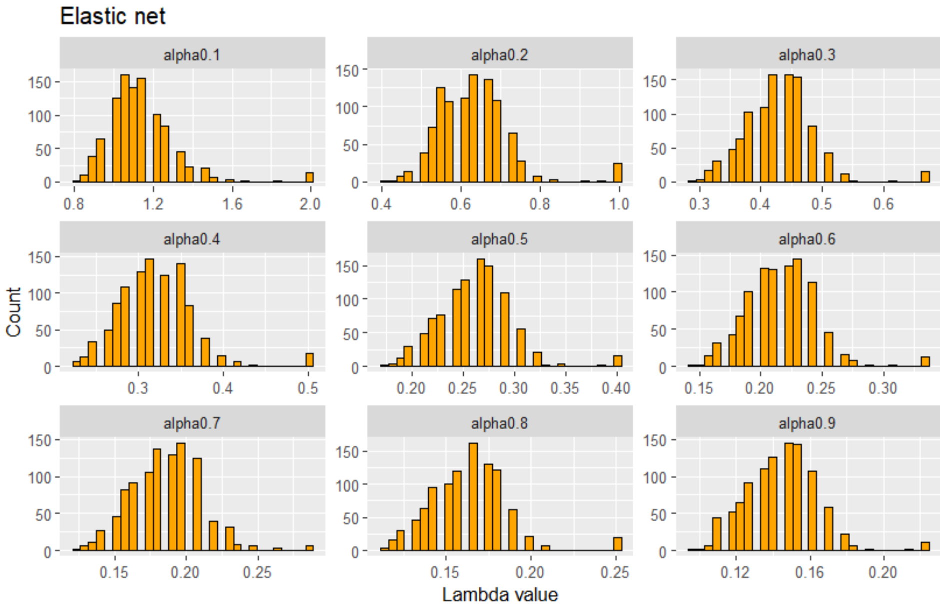

Figure 5: Histogram over λ values for Elastic net with different values for α.

This results in a total of 159 different values for λ and as such 159 different portfolios. These portfolios are quite similiar to the ones provided by Lasso regularization and their metrics are shown in table 3 with a market index added at the top. Table 3 does not show all of the 159 different portfolios. Since metrics seem to follow the same structure as for the Lasso metrics, the Elastic net metrics table has also been ordered by paramters, then the first and last ten is shown in 3.

Table 3: Elastic net evaluation metrics together with SP500, Equal weights and Britten-Jones.

"Model"	"Sharpe ratio"	"Alpha"	"Beta"	"ROI"	"Expected return"	"Risk"	"Short percentage"	"Parameters"
"Equal"	0.077112	-0.205538	0.916144	20.004844	0.002001	0.025943	0	468
"SP500"	0.132795	0	1	40.172172	0.003504	0.026384	0	468
"Britten"	0.086302	-0.109499	0.093441	128.61194	0.006996	0.081069	2914.265	468
"Elastic 1"	0.205182	1.747447	2.044611	264.570987	0.014631	0.071309	0	1
"Elastic 19"	0.205182	1.747447	2.044611	264.570987	0.014631	0.071309	0	1
"Elastic 20"	0.205182	1.747447	2.044611	264.570987	0.014631	0.071309	0	1
"Elastic 21"	0.205182	1.747447	2.044611	264.570987	0.014631	0.071309	0	1
"Elastic 38"	0.205182	1.747447	2.044611	264.570987	0.014631	0.071309	0	1
"Elastic 39"	0.205182	1.747447	2.044611	264.570987	0.014631	0.071309	0	1
"Elastic 55"	0.205182	1.747447	2.044611	264.570987	0.014631	0.071309	0	1
"Elastic 71"	0.205182	1.747447	2.044611	264.570987	0.014631	0.071309	0	1
"Elastic 72"	0.205182	1.747447	2.044611	264.570987	0.014631	0.071309	0	1
"Elastic 89"	0.205182	1.747447	2.044611	264.570987	0.014631	0.071309	0	1
"Elastic 67"	0.143162	0.409987	1.335875	68.715128	0.005662	0.039551	0	6
"Elastic 124"	0.143016	0.354089	1.279765	65.494645	0.005419	0.037893	0	7
"Elastic 105"	0.142851	0.386637	1.31203	67.34861	0.005563	0.038941	0	5
"Elastic 135"	0.142826	0.443184	1.365225	70.806757	0.005826	0.040793	0	4
"Elastic 122"	0.142571	0.392958	1.318391	67.655731	0.00559	0.03921	0	5
"Elastic 138"	0.142496	0.398724	1.323608	68.006916	0.005618	0.039426	0	5
"Elastic 119"	0.142397	0.437263	1.361942	70.207308	0.005787	0.040642	0	4
"Elastic 157"	0.142356	0.403082	1.327792	68.240636	0.005638	0.039604	0	5
"Elastic 85"	0.141907	0.420426	1.349488	68.862086	0.005692	0.040111	0	5
"Elastic 102"	0.141858	0.430031	1.357934	69.475204	0.00574	0.04046	0	4

