DBA3803
Predictive Analytics Project 1

Written Report

Portfolio Optimization
10 October 2023

A0173269W Darren Darius Tan Jia Wei
A0278460X Van Erven Oliver Daniel

Section 1: Introduction

Comprising 100 unique assets, this dataset classifies stocks based on two financial metrics: market capitalization (size) and book-to-market (B/M) ratios. Each asset represents a specific combination of size and B/M, resulting in a 10x10 grid of asset returns. By Markowitz Portfolio Theory, it suggests that each investor should aim to create a diversified portfolio of assets that have low or negative correlation with each other. As every asset has its associated risks from returns, it suggests that the assets risks are also highly correlated with each other. Risks are derived from the variance of the assets, and to minimize risks to create a diversified portfolio of assets, the Minimum Variance Portfolio (MVP) strategy is usually employed.

As risks of the assets are highly correlated in this dataset (See Appendix 3), MVP seeks to minimize the covariance through finding the optimal weights assigned to each asset. In turn, these same derived asset weights then multiplies with its corresponding asset returns, to give the best portfolio return, albeit it may not be the highest return because risks would have been minimized. Even if the assigned weights were to be altered such that risks are not minimized but reduced, or if the weights are altered to be all equal, the increase or decrease of portfolio returns remains indeterminate. In this project, the performance of the portfolio shall be determined by Sharpe Ratio.

Section 2: Objectives and Approach

The primary objective of this project is to reduce this portfolio risk by first minimizing risk/variance among the assets. This may introduce overfitting, which we then aim to prevent by exploring the 4 components of a prediction framework: **Data + Loss + Structure + Constraints** and modifying relevant components to explore the prediction model that gives the best Sharpe Ratio. The secondary objective of the project would then be to evaluate if the weights derived from minimizing the variance of the portfolio (minvar_portfolio) would give better asset return results if the weights were all equally assigned to each asset (EW portfolio)

Structure:

The structure of the prediction model has been determined by the big idea to establish a mathematical equivalence between the Optimization Expression (LHS) and the Linear Regression with Ordinary Least Squares (OLS) framework (RHS):

$$Rw = Rw_{EW} - RN\beta$$

Hence, minimizing the LHS is the same as minimizing the RHS below:

$$\text{LHS} = \text{Minimising } \|Rw\|_2^2 \text{ where}$$

$$\text{(1)} \quad w = w_{EW} - N\beta$$

$$\text{(2)} \quad R = w^T E$$

$$\text{RHS} = \text{Minimising } \|y - X\beta\|_2^2 \text{ where}$$

$$y = Rw_{EW}$$

$$X = RN$$

Based on RHS, the prediction problem can be expressed as a linear regression equation without an intercept, which minimizes $y - x\beta$, or the unexplained variance in the model's predictions. In notation form, the mathematical equivalence above translates to the idea that minimizing the weights of the demeaned returns in the LHS naturally gives the minimized loss function of the residuals in the linear regression given that can be solved using OLS in the RHS.

Data

The following data manipulation steps will therefore need to be taken to arrive at the OLS structure outlined in the RHS:

(1) Construct y and X for linear regression.
(2) Obtain 99 different β for the 100 assets.
(3) Convert each β to a portfolio of 100 w.
(4) Save the test performance of all the 100 w in an array.

Loss

In general, the train loss function applied in this problem is L2 Loss due to the use of OLS to arrive at our optimal portfolio weights. On the other hand, the Sharpe Ratio is used as the model performance evaluation metric (test loss). This is because it suits the business objective best by taking risk-adjusted returns, balancing both risks and returns, while its equation also takes into account variance, which mathematically accounts for L2 loss.

Constraints

The minimization of the RHS in the linear regression equation results in a linear regression with no intercept. As such, adding L1 and L2 regularization parameters, we end up with the following:

$$\text{LASSO: } \min_{\beta} \|y - X\beta\|_2^2 + \lambda_1 \|\beta\|_1$$

$$\text{Ridge: } \min_{\beta} \|y - X\beta\|_2^2 + \lambda_2 \|\beta\|_2^2.$$

Hence we can see that the intercept of the linear equation actually contains the regularization parameter, lambda, that shall introduce the constraint. However, based on minimization of the LHS, the RHS of the equation would translate into an OLS solution first without constraints. Constraints shall then be explored in later below.

Section 3: Preparation of Data into Model

The data is first pre-processed by taking only the window period from December 31, 2021, and December 30, 2022. This helps to create the first window (Window #1) of 252 days to determine a portfolio with resultant asset weights of minimum variance, which shall then be explored further with a “rolling window” approach in exploring the dataset for better modelling predictions.

