Optimizing Investment Strategies: A Dual Approach Featuring Performance-Based Regularization and Recession Prediction Modeling

Master Thesis Quantitative Finance

Thijs Hagen (471695)

Supervisor: Dr. Rasmus Lönn

Second Assessor: Dr. Gustavo Freire

February 7, 2024

Abstract

In this research, we evaluate multiple portfolio optimization strategies complemented with performance-based regularization (PBR), a machine learning regularization procedure proposed by Ban et al. (2018). We first analyse the addition of a new optimized threshold on the uncertainty in the sample variance in a mean-variance framework. We then evaluate multiple additions and robustness checks to the model and find that the PBR procedure significantly improves portfolio performance for some portfolios. Especially during times of economic downturn, the PBR portfolios perform well compared to the regular MV or GMV portfolios and much better than the 1/N portfolio. Additionally, we forecast the state of the economy with a random forest model and combine a conservative with a more aggressive portfolio based on this. We find that our model can forecast accurately whether there will be a recession the next month. Furthermore, the combined portfolios tend to perform better than their individual counterparts, and the combined model with an optimized weighting scheme performs best of all evaluated portfolios, including transaction costs.

Contents

1 Introduction

According to Statista, the assets of Dutch banking groups amounted to a total of 2.6 trillion euros in 2023, which is more than twice the size of the Dutch GDP. A lot of research has been conducted on the topic of portfolio management, and it turns out to be rather difficult to outperform the 1/N portfolio with equal weights over all assets. DeMiguel et al. (2009) find that out of 14 evaluated models, not one model consistently outperforms the 1/N portfolio in terms of Sharpe ratio, certainty-equivalent return or turnover. This indicates that the gain from optimal diversification is outweighed by estimation error. Since the research from DeMiguel et al. (2009), new papers on the topic have been released with different portfolio optimization methods in a mean-variance framework, but the investigation of diverse optimization techniques is still active literature.

This research aims to evaluate and compare several portfolio optimization techniques on both returns and risk. The research contains two major parts. The first part comprises an analysis of portfolio performance when the sample variance of the sample variance is restricted in a mean-variance framework. This is particularly interesting since most research focuses on plug-in estimates. A widely applied procedure is to shrink the sample mean, like Jorion (1986) do in their research. Other research focuses on restricting the sample variance, for example, Ledoit and Wolf (2003). However, Restricting the sample variance's uncertainty is not a widely adopted method yet, but research by Ban et al. (2018) shows promising results, making it interesting to investigate further.

For this purpose, an analysis is performed on the Fama-french five industry (FF5), FF10 and FF49 datasets. The optimal restriction thresholds are obtained from a k-fold cross-validation procedure where the Sharpe ratio of the hold-out sample is optimized. For the optimization of the threshold, different optimization procedures are evaluated, which are backtracking line search (LS), grid search (GS) and random search (RS). The performance-based regularization (PBR) portfolio is compared with other widely used portfolio optimization frameworks in both a mean-variance and a minimum-variance context. Thereafter, several additions to the model are evaluated, which are boundaries on weights, additional optimization methods for the threshold, releasing a restriction and taking a moving target return equal to the 1/N portfolio in the mean-variance framework. Furthermore, the impact of transaction costs is assessed.

The PBR technique turns out to be especially valuable in times of economic downturn because then losses are relatively small as a result of the added restriction. Therefore, we incorporate this characteristic in the second part of our research. In this part, we first make forecasts on the state of the US economy with a random forest model. As input for the random forest model, we evaluate a variety of mainly macroeconomic variables. We then analyse the impact of all features on model output using Shapley additive explanation (SHAP) values and evaluate the performance of our model in predicting recession or expansion. After having obtained probabilities on a US recession, we invest in a weighted combination portfolio constructed of a conservative and aggressive PBR portfolio.

For the first part, we find that the PBR portfolio sometimes significantly outperforms its SAA counterpart for the FF5 dataset, depending on model specifics. When we further improve our models by bounding weights, releasing a restriction and setting the target return in the mean-variance framework equal to the moving 1/N portfolio return, Sharpe ratios are increased further, and many models do significantly outperform the base SAA models. Taking into account transaction costs impacts portfolio performance, but the MV PBR-LS model with a target return of the 1/N portfolio still significantly outperforms its SAA counterpart. Besides this, we find that also when transaction costs are included our best models significantly outperform the 1/N portfolio, which has a much lower Sharpe ratio.

Regarding recession prediction with the random forest model, we find that relatively few features have much impact on the model output. Whether there was a recession in the previous month and the composite leading indicator are most important followed by business tendency, mortgage rate and inflation. We further find that our model performs very well in predicting a recession and also in predicting an expansion, which means it does not predict a recession too often. In terms of Sharpe ratio, four out of five combined portfolios based on forecasted recession probabilities outperform their individual counterparts (well-performing individual portfolios from part 1) excluding transaction costs and two also including transaction costs. Two combined portfolios also outperform the aggressive individual portfolio in terms of wealth accumulation while decreasing volatility through partial investments in the conservative portfolio. Including transaction costs, the combined portfolio with optimized weights based on a time series cross-validation procedure performs best of all evaluated portfolios regarding Sharpe ratio.

Next to adding to the literature on PBR and model combination, this research also adds value on a practical level, since asset managers can use the research to evaluate several techniques and incorporate well-performing methods in investment strategies. Particularly the PBR portfolios with high Sharpe ratio and low volatility are interesting for conservative asset managers like pension funds since these portfolios perform relatively well in crises and also achieve a decent return. Moreover, the combined portfolio strategy might be valuable for investors since they can take advantage of times of surging stock prices while taking on low-risk positions in times of economic downturn.

In the rest of this paper, we first discuss the relevant literature with regard to our research in Section 2. Thereafter, we discuss the data that we use and where it can be obtained in Section 3. Then, we present the used methods for the PBR optimization framework in Section 4, after which we explain all procedures and techniques corresponding to recession forecasting and portfolio combination in Section 5. Results of the PBR part are shown and interpreted in Section 6 and results of the recession forecasting and model combination in Section 7. We finish the main part of the paper with a conclusion in Section 8 after which the Appendix follows.

2 Literature Review

DeMiguel et al. (2009) evaluate the out-of-sample performance of the sample-based mean-variance model and extensions on this model, compared to the 1/N portfolio. They find that none of the 14 models (sample-based mean-variance model, Bayesian models, moment restricted models, portfolio-constrained models and combinations of portfolios) they analyse consistently outperforms the 1/N portfolio in terms of Sharpe ratio, certainty-equivalent return, or turnover. This implies that, out of sample, the gain from optimal diversification is often offset by estimation error. Besides this, they find that for the sample-based mean-variance strategy, under normality, the estimation window needs to be around 3000 months for a portfolio with 25 assets in order to outperform the 1/N portfolio. This indicates that it is still very challenging to outperform the 1/N portfolio based on mean-variance optimal portfolio choice. Kirby and Ostdiek (2012) find that implementing the mean-variance model with a target return equal to that of the expected return on the 1/N portfolio generally outperforms naive diversification when transaction costs are not taken into account. Therefore, setting the target return of the mean-variance portfolio equal to the expected return of the 1/N portfolio and using this for different variations of the regular mean-variance portfolio is also evaluated in this research.