Just as for the evaluation metrics calculated for the Lasso portfolios there are quite a few portfolios that outperform the market index, but just as for the Lasso portfolios these are not diversified portfolios and investors should approach them with care. Since the portfolios are quite similiar to the ones constructed using Lasso we will plot the risk-return relationship for the Elastic net together with the portfolios from Lasso in figure 6. We will also do the same thing for the risk-adjusted return and parameter relationship in figure 7. The similiarities between the portfolios constructed by Lasso and Elastic net can clearly be seen in the figures.

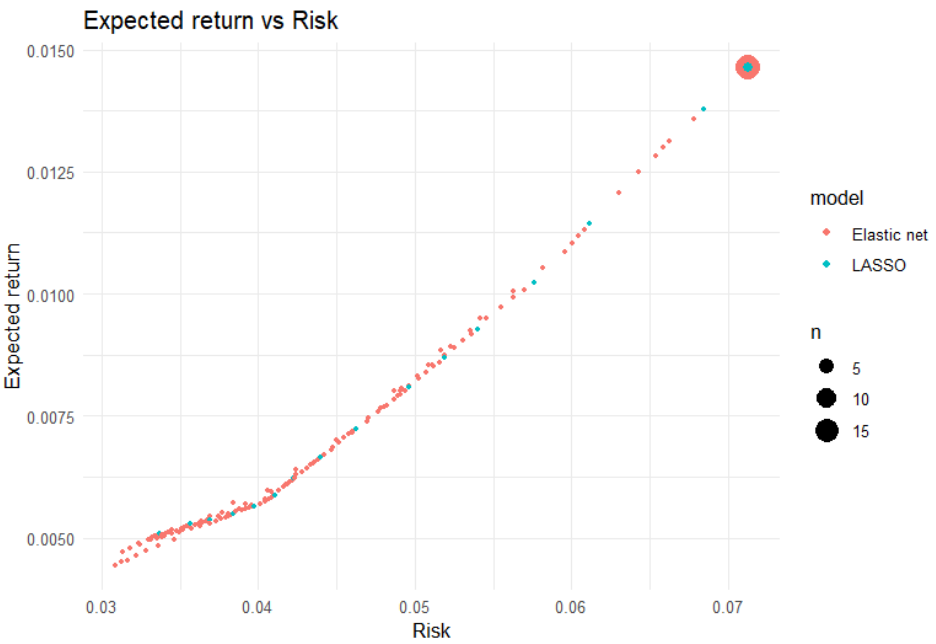

Figure 6: Expected return vs Risk for Elastic net and Lasso portfolios

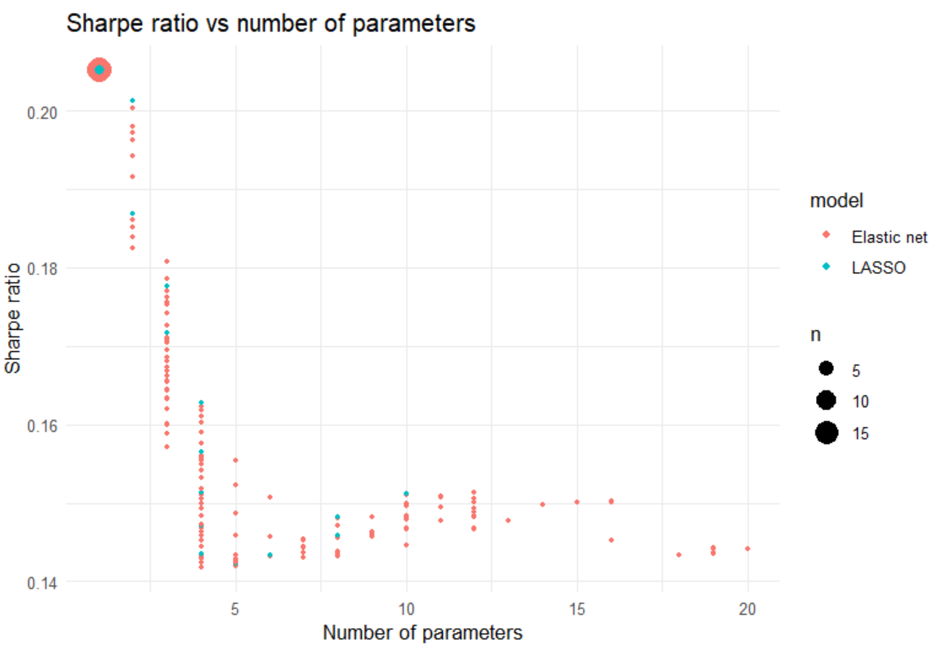

Figure 7: Sharpe ratio vs the number of parameters for Elastic net and Lasso portfolios.

7 Discussion

While the portfolios constructed using Ridge regression underperformed when compared to the market the portfolios constructed using Lasso and Elastic net regularization methods demonstrate promising performance when compared to the market index. For instance, the portfolio with the highest Sharpe ratio has a ratio 30% higher than the market while also showing favorable *Alpha* and *Beta* values along with nearly a 500% increase in the expected return. Figure 8 shows the growth of capital, which adds to the appeal of these portfolios.