Within Window #1, we first try to determine the “y” and “X” of the Linear Regression equation. Making “y” the subject of equation in the Linear Regression model, it is calculated with y = R * wEW, where wEW is an equally weighted portfolio (wEW) of each asset having a weight of 0.01. R is then determined with the de-meaned Return Data, subsequently multiplied with wEW to perform a matrix multiplication.

Next, we calculate “X” where X = RN. In constructing matrix N, the identity matrix of (Ip-1) and a vector of all ones (1p-1) is included. Then “X” is derived through a matrix multiplication with the demeaned Return Data and N.

Then, we initialize the linear regression model without intercept, fitting the model and extracting the beta coefficients. We then calculate w for MinVar Portfolio = wEW - N * beta. Beta is then “reshaped” to make it a column vector for matrix multiplication compatibility.

Therefore, in Window #1, based on the first iteration of weights calculated for the MinVar portfolio between December 31, 2021 and December 30, 2022, a sample heatmap of the 100 assets’ weights can be summarized in the table below:

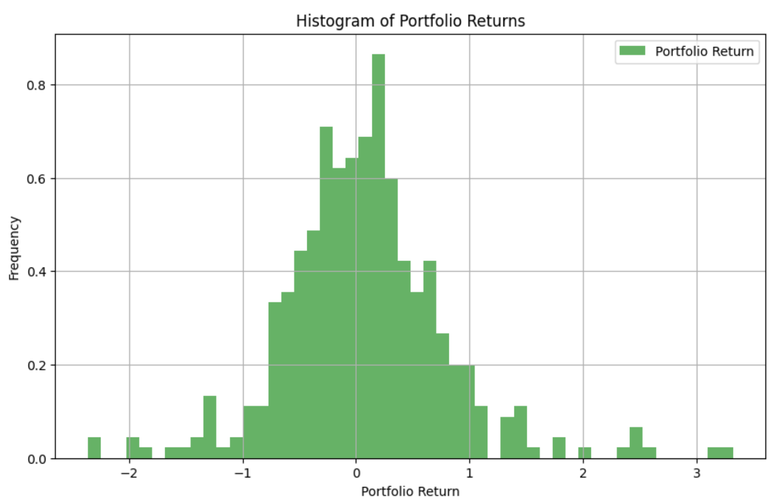
Weights To Assets	ME1	ME2	ME3	ME4	ME5	ME6	ME7	ME8	ME9	ME10
BM1	0%	-1%	4%	-7%	-4%	1%	0%	-6%	9%	-3%
BM2	-2%	8%	6%	4%	3%	-5%	-25%	13%	10%	-6%
BM3	-6%	-11%	1%	-10%	-4%	-17%	31%	4%	15%	8%
BM4	-11%	1%	-2%	5%	12%	-6%	-40%	-6%	-6%	14%
BM5	10%	9%	-17%	1%	15%	19%	-5%	12%	-12%	4%
BM6	-3%	-20%	-27%	2%	31%	-22%	17%	-8%	-18%	-7%
BM7	6%	3%	-3%	-4%	-2%	20%	29%	-6%	14%	9%
BM8	14%	-13%	9%	14%	-17%	7%	-22%	4%	12%	13%
BM9	72%	1%	0%	48%	-26%	-27%	15%	-37%	9%	2%
BM10	40%	-11%	7%	3%	-6%	-8%	-3%	6%	-19%	-4%

Compared to an Equal Weight (EW) portfolio where each weight assigned to each asset would be 1% for 100 assets, the MinVar portfolio shows both positive and negative weights. Positive weights represent a long position in an asset where one is buying the asset with the expectation that its value will increase. Negative weights represent a short position, where one is essentially borrowing the asset to sell it with the plan to buy it back later at a lower price, betting on the asset's value to decrease. (Elton, et. al, 2014)

With the above weights in MinVar Portfolio, we then calculate the portfolio returns for the first date in Day 253 on 3 Jan 2023, with a return of 1.096379.