Stein (1956) introduces the concept of shrinking the mean and thereby decreasing estimation error, and Jorion (1986) uses the mean of the GMV portfolio as a shrinkage target for the mean-variance portfolios. Ledoit and Wolf (2003) evaluated the impact of shrinking the covariance matrix by taking a weighted average of the sample covariance matrix and the single-index covariance matrix (based on the market factor). They find that the portfolio constructed with the shrunk covariance matrix has a lower standard deviation than portfolios constructed with other widely used estimators for the covariance matrix. Ledoit and Wolf (2004) introduce an estimator that is well-conditioned (inverting does not increase estimation error) and more accurate than the sample covariance matrix asymptotically. They find that their estimator has smaller risk and is better conditioned than the sample covariance matrix.

Brandt (2010) discusses the econometric treatment in portfolio choice problems. Specifically, the paper covers three major parts, namely the theoretical aspects of constructing portfolios, traditional econometric approaches (plug-in estimation and decision theory) and an alternative econometric approach in the form of parametric portfolio weights. DeMiguel et al. (2013) performs an extensive investigation of shrinkage estimators for asset allocation and find that the shrinkage intensity plays a significant role in the performance of their portfolios. Moreover, they highlight the importance of calibrating shrinkage estimators for constructing optimal portfolios.

Goto and Xu (2015) propose a sparse estimator of the inverse covariance matrix, which means that a significant fraction of the off-diagonal elements is zero. They find that implementing this sparse estimator results in a significant out-of-sample risk reduction, especially when the regular sample covariance matrix estimator is ill-conditioned. Their approach also mitigates estimation error. Ledoit and Wolf (2022) discuss their findings in the estimation of large-dimension covariance matrices over the last 15 years. They review linear and non-linear shrinkage methods and state that their linear shrinkage techniques improve upon the original estimator. Their non-linear shrinkage method does not modify the sample covariance matrix but works on its eigenvalues. They state that they have found a way to overlay a particular structure (e.g. a factor model) on non-linear shrinkage to increase the performance of regular non-linear shrinkage techniques.

Ban et al. (2018) implement two machine learning methods, regularization and cross-validation, to improve portfolio performance of a regular mean-variance portfolio. They first conduct a performance-based regularization (PBR) method with the idea of constraining the sample variances of the estimated portfolio risk and return, which adjusts the solution towards one with less estimation error. They apply PBR on both a mean-variance and a mean-conditional-value-at-risk (CVAR) framework. For the mean-variance problem, the first PBR method restricts the uncertainty in the sample variance, which results in shrinking the optimal regular mean-variance portfolio weights. The second PBR method shrinks the sample covariance matrix. To calibrate the threshold on the sample variance of the estimated portfolio risk, they apply the k-fold cross-validation method where they evaluate the Sharpe ratio on the hold-out samples for different values for the threshold. They find that the PBR portfolios dominate all other portfolios (including the 1/N portfolio) in terms of Sharpe ratio for two out of three Fama-French data sets.

Because of the innovative methodology and promising results of the PBR technique, this is particularly interesting to further investigate and extend. Therefore, in the first part of this research, we evaluate the PBR method and contribute to it by introducing a restriction on portfolio weights, evaluating additional optimization methods for the threshold, releasing a technical optimization restriction, imposing a moving target return equal to the 1/N portfolio in the mean-variance framework and incorporating transaction costs. Furthermore, we perform an analysis on the thresholds through time.

In terms of which optimization procedure to use for finding optimal hyperparameters, there are several options, according to academic research. For finding the optimal threshold of the sample variance of the risk, Ban et al. (2018) use a backtracking line search algorithm. Moreover, Bergstra and Bengio (2012) state that grid search and manual search are the most widely-used methods for hyperparameter optimization. In their research, they compare grid search with random search and find that random search is more efficient than grid search. In this research, we will evaluate both grid search and random search to find the optimal threshold for the sample variance of the risk. Snoek et al. (2012) analyses Bayesian optimization for hyperparameter selection of general machine learning algorithms and find that Bayesian optimization selects better hyperparameters than other evaluated methods. However, since our main focus is on optimizing only the threshold on the sample variance of the risk and not optimizing many hyperparameters simultaneously, the use of Bayesian optimization might cause more complications and runtime than that it benefits our case and is therefore not incorporated.

Next to the analysis of PBR portfolios, this research also contributes in the field of portfolio combination. DeMiguel et al. (2009) evaluate many different portfolios in their research, among which also some portfolios that are combinations of other portfolios. Specifically, they evaluate the three-fund portfolio from Kan and Zhou (2007), which is a portfolio consisting of two risky assets and the risk-free asset. Weights are optimized here such that the expected utility is maximized. They also analyse a combination of the 1/N portfolio and the minimum-variance portfolio, where weights in both portfolios are again chosen such that the expected utility is maximized. Tu and Zhou (2011) study the combination of the 1/N portfolio with four portfolios, namely the Markowitz portfolio and the portfolios proposed by Jorion (1986), Craig MacKinlay and Pástor (2000) and Kan and Zhou (2007). The optimal combination between the 1/N portfolio and the other portfolios is determined through a coefficient that optimizes the bias-variance trade-off. They find that the combined strategies significantly improve the four individual portfolios and also outperform the 1/N portfolio in most scenarios.

The above-mentioned portfolio combination studies do not take into account the state of the economy when forming model combinations. In terms of state-adjusted modelling, Elliott and Timmermann (2005) optimize combination weights based on a regime-switching model in order to make forecasts on macroeconomic variables and find that this performs well for some of the scenarios. Furthermore, Clarke and de Silva (1998) create efficient frontiers for different states of the economy.

In this research, we explicitly take into account the economic state in the construction of combined portfolios. First, we make a prediction on the state of the economy for the next month. Based on this value, we form a model combination between a conservative and a risky portfolio optimization strategy. To the best of our knowledge, integrating this approach of forecasting the state of the economy with a machine learning model and determining portfolio combination weights based on this has not yet been performed before and is therefore particularly interesting to investigate.

We make use of a wide variety of explanatory variables to forecast economic state. Regarding this topic, Gogas et al. (2015) use yield curve and GDP data to forecast output fluctuations around its long-run trend. They focus on correctly forecasting output gaps, which they refer to as recessions, and apply a support vector machine technique for classification. They obtain an overall forecasting accuracy of 66.7% and an accuracy of 100% for recession forecasting. Nyberg (2010) evaluates the performance of a dynamic probit model in forecasting a recession in the US and Germany. As input, he uses a variety of financial variables, and finds that the dynamic probit model outperforms a standard static model.

Random forest models are widely used in financial applications, and we also use them in this research for reasons explained later on. The random forest model needs features as input and we aim to evaluate importance of these features. For this purpose, Lundberg and Lee (2017) state that diverse methods have been proposed to interpret predictions of complex models but it is not clear when one method is preferred over another. Therefore, they propose Shapley additive explanations (SHAP) to evaluate absolute feature impact and the relation between the direction of a feature and the output. SHAP values were initially proposed by Shapley et al. (1953) and are based on game theory. Jabeur et al. (2021) also make use of SHAP values to analyse feature importance in forecasting gold prices with six machine learning models, among which the random forest. We also use SHAP values to evaluate the impact of our features.