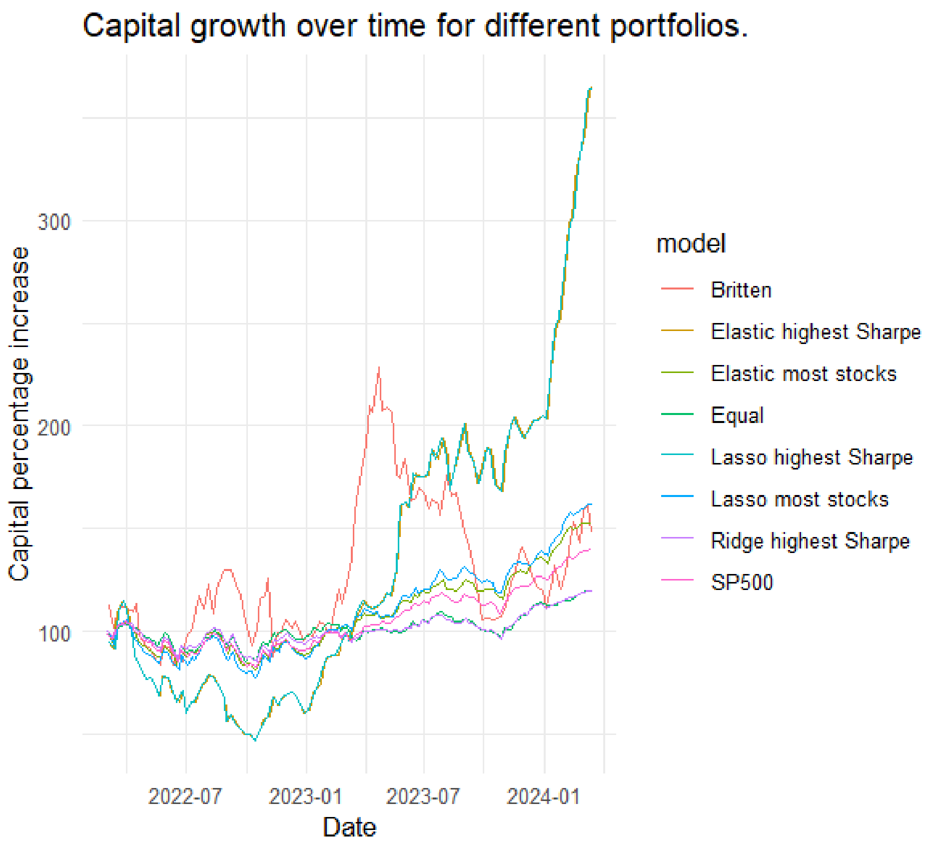

Figure 8: Capital growth for different portfolios together with market index.

It is however crucial to consider the elevated risks associated with these portfolios. They carry a significantly higher risk profile, around 400% greater than the market index. Additionally, their heavy reliance on single-stock allocations challenges one of the core principles of modern portfolio theory, diversification.

It is also of interest to take a closer look at the stock that make up these single stock portfolios and the stock in question is NVDA. The stock has made quite the journey during the sample period for this thesis and these single asset portfolios are easy to understand when looking at the development of NVDA as seen in figure 9.