Section 4: Generating Plots for Clear Analysis

Before sliding this “rolling window”, we are also able to determine the returns based on given data from Day 1 (31 Dec 2021) until Day 396 (31 July 2023), with a visualization of portfolio returns of these 396 days over time in Appendix 1. Next, we then generate Synthetic Predicted Returns with R^2 = 0.02 with add a large amount of noise; and for R^2 = 0.9 with adding a small amount of noise, illustrated in Appendix 2. We then generate a cumulative frequency plot to visualize the likelihood of Portfolio Returns:



Section 5: Rolling Window Approach

With Window #1 established from Day 1 to 252, we then iterate the next Window #2 from Day 2 to 253, until Day 144 to 396. With 396 days of training data, the regression lines for the 144 iterations are constructed through repeating the same steps above when constructing "y" and "X" in Window #1. Employing such a rolling window approach offers several advantages. Primarily, it facilitates extensive out-of-sample testing, as each window yields a distinct testing period, enabling a comprehensive performance evaluation of the model. Given the significant influence of recent market developments on financial asset returns, we've chosen to only use a subset of the entire dataset (most recent 396 days). This ensures the incorporation of the most current data, at the cost of having less data points to train the model. However, the rolling window methodology compensates for this by continually training the model on varied data subsets, maximizing both training and testing opportunities on such a small dataset.

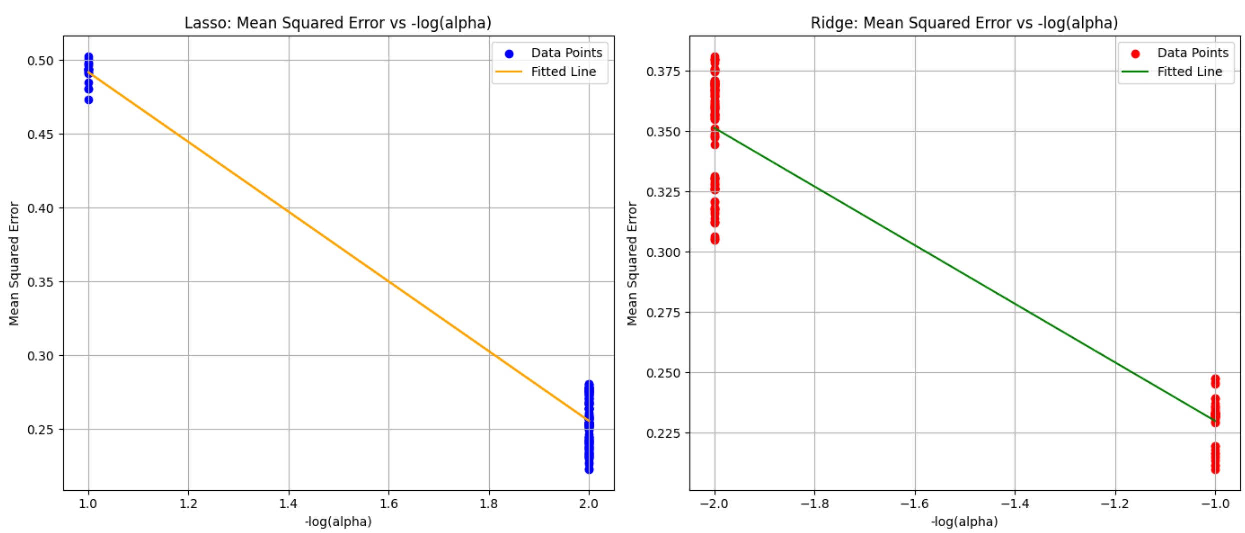
Within each of these training sets, we introduce cross validation with step size = 5. As a result of introducing this, we aim to find the optimal regularization hyperparameter lambda (alpha in python) for the LASSO and Ridge constraints. In this portfolio context, the different lambdas in each rolling window sets a cap on how large each beta should be, and thus determines the maximum deviation from the equally weighted portfolio our optimal portfolio should be.

By comparing the training set and validation set with the test data of the next day in 1 window, regularization helps to reduce overfitting as the OLS solution first derived from the linear regression expression was maximally overfitted. Each iteration then returns a unique lambda which is then stored to then construct the unique betas and derive the unique new weights assigned to each asset in the portfolio. In here, lambda, beta and weights are stored.

The stored weights are then used to calculate the stored portfolio returns for the test data as the measure of its test performance. Next, the mean squared error for different lambdas (regularization parameters) are calculated and stored, and for each log value of lambda, the corresponding mean squared error is also stored. This helps to identify the lowest test loss of all 144 logged lambdas.

<u>Section 6: Varying Constraints: LASSO and Ridge vs OLS</u>

After obtaining the regularization parameters, we obtain the optimal LASSO and ridge regression parameters which minimize the test loss the most in our rolling windows. Below is a summary of mean squared error vs the regularization parameter [lambda] (alpha in python) used.