3 Data

We evaluate monthly portfolio returns for the Fama-French 5, 10 and 49 industry portfolios over a period ranging from January 1994 until December 2013. The portfolio returns can be obtained from Kenneth R. French's website[1]. The portfolios are formed by assigning each stock from the NYSE, AMEX, and NASDAQ to an industry portfolio at the end of June every year. Hence, the analysed portfolios all consist of US stocks. For the five industry portfolios, the stocks are divided over four industries and the residual stocks are captured in the last portfolio which is called 'other'. Specifically, the four industries are consumer, manufacturing, high-tech and healthcare. For the ten industry portfolios, the same is done but then a split of ten industries is made and for the 49 industry portfolios the stocks are divided over 49 industries. In this paper, value-weighted returns are analysed for the different portfolios.

Next to the industry portfolios, we use the risk-free rate, which is also taken from the website of Kenneth. R. French. Specifically, we take the risk-free rate given in the Fama/French 3 Factors monthly returns file. We analyse data from January 1994 until December 2013, so 240 observations in total. This is the same number of observations used by Ban et al. (2018) for their research. The 240 observations are split into a train set and a test set where the train set consists of $T_{train} = 120$ observations, and the test set also of $T_{test} = 120$ observations. A rolling window is used to obtain optimal weight forecasts every month.

In the tables below, annualized descriptive statistics for the different data sets are shown for the returns over the period from January 1994 until December 2013. In Table 1, it can be observed that the manufacturing industry portfolio has the highest Sharpe ratio and the consumer industry portfolio the lowest volatility. In Table 2, the consumer nondurables industry portfolio attains the highest Sharpe ratio and the lowest volatility. For the 49 industry portfolio, there are too many portfolios to show the descriptive statistics for all portfolios but we report the statistics for the portfolios with the minimum, first quartile (Q1), median, third quartile (Q3) and maximum Sharpe ratio of all portfolios. We observe that the maximum Sharpe ratio does not differ much from the maximum Sharpe ratio for the FF5 and FF10 portfolios, but the minimum is much lower. These are the statistics for the entire data, so including both the train set and the test set. We only analyse portfolio performance on the test set. Therefore, also descriptive statistics of the data on only the test set are added in Tables 16, 17 and 18 in the Appendix.

	FF 5 industry				
	Consumer	Manufacturing	Hi-Tech	Health	Other
Mean	7.8535	9.3560	8.6895	9.7385	6.9215
SD	13.5542	15.0840	23.1326	14.7357	18.5991
SR	0.5794	0.6203	0.3756	0.6609	0.3721

Table 1: Annualized descriptive statistics for excess returns on the 5 industry portfolios for the period ranging from January 1994 until December 2013.

[1] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

	FF 10 industry									
	Cnsmr Nond.	Cnsmr Dur.	Manuf.	Energy	Hi-Tech	Telecom	Shops	Health	Utilities	Other
Mean	8.7560	6.2575	9.8880	11.1645	10.6055	5.7055	8.1525	9.6395	6.8840	6.8910
SD	12.7637	25.8389	17.2506	19.1959	26.3749	18.8912	15.7853	14.7623	14.4808	18.4962
SR	0.6860	0.2422	0.5732	0.5816	0.4021	0.3020	0.5165	0.6530	0.4754	0.3726

Table 2: Annualized descriptive statistics for excess returns on the 10 industry portfolios for the period ranging from January 1994 until December 2013.

	FF 49 industry				
	Min.	Q_1.	Median	Q_3	Max.
Mean	1.8705	9.6805	10.05	8.464	10.4135
SD	39.3025	30.5676	23.7115	17.1075	15.715
SR	0.0476	0.3167	0.4238	0.4948	0.6626

Table 3: Annualized descriptive statistics for excess returns on the 49 industry portfolios for the period ranging from January 1994 until December 2013. Due to the large number of industries, the minimum, $Q1$, median, $Q3$ and maximum values of the industry portfolios are shown.

For the second part of this research, we try to estimate whether the economy is in a good or a bad state the next month and make investment decisions based on this forecast. Therefore, we need data that represents the state of the economy. For this, we use the "NBER-based Recession Indicators for the United States from the Period following the Peak through the Trough" dataset from the FRED, which is monthly binary data that takes on value one if the US economy is in a recession and value zero if it is in a normal state. Specifically, for a recession they take the first month after a peak (a point in time at which a variety of economic indicators reaches its highest level followed by a substantial decline in economic activity). The recession then continues until the first month with a trough. A trough is a point when economic activity is at its lowest, after which a considerable increase in economic activity occurs. Hence, recession is not defined in the conventional way of two consecutive quarters of economic downturn.

In order to forecast the binary recession variable, we make use of a random forest model with features. Most of these features are US macroeconomic variables, and data can be found on the FRED website. A table with all the used variables and their source is presented in Section 5.2. For all the variables, we use monthly data from June 1976 until December 2013. June 1976 is chosen as the start date because yield spread data does not reach further back, and every variable needs to start at the same date to be incorporated in the random forest model.

4 Performance-Based Regularization

In this research, we analyse a performance-based regularization procedure inspired by Ban et al. (2018). We evaluate different portfolio optimization methods on the monthly data from the Fama-French (FF) industry portfolios. Specifically, we evaluate whether setting a threshold on the sample variance of the sample variance of the portfolio in a mean-variance framework increases performance compared to widely used other optimization methods. The time-varying threshold is determined based on a k-fold cross-validation procedure where the Sharpe ratio of the hold-out sample is optimized. In this Section, we first briefly discuss the different evaluated models, among which is the PBR model. Thereafter, we explain in detail the k-fold cross-validation procedure to obtain optimal thresholds, after which we discuss several optimization methods used within the k-fold cross-validation. Then we elaborate on all used evaluation metrics. All programming is performed in Python except for the computation of the HAC and studentized bootstrap p-values, which is in R.

4.1 Evaluated Portfolios

In this research, we analyse the effect of constraining the sample variance of the variance for the GMV portfolio and the mean-variance portfolio with different target returns. In order to evaluate the performance of these regularized models, we compare them with the regular sample average approximation (SAA) portfolio, the SAA portfolio with no short-selling constraint, the Lasso portfolio, the Ridge portfolio, the James Stein portfolio with a shrunk mean and the Ledoit and Wolf portfolio with a shrunk covariance matrix. These other portfolios are solely included for comparison purposes and not an application of the PBR methodology. The mentioned portfolios are elaborated on in the next paragraphs.