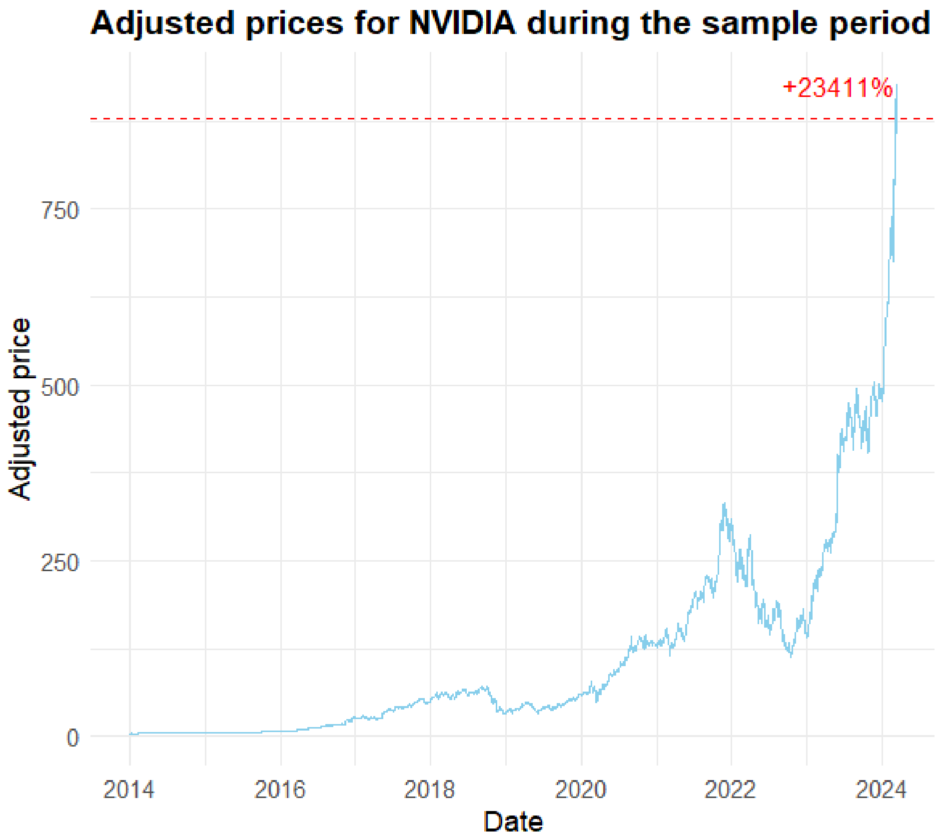

Figure 9: Growth of Nvidias (NVDAs) price during the sample period. The dashed red line indicate the percentage increase at the end of the sample period.

It will be seen that this stock is included in the constructed portfolios and it is easy to understand why seeing it has increased over 23,000% during the sample period.

There is some middle ground to explore however, the portfolios containing 8 to 20 different stocks strike a balance between sparsity and diversification. These