The overall idea of this process is to decrease the extremely high variance which the regular OLS model is overfitting to; at the cost of adding a small amount of bias into our predictions. In general, LASSO, or L1 regularization, augments the linear regression model by adding a penalty term based on the absolute values of the coefficients. This regularization method promotes sparsity within the model, effectively selecting a subset of the most influential features while encouraging simplicity, by driving the beta coefficient along each "X" covariate to 0. Conversely, Ridge, or L2 regularization, supplements the linear regression model by introducing a penalty term based on the squared magnitudes of the coefficients. This serves to control the magnitude of the coefficients and counteract overfitting by constraining their growth.

Therefore, based on the above stored alphas, the average test returns and variance of test returns are calculated to determine the best Sharpe Ratios below:

(a) Sharpe ratio for LASSO: 0.079 (L1 Regularization)
(b) Sharpe ratio for Ridge: 0.029 (L2 Regularization)
(c) Sharpe ratio OLS: 0.039 (No Regularization)

For additional visualization, the returns and cumulative returns of the OLS, LASSO and Ridge regression portfolios can be found in Appendix 4 and Appendix 5 respectively.

<u>Section 7: Improvements To Prediction Framework and Results</u>

Motivated by LASSO's good performance in maximizing our Sharpe Ratio, Elastic Net emerges as an interesting alternative to change the constraints in our model for 2 reasons below:

(1) By combining the penalties of both Lasso (L1) and Ridge (L2) regularization, Elastic Net can capitalize on Lasso's ability to eliminate irrelevant features and Ridge's efficacy in countering overfitting (Zou et. al, 2005). This is especially pertinent when addressing collinear features in financial data, as Elastic Net remains stable, a challenge that Lasso tends to struggle with, whereas Ridge does not. Particularly, this dataset used in this project contains asset returns that are highly correlated with each other, as seen in the correlation matrix in Appendix 3.
(2) Elastic Net helps in dealing with high dimensionality, where "dimensionality" refers to the number of assets (covariates) included in the portfolio. Each asset can be thought of as a dimension in the vector space where the portfolio resides. In general, in high dimensions, models are more likely to fit the noise rather than the underlying data-generating process, leading to poor out-of-sample performance.

A summary of the application of Elastic Net into this project is seen below after all 144 iterations:

- Sharpe ratio Elastic Net: 0.04
- Average optimal L1_ratio: 0.16215277777777778

An average optimal L1_ratio of around 0.162 suggests that the model leans more towards L2 regularization (Ridge) than L1 regularization (Lasso).

<u>Section 8: Optimal Number of Assets</u>

Based on regularization through cross validation, the 6 assets identified in the python code would also be the optimal number of assets to work with. This reduces the number of features (assets) and ultimately reduce the dimensionality of the linear equation for greater interpretability purposes. The weights of the 6 assets in this minimum variance portfolio is ([-0.37557609, -0.04628394, -0.19910571, 0.10350876, 0.15339477, 1.36406221]).

Based on these 6 given weights, the asset returns can also be derived (refer to Appendix 6) with a visualization of a time-series plot over time (refer to Appendix 7).

Section 9: Evaluation

In summary, the best solution that maximizes the Sharpe ratio would be LASSO regression, which is reflective of the measure assigned to the project/business objective. Although portfolio weights w are expressed as $w = w_{\mathrm{EW}} - \beta,$ which may mean that a sparse β may not necessarily lead to sparse w, it has produced the best Sharpe Ratio of 0.079 with the following advantages:

(1) Regularization: Even if LASSO doesn't produce sparse w, the regularization still helps in reducing overfitting, which is a common problem in high dimensions. This can improve the portfolio's out-of-sample performance.

(2) Stability: By shrinking the coefficients towards zero, LASSO tends to produce a more stable and robust portfolio, which can be particularly beneficial when dealing with high-dimensional data.