4.1.1 Sample Average Approximation

The SAA portfolio is simply the regular portfolio optimization model where the mean constraint between brackets is left out for the GMV portfolio and is included for the mean-variance portfolios. The mean-variance portfolios are analysed for monthly portfolio target returns ($\hat{\mu}_{target}$) of 0.8%, 1.0% and 1.2% because these returns are in line with the returns on the individual assets. Following Kirby and Ostdiek (2012) who use $\hat{\mu}_{target} = \frac{\hat{\mu}_t \iota}{N}$, we also analyse portfolio performance for a moving target portfolio return equal to that of the 1/N portfolio. This approach guarantees that the aggressiveness of the target return is in line with the return on the 1/N portfolio for each period. The regular mean-variance optimization framework is shown below.

$$\begin{aligned} \min_{w \in R^N} \quad & w'\hat{\Sigma}w \\ \text{s.t.} \quad & w'\iota = 1 \\ & (w'\hat{\mu} = \hat{\mu}_{target}) \end{aligned} \tag{1}$$

4.1.2 PBR

In order to perform performance-based regularization, we add the last constraint in equation (2), similar to Ban et al. (2018). This constraint restricts the sample variance of the sample variance and the threshold U is optimized with regards to the Sharpe ratio for every iteration. This optimization procedure is explained in detail in Sections 4.2 and 4.3.

$$\begin{aligned} \min_{w \in R^N} \quad & w'\hat{\Sigma}w \\ \text{s.t.} \quad & w'\iota = 1 \\ & (w'\hat{\mu} = \hat{\mu}_{target}) \\ & Svar(w'\hat{\Sigma}w) \leq U \end{aligned} \tag{2}$$

Here, the sample variance of the variance is given by

$$Svar(w'\hat{\Sigma}w) = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N} w_i w_j w_k w_l \hat{Q}_{ijkl} \tag{3}$$

with $\hat{Q}$ as follows

$$\hat{Q}_{ijkl} = \frac{1}{T}(\hat{\mu}_{4,ijkl} - \hat{\sigma}^2_{ij}\hat{\sigma}^2_{kl}) + \frac{1}{T(T-1)}(\hat{\sigma}^2_{ik}\hat{\sigma}^2_{jl} + \hat{\sigma}^2_{il}\hat{\sigma}^2_{jk}) \tag{4}$$

and $\hat{\mu}_{4,ijkl}$ the sample average estimator for the fourth central moment $\mu_{4,ijkl}$, which is given by

$$\mu_{4,ijkl} = E(X_i - \mu_i)(X_j - \mu_j)(X_k - \mu_k)(X_l - \mu_l) \tag{5}$$

and $\hat{\sigma}^2_{ij}$ the sample average estimator of the covariance estimator σ^2_{ij}, given by

$$\sigma^2_{ij} = E(X_i - \mu_i)(X_j - \mu_j) \tag{6}$$

Hence, the constraint that states that the sample variance of the sample variance is smaller than a certain threshold U is a quartic polynomial constraint in the decision vector w. It is not clear whether $Svar(w'\hat{\Sigma}w)$ is a convex function in w, and hence it is uncertain if the problem is convex. Similar to Ban et al. (2018), we thus consider a convex approximation. In order to do this, we make a rank-1 approximation and write the quartic polynomial constraint as follows.

$$(w'\hat{\alpha})^4 \approx \sum_{ijkl} w_i w_j w_k w_l \hat{Q}_{ijkl} \tag{7}$$

Now we compute $\hat{\alpha}$ by

$$\hat{\alpha}_i = \sqrt[4]{\hat{Q}_{ijkl}} = \sqrt[4]{\frac{1}{T}\hat{\mu}_{4,iiii} - \frac{T-3}{T(T-1)}(\hat{\sigma}^2_{ii})^2} \tag{8}$$

From this, we obtain the following convex approximation

$$
\begin{aligned}
\min_{w \in R^N} \quad & w'\hat{\Sigma}w \\
\text{s.t.} \quad & w'\iota = 1 \\
& (w'\hat{\mu} = \hat{\mu}_{target}) \\
& w'\hat{\alpha} \leq \sqrt[4]{U}
\end{aligned} \tag{9}
$$

with α as in equation 8.

The solution of the PBR problem can be written as an adjustment to the unconstrained mean-variance portfolio problem, as shown below.

$$
\hat{w}_{PBR} = \hat{w}_{MV} - \frac{1}{2}\lambda^*\hat{\Sigma}^{-1}(\beta_1\iota + \beta_2\hat{\mu} + \hat{\alpha}) \tag{10}
$$

Where $\hat{w}_{MV}$ is the sample average approximation (SAA) solution, λ^* the optimal Lagrange multiplier for the PBR constraint $w'\hat{\alpha} \leq \sqrt[4]{U}$ and

$$
\beta_1 = \frac{\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\mu} \cdot \hat{\mu}\hat{\Sigma}^{-1}\iota - \hat{\alpha}'\hat{\Sigma}^{-1}\iota \cdot \hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu}}{\iota'\hat{\Sigma}\iota \cdot \hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu} - (\hat{\mu}'\hat{\Sigma}^{-1}\iota)^2} \tag{11}
$$

$$
\beta_2 = \frac{\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\mu} \cdot \iota'\hat{\Sigma}\iota - \hat{\alpha}'\hat{\Sigma}^{-1}\iota \cdot \iota'\hat{\Sigma}^{-1}\hat{\mu}}{\iota'\hat{\Sigma}\iota \cdot \hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu} - (\hat{\mu}'\hat{\Sigma}^{-1}\iota)^2} \tag{12}
$$

The solution (without the mean constraint) is now given by

$$
\hat{w}_{PBR} = \hat{w}_{MV} - \frac{1}{2}\lambda^*\hat{\Sigma}^{-1}(\beta\iota + \hat{\alpha}) \tag{13}
$$

with β as follows

$$
\beta = -\frac{\iota'\hat{\Sigma}^{-1}\hat{\alpha}}{\iota'\hat{\Sigma}^{-1}\iota'} \tag{14}
$$

Hence, the threshold and the rank-1 approximation result in shrinking the SAA optimal weight by an amount scaled by λ^*, towards a direction that depends on the (approximated) fourth moment of the asset returns.

4.1.3 SAA Without Short Selling

For this portfolio, the no short-selling constraint is added, which implies that all weights should be larger than zero. Again, the mean constraint between brackets is included for the mean-variance portfolio and excluded for the GMV portfolio.

$$
\begin{aligned}
\min_{w \in R^N} \quad & w'\hat{\Sigma}w \\
\text{s.t.} \quad & w'\iota = 1 \\
& (w'\hat{\mu} = \hat{\mu}_{target}) \\
& w_i \geq 0, \quad \forall i = 1, \ldots, N
\end{aligned} \tag{15}
$$

4.1.4 Lasso Portfolio

The Lasso portfolio is the same as the SAA portfolio with the additional constraint that the sum of the absolute values of the weights is smaller than the chosen value ℓ_1. This ensures that portfolio weights are shrunk and some are set to zero, which reduces the variance in our portfolio returns. The technique is especially promising when many variables are taken into account because then unimportant variables can be shrunk heavily, and the most important variables have a large impact. It is interesting to compare the Lasso portfolio with the PBR portfolio since both techniques aim to reduce the risk of the portfolio. The value for ℓ_1 is computed with a k-fold cross-validation optimization procedure where the Sharpe ratio is optimized, similar to the one used for PBR, which is explained in Section 4.2. Within this procedure, a grid search is used to obtain an optimal value for ℓ_1 for each iteration. Grid search is further explained in Section 4.3.2.