portfolios still outperform the market while offering a more varied mix of assets.

The portfolios constructed using Lasso and Elastic net only included a range of 20 different stocks all together. In figure 10 the inclusion of these stocks in the portfolios constructed using Lasso and Elastic net is shown. In table 4 the companies behind the stocks are also shown.

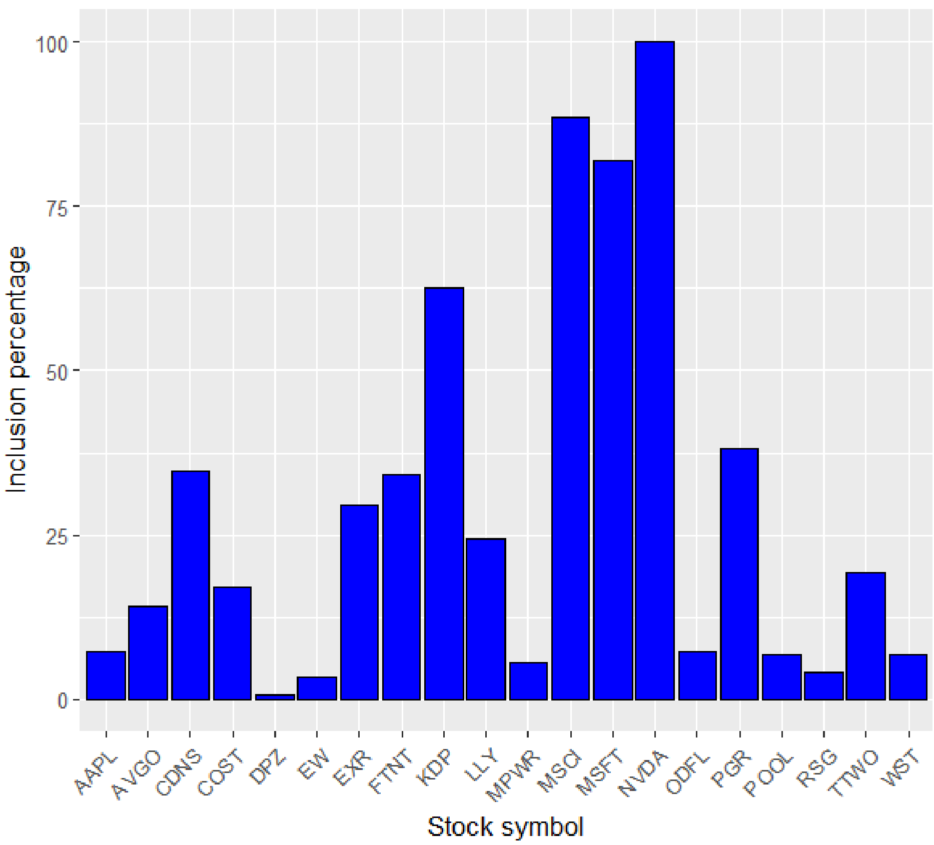

Figure 10: The twenty stocks comprising the Lasso and Elastic net portfolios and their inclusion percentage.