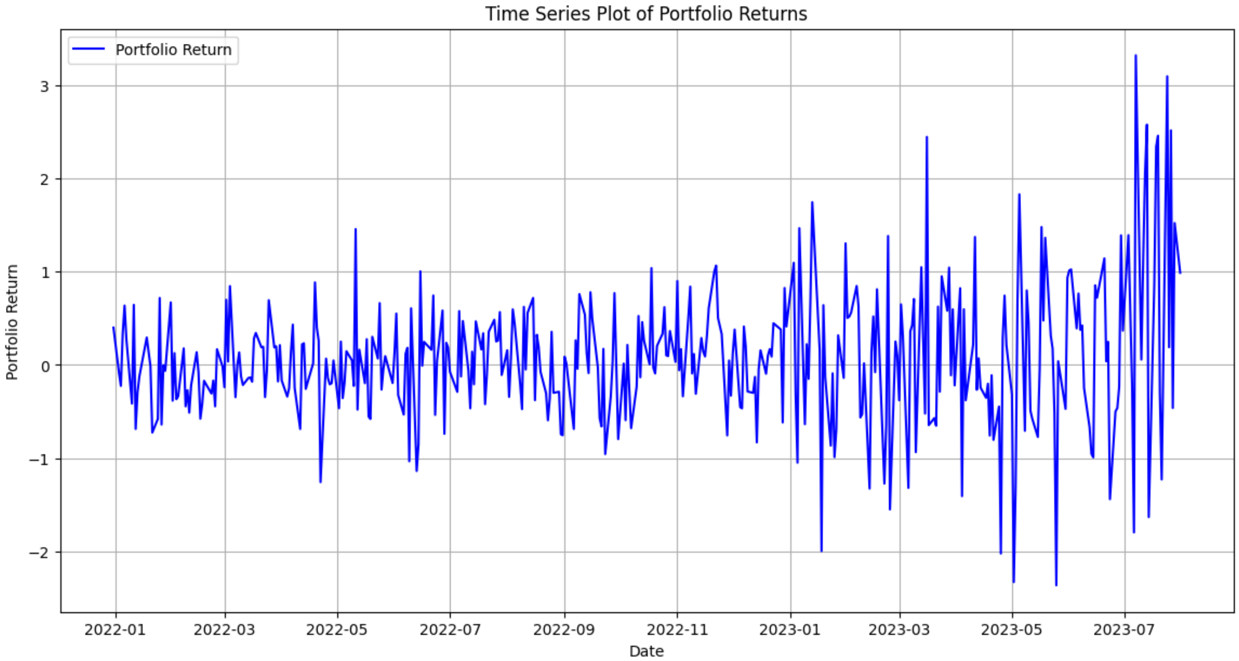
(3) Interpretable Models: While it may not lead to sparse w, a sparse β can still make the model more interpretable by highlighting which assets' deviations from the equally-weighted portfolio are most influential.

(4) Computationally Efficient: LASSO is computationally efficient even for high-dimensional problems, making it a practical choice for large portfolios.

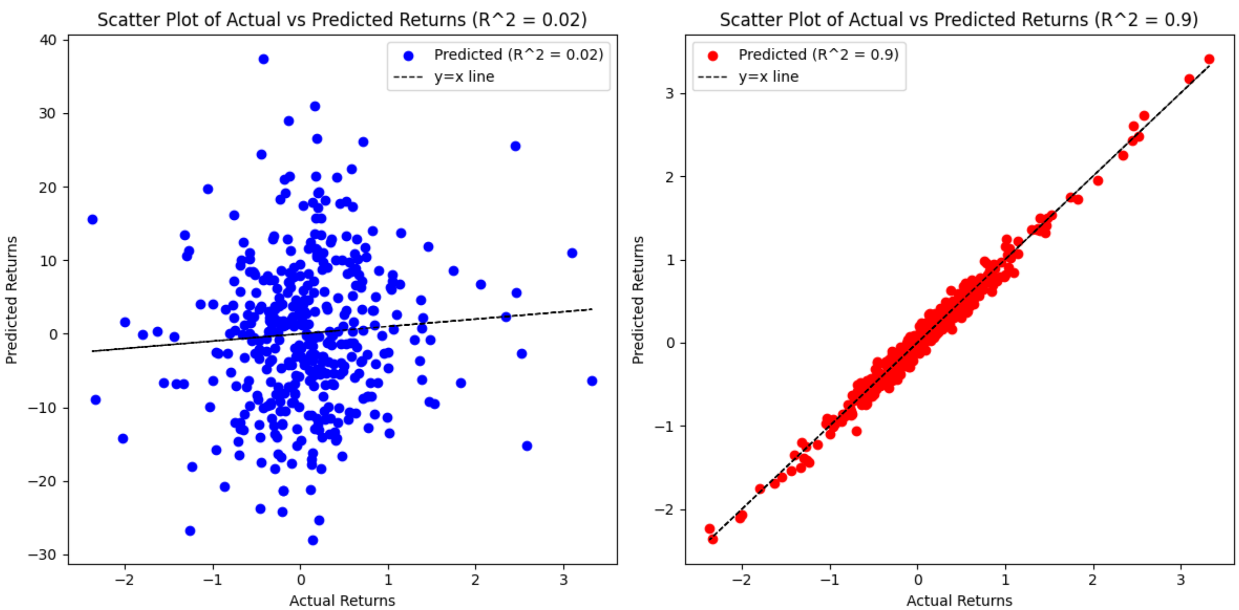
Additionally, the component of “Constraint” in the predictive model equation has been varied, through the experimentation of Elastic Net. Elastic Net adjusts the balance between L1 and L2 regularization by the hyperparameter ‘l1_ratio’, making it adaptable for specific portfolio modeling requirements. However, Elastic Net’s computational demands are heightened due to the extra hyperparameter, potentially impacting real-time portfolio decisions (Zou, et. al, 2005). While it may yield better predictive outcomes compared to Ridge and OLS in Section 6, the interpretability of the resultant model may not be as clear as those produced solely by LASSO or Ridge. It is therefore important to ensure thorough validation of Elastic Net to prevent potential overfitting in the intricate landscape of financial markets.