$$
\begin{aligned}
\min_{w \in R^N} \quad & w'\hat{\Sigma}w \\
\text{s.t.} \quad & w'\iota = 1 \\
& (w'\hat{\mu} = \hat{\mu}_{target}) \\
& \sum_{i=1}^{N} |w_i| \leq \ell_1
\end{aligned}
\tag{16}
$$

4.1.5 Ridge Portfolio

The Ridge portfolio is also similar to the SAA portfolio, only now the constraint that the sum of squares of all weights is smaller than a threshold ℓ_2 is added. This implies that the weights are shrunk, but this time, no variable selection is performed (i.e. no weights are set to zero). However, Ridge is stable in the case of multicollinearity, whereas Lasso is not. Because the Ridge method reduces risk in a different way than the Lasso technique, it is also interesting to compare performance with the PBR portfolio. Again, the threshold is optimized with the k-fold cross-validation optimization procedure with regard to the Sharpe ratio and a grid search is used to evaluate different values for ℓ_2 and find the optimal one for each iteration.

$$
\begin{aligned}
\min_{w \in R^N} \quad & w'\hat{\Sigma}w \\
\text{s.t.} \quad & w'\iota = 1 \\
& (w'\hat{\mu} = \hat{\mu}_{target}) \\
& \sum_{i=1}^{N} w_i^2 \leq \ell_2
\end{aligned}
\tag{17}
$$

4.1.6 James Stein Portfolio

In this portfolio, the mean is shrunk towards a target mean. This procedure was introduced by Stein (1956) and aims to reduce estimation errors in the expected returns by shrinking them towards a constant. Since the mean return is not taken into account for the GMV portfolio, it is only relevant for the mean-variance portfolios. Practically this means that in equation 1, instead of the sample mean we use the shrunk mean to obtain optimal portfolio weights. For the covariance matrix, we use the regular sample covariance matrix. The shrunk mean is obtained by multiplying the sample mean and the target mean (μ_0) with a constant which is represented by δ, as in the formula below.

$$\hat{\mu}_{\text{shrunk}} = (1 - \delta)\hat{\mu} + \delta\hat{\mu}_0\iota \tag{18}$$

For this formula, δ is computed as follows.

$$\delta = \min\left\{1, \frac{\frac{N-2}{T}}{(\hat{\mu} - \mu_0\iota)'\Sigma^{-1}(\hat{\mu} - \mu_0\iota)}\right\}$$

We evaluate performance for three target means. Similar to DeMiguel et al. (2009) and Jorion (1986), we shrink the mean towards the GMV mean, so $\mu_0 = \hat{w}'_{gmv}\hat{\mu}$. Additionally, we analyse performance when the target mean is equal to the grand mean of all assets, $\mu_0 = \frac{1}{N}\hat{\mu}'\iota$. Lastly, we analyse the base case where the mean is shrunk towards zero, so the target mean takes on value $\mu_0 = 0$.

4.1.7 Ledoit and Wolf Portfolio

Next to shrinking the mean, we also evaluate performance when the covariance matrix is shrunk. Ledoit and Wolf (2003) find that shrinking the sample matrix towards a single-index covariance matrix estimator performs best in terms of standard deviation. Therefore, we also evaluate this model and compare it to the other models. Complementary to their paper, Ledoit and Wolf (2003) provide Python code[2]. We use their code to compute the shrunk covariance matrix for our data. As for specifics on the procedure, the single-index model assumes returns that are generated by:

$$x_{it} = a_i + \beta_i x_{0t} + \epsilon_{it} \tag{19}$$

With x_{it} an asset return at time t and x_{0t} the market return (in our case the average return of the industry portfolios). The estimator for the covariance matrix of the single-index model is now given by

$$\hat{\Sigma}^* = s_{00}^2 bb' + D \tag{20}$$

Where s_{00}^2 represents the variance of the market returns, b is the vector of regression coefficients and D is the diagonal matrix containing residual variance estimates. Now the sample covariance matrix is shrunk towards this target single-index matrix with shrinkage intensity ϕ as follows.

$$\hat{\Sigma}_{\text{shrunk}} = (1 - \phi)\hat{\Sigma} + \phi\hat{\Sigma}^* \tag{21}$$

[2] Available on: https://www.econ.uzh.ch/en/people/faculty/wolf/publications.htmlProgramming$_{C}ode$

4.2 Establishing Thresholds with K-fold Cross-Validation

In the previous section, we explained that introducing a threshold on the sample variance of the sample variance of the portfolio and taking a rank-1 approximation, implies shrinking the regular mean-variance optimal weights. In order to do this, we need an accurate value for the threshold which is not too low and leads the problem to become infeasible and not too high which would result in a meaningless addition to the model. In order to find optimal thresholds through time, we apply the k-fold cross-validation procedure where the Sharpe ratio is maximized. This is a similar procedure as is used by Ban et al. (2018). The model is evaluated with a rolling window, whereby we train the model with $T_{train} = 120$ observations and test the model on observation $T_{train} + 1$. After this, we roll the window by one period where we again have a training set of 120 observations and this time we test the model on the next observation. With the k-fold cross-validation procedure, we evaluate which threshold to implement based on maximizing the Sharpe ratio for a hold-out sample in sample. In order to do this, we first carefully need to set the search boundaries for the threshold U to obtain a solution that is feasible (not too small) and has effect (not too large).

Let $D = [X_1, \ldots, X_{T_{\text{train}}}] \in \mathbb{R}^{N \times T_{\text{train}}}$ define the training data set of stock returns. This data set is split into k equally sized bins, $D_1, D_2, \ldots, D_k$. Let $P_{-b}(U_{-b})$ denote the PBR problem on the data set $D \setminus D_b$ with corresponding threshold $U = U_{-b}$. We now find the optimal U, denoted by U^*, on the whole train data set D with the following procedure. First, we set a search boundary for U_{-b}, which is $[\underline{U}_{-b}\ , \bar{U}_{-b}]$. When we have established this search boundary, we estimate optimal weights for $P_{-b}(U_{-b})$ for multiple values of U_{-b} With these optimal weights, we compute the Sharpe ratio of the solution on D_b. Then we repeat this process for other U_{-b} and we find the optimal $\text{U}^*_{-b} \in [\underline{U}_{-b}\ , \bar{U}_{-b}]$. We perform this k times, so all bins are left out of the train set one time. Finally, we take the average of the k optimal thresholds, $U^* = \frac{1}{k}\sum_{i=1}^{k} U^*_{-b}$

To be more specific on setting the correct search boundary for $P_{-b}(U_{-b})$, it cannot exceed the search boundary of the entire data set, meaning $[\underline{U}_{-b}\ ,\ \bar{U}_{-b}] \subset [\underline{U}\ , \bar{U}]$. This is because we take the average over the k bins to obtain the optimal threshold that we use for the entire train set, and if this average exceeds the boundaries, it has either no effect (upper boundary) or the problem becomes infeasible (lower boundary). The upper bound on U is as follows.