Table 4: Stock symbol with corresponding company.

"symbol"	"company"
"AAPL"	"APPLE INC"
"AVGO"	"BROADCOM INC"
"CDNS"	"CADENCE DESIGN SYS INC"
"COST"	"COSTCO WHOLESALE CORP"
"DPZ"	"DOMINO S PIZZA INC"
"EW"	"EDWARDS LIFESCIENCES CORP"
"EXR"	"EXTRA SPACE STORAGE INC"
"FTNT"	"FORTINET INC"
"KDP"	"KEURIG DR PEPPER INC"
"LLY"	"ELI LILLY + CO"
"MPWR"	"MONOLITHIC POWER SYSTEMS INC"
"MSCI"	"MSCI INC"
"MSFT"	"MICROSOFT CORP"
"NVDA"	"NVIDIA CORP"
"ODFL"	"OLD DOMINION FREIGHT LINE"
"PGR"	"PROGRESSIVE CORP"
"POOL"	"POOL CORP"
"RSG"	"REPUBLIC SERVICES INC"
"TTWO"	"TAKE TWO INTERACTIVE SOFTWRE"
"WST"	"WEST PHARMACEUTICAL SERVICES"

It is also worth noting that using other data frequencies such as daily returns, or expanding sample period give will give significantly different results. In table 5 we can see an example when using daily data instead of weekly. In the table portfolios have been constructed using Lasso with daily data and are compared to the S&P500, equal weighted and Britten-Jones portfolio. With the daily data it can be seen that some portfolios constructed using Lasso regression include negative weights, which none of the one constructed using weekly data did. These portfolios are also constructed with different assets than the ones constructed from weekly data.

Table 5: Daily data portfolio metrics.

"Model"	"Sharpe.ratio"	"Alpha"	"Beta"	"ROI"	"Expected.return"	"Risk"	"Short.percentage"
"Britten-Jones"	0.07378	0.5371	0.54404	155.11085	0.00233	0.03159	-1818.59131
"SP500"	0.0588	0	1	41.77323	0.00076	0.01287	0
"Lasso"	0.3194	2.86403	0.57062	2922.41163	0.00684	0.0214	-128.76561
"Lasso"	0.30209	2.62664	0.60356	2245.16577	0.00634	0.02097	-104.55242
"Lasso"	0.2828	2.40266	0.64421	1733.06172	0.00585	0.02069	-82.08018
"Lasso"	0.26277	2.16251	0.68291	1309.57139	0.00534	0.02032	-60.12822
"Lasso"	0.23993	1.86678	0.71073	935.39311	0.00473	0.01971	-36.0015
"Lasso"	0.21951	1.57026	0.72553	673.92771	0.00413	0.01884	-16.56245
"Lasso"	0.19697	1.31298	0.75224	495.27125	0.0036	0.01829	0
"Lasso"	0.1931	1.43183	0.78321	551.45283	0.0038	0.01967	0
"Lasso"	0.1897	1.57082	0.82356	619.09346	0.00402	0.02121	0

In table 6 another example is shown for an extended sample period, this period only extend 5 years so it is from 2009 to 2024. But the results are significantly different.

Table 6: Daily data portfolio metrics.