As a final evaluation, there is a mismatch between test and train loss. This is because the models are trained by minimizing the L2 loss aiming to achieve the best fit between predicted and actual returns in the training data (as this is the OLS). However, the evaluation of the portfolio's performance is based on the Sharpe ratio, a measure of risk-adjusted return. This presents a slight disconnect between the training objective and the evaluation metric. Our OLS problem focuses on reducing the squared differences between predicted and actual returns, essentially minimizing variance. On the other hand, the Sharpe ratio balances both maximizing the return AND minimizing the risk, offering a holistic view of the portfolio's performance. There could for example be portfolios with higher variance, but an even higher expected return, resulting in a more optimal Sharpe ratio. As such, we may not necessarily be finding the portfolio with the highest Sharpe ratio, despite this being our main evaluation metric.

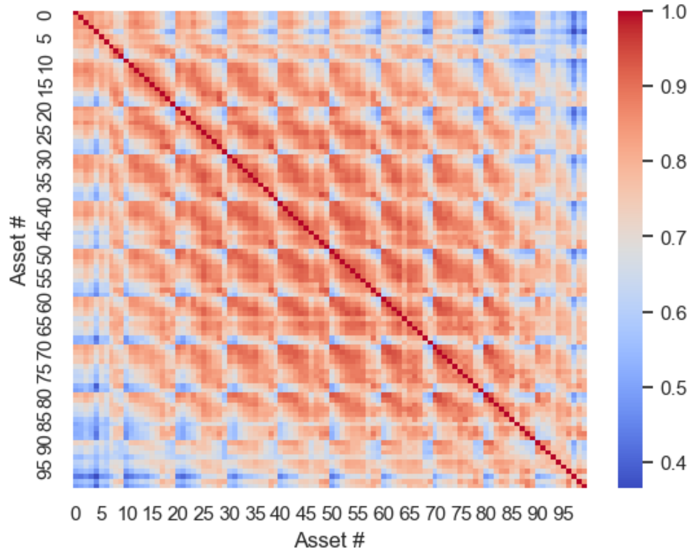
Appendix:



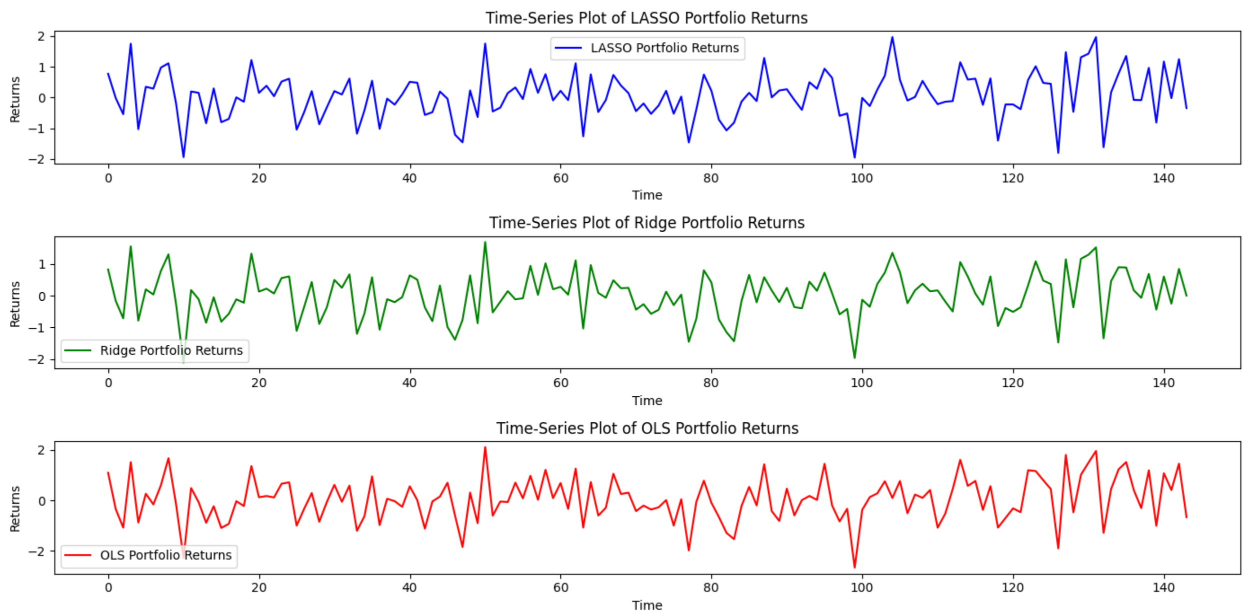
Appendix 1: Portfolio Returns from 31-12-2021 to 31-07-2023