$$\bar{U} = (\hat{w}'\alpha)^4 \tag{22}$$

Where $\hat{w}$ is the SAA solution without the additional constraint and α is computed as mentioned in equation (8). Implementation of this upper bound would lead to a problem which is the same as the mean-variance problem without the threshold. To find the lower bound on the threshold $\underline{U}$, we compute

$$\begin{aligned} \min_{w \in R^N} \quad & w'\alpha \\ \text{s.t.} \quad & w'\iota = 1 \\ & (w'\hat{\mu} = \hat{\mu}_{target}) \\ & w'\alpha > 0 \end{aligned} \tag{23}$$

The last constraint is added because when weights are unbounded, the objective $w'\alpha$ can take on large negative values and then taking it to the power four gives a very high lower bound for U. This lower bound might then even become larger than the upper bound, which makes it impossible to find an optimal value for the threshold within the bounds. With the positive constraint, this problem is resolved, and the lower bound takes on lower values than the upper bound. For some particular situations, this constraint is not necessary and therefore left out, which is further explained in Section 6.3. To find the upper bound for the partial problem b, we perform a similar procedure as for the upper bound on the entire trainset, only now different data is used. SAA is performed on the trainset $D \backslash D_b$ to obtain $\hat{w}_{-b}$ and it is evaluated whether the upper bound for the subproblem, $(\hat{w}'_{-b}\alpha_{-b})^4$, is below the lower bound of the threshold for the whole train set $\underline{U}$. If this is the case, then $U^*_{-b} = \underline{U}$ and if not, then we set the upper bound as follows.

$$\bar{U}_{-b} = min\,[\bar{U}, (\hat{w}'_{-b}\alpha_{-b})^4] \tag{24}$$

So if the upper bound for subproblem $P_{-b}(U_{-b})$ is greater than the lower bound for the entire train set, it is set to the minimum of the upper bound on the threshold for the whole train set and the upper bound of the train set without bin b.

To find $\underline{U}_{-b}$, we first solve the optimization problem below, where $\hat{\mu}_{-b}$ and α_{-b} are computed on D/D_{-b}.

$$\begin{aligned} \min_{w} \quad & w\ '\alpha_{-b} \\ \text{s.t.} \quad & w'\iota = 1 \\ & (w'\hat{\mu}_{-b} = \hat{\mu}_{target}) \\ & w'\alpha_{-b} > 0 \end{aligned} \tag{25}$$

Then we check whether the obtained lower bound is above the upper bound for the threshold on the whole trainset. If this is the case, the optimal threshold for this bin will be set equal to the upper bound of the threshold for the entire train set. If the lower bound for the subproblem corresponding to bin b is below the upper bound for the train set, we proceed by setting the lower bound for bin b equal to the maximum of the lower bound for problem $P_{-b}(U_{-b})$ and the lower bound on the whole train set, which is shown in the formula below.

$$\underline{U}_{-b} = max(\underline{U}_{-b}, \underline{U}) \tag{26}$$

Pseudocode for the procedure to find bounds for the optimal threshold U is presented on the next page.

Algorithm 1 Pseudocode for obtaining boundaries for thresholds U_{-b} for k bins, for one iteration

Solve SAA in equation (1) on D_{train} to get $\hat{w}_{\max}$
Compute α for D_{train}
Set $\bar{U} = (\hat{w}'_{\max}\alpha)^4$
Solve equation (23) on D_{train} to obtain $\hat{w}_{\min}$
Set $\underline{U} = (\hat{w}'_{\min}\alpha)^4$
Divide up D_{train} randomly into k equal bins, D^b_{train}, $b = 1, \ldots, k$. Let D^{-b}_{train} denote the training data minus the b-th bin.
for $b = 1$ to k **do**
 Solve SAA on D^{-b}_{train} to get $\hat{w}^{-b}_{\max}$
 Compute α_{-b} for D^{-b}_{train}
 Set $\bar{U}_{-b} = (\hat{w}'_{\max}\alpha_{-b})^4$
 if $\bar{U}_{-b} \leq \underline{U}$ **then**
  $U^*_{-b} = \underline{U}$ and terminate
 else
  Solve equation (25) on D^{-b}_{train} to obtain $\hat{w}_{\min}$
  Set $\underline{U}_{-b} = (\hat{w}'_{\min}\alpha_{-b})^4$
 end if
 if $\underline{U}_{-b} \geq \bar{U}$ **then**
  $U^*_{-b} = \bar{U}$ and terminate
 else Compare and update boundaries:
  $\bar{U}_{-b} = min(\bar{U}_{-b}, \bar{U})$
  $\underline{U}_{-b} = max(\underline{U}_{-b}, \underline{U})$
 end if
end for

4.3 Optimization Methods in K-fold Cross-Validation

After having set the search boundaries for problem $P_{-b}(U_{-b})$, we proceed to compute the optimal U^*_{-b} within the boundaries. In order to do this, we compute the Sharpe ratio of the hold-out sample for different values of the threshold. When the Sharpe ratio is maximized, the corresponding threshold is the optimal one for the subproblem. Concretely, we employ three different methods and compare their results. Firstly, we use a backtracking line search algorithm, and thereafter we evaluate the performance while using a grid search and random search. Since we only use the optimization procedures to find an optimal value for the threshold U and not for many parameters simultaneously, these relatively straightforward optimization methods suffice and alternative more extensive methods do not have much added value and are therefore not incorporated.

4.3.1 Backtracking Line Search Algorithm

We apply a similar backtracking line search algorithm as explained by Boyd and Vandenberghe (2004). However, we employ it in a maximization context, whereas they use it in a minimization context. With the backtracking line search algorithm procedure, we try to find the value for the threshold U that optimizes the Sharpe ratio with an algorithm that evaluates step-by-step whether a stopping criterion is met. We first compute the Sharpe ratio when we set the threshold equal to $U - t\Delta U$, with $t\Delta U = t(\bar{U}_{-b} - \underline{U}_{-b})/q$. If this value is above a certain value, equal to the right-hand side of equation (27), then the stopping criterion is directly met, and we take $U - t\Delta(U)$ as the optimal value for the threshold U. If this is not the case and the value is below the right-hand side of the equation, we descend t at a fixed rate. A graphical representation of the procedure is shown in Figure 1, where we strive to find the optimum of the blue line, which represents the Sharpe ratio for a certain threshold U. When the stopping criterion is met, we take the corresponding value of $U - t\Delta(U)$ to be the optimal threshold. We continue with decreasing t as long as the formula below holds.

$$Sharpe(U - t\Delta U) \leq Sharpe(U) + \alpha t\Delta U \frac{dSharpe(U)}{dU} \tag{27}$$

In the formula above, computing the marginal change in the out-of-sample Sharpe ratio is challenging. Therefore, we compute it numerically, similar to Ban et al. (2018). To do this, we apply the chain rule as follows:

$$\frac{dSharpe(U)}{dU} = \nabla_{\hat{w}^*} Sharpe(\hat{w}^*(U))' \left[\frac{d\hat{w}^*(U)}{dU}\right] \tag{28}$$

With $\hat{w}^*(U)$ the optimal regularized solution when the right-hand side is set to U. We know the derivative of the Sharpe ratio with respect to the weight which is as presented below.