"Model"	"Sharpe.ratio"	"Alpha"	"Beta"	"ROI"	"Expected.return"	"Risk"	"Short.percentage"
"SP500"	0.14625	0	1	56.34725	0.00327	0.02238	0
"Britten"	0.21213	-1.07752	0.68905	972.45592	0.01909	0.08997	-2766.455
"Equal weight"	0.08569	-0.57956	0.96461	29.55971	0.00199	0.02317	0
"Lasso"	0.04638	0.59841	1.08046	32.42212	0.00134	0.02892	0
"Lasso"	0.04085	0.71771	1.09318	32.4078	0.0012	0.02944	0
"Lasso"	0.03913	0.74097	1.09583	32.50179	0.00116	0.02972	0
"Lasso"	0.0372	0.77436	1.09947	32.73091	0.00112	0.03007	0
"Lasso"	0.03488	0.83839	1.106	33.04665	0.00107	0.03056	0
"Lasso"	0.03397	1.4769	1.1624	35.7246	0.00119	0.03509	0
"Lasso"	0.03277	1.39031	1.15578	35.42787	0.0011	0.03352	0
"Lasso"	0.0326	0.99454	1.12085	33.50937	0.00102	0.03124	0
"Lasso"	0.02988	1.21189	1.14135	34.21989	0.00096	0.03228	0

The portfolios constructed with Lasso regression on the extended sample period underperforms when compared to the comparison portfolios. The reason for these varying results extends beyond the confines of this thesis.

In conclusion, using regularization methods together with a constant response vector shows promise in enhancing efficiency and performance. Further testing is needed to confirm their broader applicability and reliability.

References

[1] Markowitz, H. (1952) Portfolio Selection.
The Journal of Finance, Vol. 7, No. 1, pp. 77-91. https://www.jstor.org/stable/2975974

[2] Sharpe, W. (1966) Mutual Fund Performance The Journal of Business, Vol. 39, No 1, Par 2: Supplement on Security Prices (Jan., 1966), pp. 119-138. https://www.jstor.org/stable/2351741?seq=13

[3] Sharpe, W. (1994) The Sharpe Ratio. The Journal of Portfolio Management, 1994, 21, pp. 49-58. https://web.stanford.edu/~wfsharpe/art/sr/SR.htm

[4] Zou, H., Hastie, T. (2005) Regularization and Variable Selection via the Elastic Net. Journal of the Royal Statistical Society, Series B (Statistical Methodology) Vol. 67, No. 2, 2005, pp. 301-320.

[5] Elton, E., Gruber, M., Brown, S., Goetzmann, W. (2014) Modern Portfolio Theory and Investment Analysis, 9th Edition Wiley, Ninth Edition, Januari 2014.

[6] Brodie, j., Daubechies, I., De Mol, C., Giannone, D., Loris, I. (2009) Sparse and Stable Markowitz Portfolios. Proceedings of the National Academy of Sciences, 2009, 106(30).

[7] Britten-Jones, M.(April 1999)
The Sampling Error In Estimate sof Mean-Variance Efficient Portfolio Weights.
The Journal of Finance, Vol. 54, No.2, pp. 655-671.
https://www.jstor.org/stable/2697722

[8] hastie, T., Tibshirani, R., Friedman, J., (2017) The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer, Second Edition, Corrected 12th printing, Januari 2017.

[9] Tibshirani, R.(1996) Regression Shrinkage and Selection via the Lasso.
Journal of the Royal Statistical Society. Series B (Methodological), Vol. 58, No. 1, pp. 267-288.
https://www.jstor.org/stable/2346178

[10] Penrose, R. (1955) A Generalized Inverse of Matrices.
Mathematical Proceedings of the Cambridge Philosophical Society, Vol. 51, Issue 3, pp. 406-413.
https://doi.org/10.1017/S0305004100030401

[11] Hastie, t., Tibshirani, R., Wainwright, M. (2015) Statistical Learning with Sparsity: The Lasso and Generalization Chapman & Hall, First Edition, 2015.

[12] James, G., Witten, D., Hastie, T., Tibshirani, R. (2021) An Introductin to Statistical Learning with Applications in R Springer, Second Edition.

[13] Tibshirani, R., Friedman, J., Hastie, T. (2010)
Regularization Paths for Generalized Linear Models via Coordinate Descent.
Journal of Statistical Software, Vol. 33, Issue 1, pp. 1-22.
https://www.jstatsoft.org/article/view/v033i01