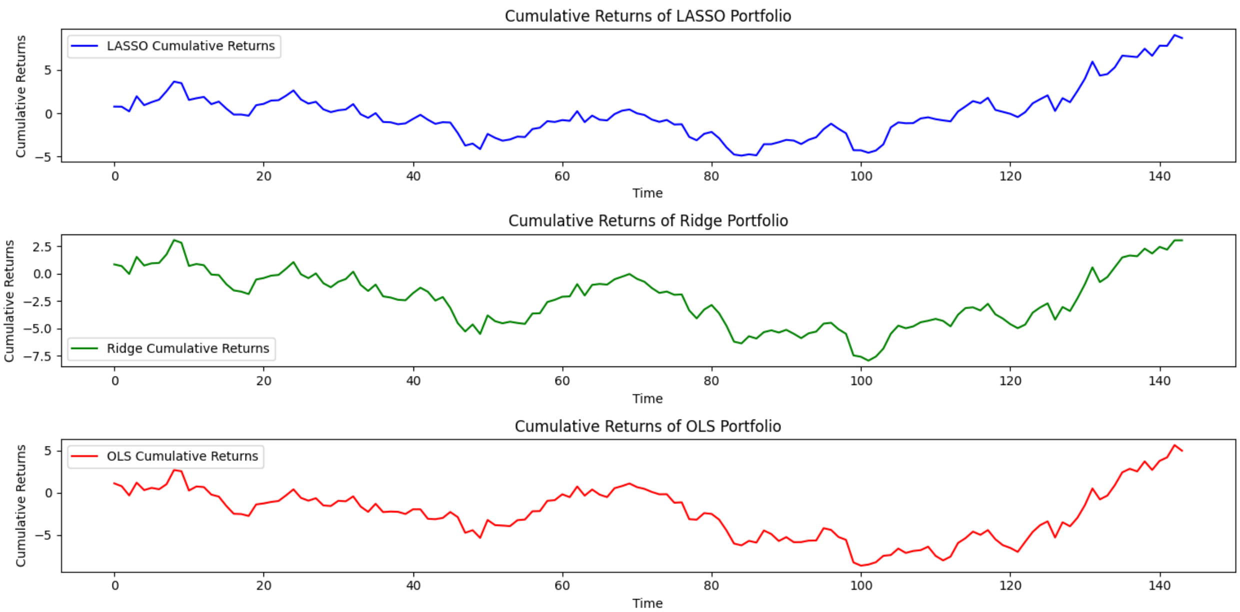
Appendix 2: Synthetic R^2 values for the Scatter Plot of 396 days of data



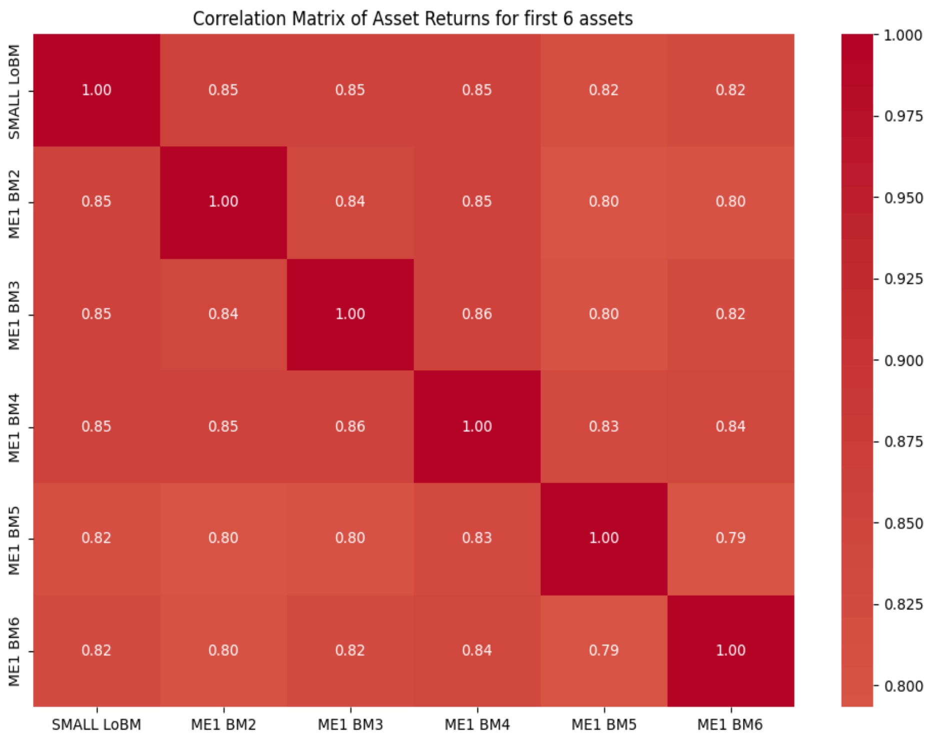
Appendix 3: Correlation Matrix between 100 Assets in Portfolio Data