$$\nabla_{\hat{w}} Sharpe(\hat{w}) = \frac{(\hat{w}'\Sigma\hat{w})\mu - (\hat{w}'\mu)\Sigma\hat{w}}{(\hat{w}'\Sigma\hat{w})^{\frac{3}{2}}} \tag{29}$$

The other part $\left[\frac{d\hat{w}^*(U)}{dU}\right]$ represents the marginal change in the optimal weight w^* when the threshold U changes. We compute this part by computing the optimal weight $w^*(U)$ for the threshold U and the optimal weight $w^*(U - v)$ for the threshold $U - v$ where v is a small value set between 0 and 1. Mathematically, we obtain the following.

$$\left[\frac{d\hat{w}^*(U)}{dU}\right] \approx \frac{\hat{w}^*(U) - \hat{w}^*((1-v)U)}{v \times U} \tag{30}$$

Now, we can compute the optimal threshold for subproblem b. We evaluate the problem for the case of 2 and 3 bins. Similar to Ban et al. (2018) t is initially set at 1 and then multiplied with fixed rate β, which is set at 0.9. α is set at 0.4, q at 5 and v at 0.05. Furthermore, in practice it turns out that the optimal values for the threshold U are closer to the lower bound than to the upper bound, therefore we set initial value for U to be $U = \frac{\bar{U}_{-b} - \underline{U}_{-b}}{2}$. When we have computed the optimal threshold for all subproblems, we take the average to be the optimal threshold for that particular trainset and compute the out-of-sample

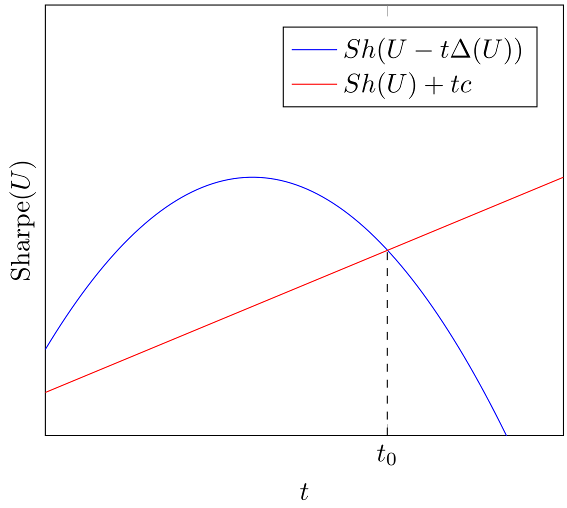

Figure 1: Graphical representation of the backtracking line search algorithm. When initial value for $t < t_0$, the algorithm stops directly, whereas if initial value is larger than t_0, t is backtracked until it is smaller than t_0. c in the formula for the red line represents the constant $\alpha\Delta(U)\frac{dSh(U)}{dU}$

.

optimal weights. Then we repeat this procedure for the next month and so forth. The pseudocode for finding the optimal value for U within the bounds is shown in Algorithm 2 below.

Algorithm 2 Pseudocode for the backtracking line search algorithm to obtain optimal threshold U^*

Choose parameters $\alpha \in (0, 0.5), \beta \in (0, 1)$
Choose stepsize q
Choose peturbation size $v \in (0, 0.5)$
for $b = 1$ to k **do**
  Set $U = \frac{\bar{U}_{-b} - \underline{U}_{-b}}{2}, \Delta U = \frac{\bar{U}_{-b} - \underline{U}_{-b}}{q}, t = 1$
  Solve equation (2) on D^{-b}_{train} to obtain $\hat{w}_{-b}(U)$
  Compute $\frac{dSharpe(U)}{dU} = \nabla_w Sharpe(\hat{w}_{-b}(U))' \left[\frac{d\hat{w}_{-b}(U)}{dU}\right]$
  With
  $\nabla_w Sharpe(\hat{w}_{-b}(U)) = \frac{(\hat{w}'_{-b}\Sigma_{-b}\hat{w}_{-b})\mu_{-b} - (\hat{w}'_{-b}\mu)\Sigma_{-b}\hat{w}_{-b}}{(\hat{w}'_{-b}\Sigma_{-b}\hat{w}_b)^{\frac{3}{2}}}$, and
  $\frac{d\hat{w}^*_{-b}(U)}{dU} = \frac{\hat{w}^*_{-b}(U) - \hat{w}^*_{-b}((1-v)U)}{v \times U}$
  while $Sharpe(U - t\Delta U) < Sharpe(U) + \alpha t \Delta U \frac{dSharpe(U)}{dU}$ **do**
    $t = \beta t$
  end while
end for
Return $U^*_{-b} = U - t\Delta U$

4.3.2 Grid Search and Random Search

Next to the backtracking line search algorithm, we also employ a grid search in order to find the optimal threshold U within the bounds. In this grid search, for every iteration and for every bin, we evaluate the Sharpe ratio for the hold-out bin for different U values within the bounds. Specifically, we evaluate five values, including the lower and the upper bound. For every left-out bin b, the lower and upper bounds and the values in between do differ and the optimal threshold U takes on the average of the optimal thresholds of the different folds. Therefore, more values than five are actually taken into account. Similar to Bergstra and Bengio (2012), we also evaluate the performance of applying random search for hyperparameter optimization. In this case, instead of taking prespecified values to evaluate for U, we take random values within the bounds and analyse for which value the Sharpe ratio is maximized. This time, we evaluate six values for every fold.

4.4 Evaluation Metrics

4.4.1 Sharpe Ratio and Tests

We evaluate and compare the different portfolios on the FF5, FF10 and FF49 datasets. Additionally, we analyse the portfolios for two different cases with regard to individual asset weights. First, the case of no boundaries on the individual asset weights, except for the constraint that the sum of asset weights should equal one. Secondly, we analyse if performance differs when individual asset weights are bounded to be between (-1,1), and the sum of weights should still equal one. This second case is created to guarantee that the portfolio does not take very large positions in one asset and takes short positions in all other assets, thereby still exposing itself to high risk. In order to evaluate and compare the performance of different methods, we take into account the Sharpe ratio, which is presented below.

$$\text{Sharpe Ratio} = \frac{\hat{\mu}_{test} - R_f}{\hat{\sigma}_{test}} \tag{31}$$

For every portfolio, optimal weights are computed for the next month, given the previous 120 months. By multiplying these weights with the asset returns, portfolio returns are obtained for month $T+1$. Then, the risk-free rate is subtracted in order to obtain excess returns. This is done 120 times to obtain 120 out-of-sample portfolio returns. Then, the mean and standard deviations are computed in order to compute the Sharpe ratio. We test the null hypothesis of equal Sharpe ratios for portfolios y and z, which is shown below.

$$H_0 : \frac{\hat{\mu}_y}{\hat{\sigma}_y} - \frac{\hat{\mu}_z}{\hat{\sigma}_z} = 0 \tag{32}$$