Appendix 4: Portfolio Returns vs Time

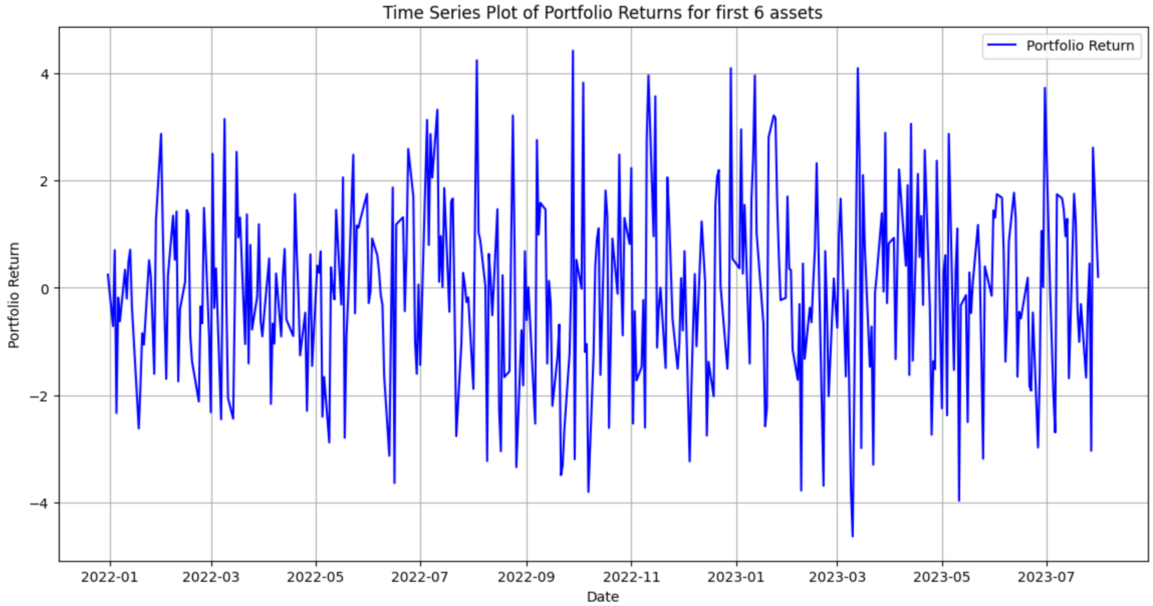


Appendix 5: Portfolio Cumulative Returns over Time



	SMALL LoBM	ME1 BM2	ME1 BM3	ME1 BM4	ME1 BM5	ME1 BM6
SMALL LoBM	1.00	0.85	0.85	0.85	0.82	0.82
ME1 BM2	0.85	1.00	0.84	0.85	0.80	0.80
ME1 BM3	0.85	0.84	1.00	0.86	0.80	0.82
ME1 BM4	0.85	0.85	0.86	1.00	0.83	0.84
ME1 BM5	0.82	0.80	0.80	0.83	1.00	0.79
ME1 BM6	0.82	0.80	0.82	0.84	0.79	1.00

Appendix 6: Correlation Matrix of Asset Returns for Optimal 6 Assets



Appendix 7: Portfolio for Optimized 6 Asset Returns over Time

References:

Elton, E. J., Gruber, M. J., Brown, S. J., & Goetzmann, W. N. (2014). Modern Portfolio Theory and Investment Analysis (9th ed.). Wiley.

Zou, H., & Hastie, T. (2005). "Regularization and variable selection via the elastic net." Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301-320.