With $\hat{\mu}_y$, $\hat{\mu}_z$, $\hat{\sigma}_y$ and $\hat{\sigma}_z$ the mean and standard deviation for portfolio y and z, respectively. In order to test this null hypothesis, there are multiple possible tests available, discussed by Ledoit and Wolf (2008). A test they mention for evaluating the difference in Sharpe ratios is based on HAC inference. This test computes the p-value for the null hypothesis stated in equation (32) as follows.

$$\hat{p} = 2\Phi\left(\frac{-\hat{\Delta}}{s(\hat{\Delta})}\right) \tag{33}$$

With Φ the c.d.f. of the standard normal distribution. Here Δ represents the difference in Sharpe ratios between the two portfolios and $s(\Delta)$ the corresponding standard error which is computed as follows.

$$s(\hat{\Delta}) = \sqrt{\frac{\nabla' f(\hat{v})\hat{\Psi}\nabla f(\hat{v})}{T}} \tag{34}$$

With Ψ the Parzen kernel and $f(a, b, c, d)$ as presented below

$$f(a, b, c, d) = \frac{a}{\sqrt{c - a^2}} - \frac{b}{\sqrt{d - b^2}} \tag{35}$$

And $v = (\mu_y, \mu_z, \gamma_y, \gamma_z)'$, where $\gamma_y = E(r_{1y}^2)$ and $\gamma_z = E(r_{1z}^2)$. For further elaboration on this procedure, we refer the reader to Ledoit and Wolf (2008) who explain the test in detail. They also openly provide code[3] for this test, which is used in this research. Furthermore, we perform the studentized bootstrap method to test for equal Sharpe ratios, also used by Ledoit and Wolf (2008). To generate bootstrap data, we use the circular bootstrap method, where we resample blocks of pairs of observations with replacement. We set the fixed block size at 4. For this test, we create a two-sided bootstrap confidence interval with nominal level $1 - \alpha$ for Δ. If this interval does not contain zero, then H_0 is rejected at nominal level α. Ledoit and Wolf (2008) provide R code for this procedure together with the code for the HAC test.

4.4.2 Turnover and Transaction Costs

Next to analysing the Sharpe ratio, we evaluate turnover because high turnover may result in high transaction costs, which should be minimized. Similar to DeMiguel et al. (2009), we compute the average turnover as follows.

$$\text{Turnover} = \frac{1}{T_{test}} \sum_{t=1}^{T_{test}} \sum_{j=1}^{N} |\hat{w}_{t+1,j} - \hat{w}_{t,j+}| \tag{36}$$

In the formula above, $\hat{w}_{t,j}$ represents the weights in asset j at time t and $\hat{w}_{t,j+}$ represents the portfolio weights before rebalancing at time $t+1$. Concretely, this means that the weights at time t (before returns at time t) are taken, then they are multiplied with the return at time t and scaled so that they sum to 1. Then, these weights are used to evaluate the difference with the new optimal weights at time $t + 1$, presented by $\hat{w}_{t+1,j}$, as is shown in the formula. So in this way the equal weights portfolio also has a non-zero turnover because of the returns that are taken into account.

With the inclusion of turnover, we can also evaluate the different portfolios when transaction costs are incorporated. Kirby and Ostdiek (2012) find that their mean-variance portfolio significantly outperforms the 1/N portfolio absent of transaction costs. However, when transaction costs are included the mean-variance portfolio struggles to significantly outperform the 1/N portfolio. Therefore, it is interesting to also evaluate the impact of transaction costs on our portfolios and their relative performance to benchmark portfolios such as the equally weighted one. Next to this theoretical aspect, including transaction costs in our analysis is also relevant because an investor in real life also encounters transaction costs when making investments. In order to include transaction costs, we start with a certain amount of wealth and

[3]Available on: https://www.econ.uzh.ch/en/people/faculty/wolf/publications.htmlProgramming$_{C}ode$

obtain the wealth of the next month by multiplying this wealth with the portfolio return and subtracting the wealth times the transaction costs made in that period. The procedure is the same as the one used by DeMiguel et al. (2009), and the formula to compute wealth for the next month is as follows.

$$W_{t+1} = W_t(1+R)\left(1 - f \times \sum_{j=1}^{N} |\hat{w}_{t+1,j} - \hat{w}_{t,j}|\right) \tag{37}$$

Here W_t represents the wealth, R the portfolio return and f the constant that symbolizes the transaction costs per unit. The last term is the turnover for a month. We set the proportional transactions costs f equal to 50 basis points, similar to DeMiguel et al. (2009) and Kirby and Ostdiek (2012). The portfolio return after transaction costs is now computed by $(\frac{W_{t+1}}{W_t} - 1) \times 100$. We then use these returns to compute Sharpe ratios and compare these with Sharpe ratios before transaction costs to find the impact of turnover on our portfolios.

5 Model Combination

After evaluating all models separately, we have a good view on which models perform best in terms of Sharpe ratio, returns and risk. It can be expected that certain models are more conservative and perform better in times of crises, whereas other models perform better in times of bull markets where stock prices are rising. Therefore, we aim to capture the best characteristics of different models and combine them into one investment strategy. Specifically, we invest in two models: one model is a relatively conservative model with lower returns and lower volatility that performs well in a crisis and another model performs well during times of surging stock prices. This second model has higher returns but also higher volatility in general. Therefore we need to know for every month in which model to invest and corresponding weights. For this, we use predictions on the state of the economy. To examine the state of the economy, we implement a random forest model with (mostly) US macroeconomic variables as input and a probability on a recession as output. Based on the output, we invest in either the conservative model, the riskier model, or a weighted combination of both. In the next sub-chapters, we elaborate on the specifics of the procedure.

5.1 Random Forest Model

In order to make predictions on the state of the US economy for the upcoming month, we make use of a random forest model, initially proposed by Breiman (2001). The random forest model is especially useful to implement in our case because the relationship between a US recession and macroeconomic variables is complex and non-linear. Furthermore, the random forest model has the capability of capturing interactions between variables, according to Lunetta et al. (2004). This is beneficial since some macroeconomic variables might have marginal individual predictive power, but joint power might be strong. Additionally, the model selects a random subset of features for every split in every tree, thereby ensuring that the correlation between trees is not too high. This aspect is advantageous since it reduces the chances of overfitting and increases robustness to extreme values for one (or multiple) features. Lastly, a random forest model is more transparent than other complex machine learning models (e.g. neural networks).

In terms of the methodology behind the random forest algorithm, it combines multiple decision trees to form a "forest". Every tree delivers a probability on a recession based on the provided values of the features for a data point. The random forest model then averages predictions over all trees B as shown in the formula below, x represents the vector of input variables.

$$\hat{f}_{\mathrm{RF}}(x) = \frac{1}{G}\sum_{g=1}^{G} \hat{f}^{*g}(x) \tag{38}$$

Our model is trained with a binary output variable (US economy in recession or not) and delivers probabilities on a recession as output.

5.2 Feature Evaluation

We first select a wide variety of variables that might be able to indicate whether the US economy will be in a recession next month or not. The variables are selected based on other academic research (e.g. yield curve and GDP are discussed in Gogas et al. (2015)), news articles and domain knowledge. In Table 4, all used variables are shown with the source where we obtained the data. All data is for the US unless indicated otherwise. Next to macroeconomic variables, we add the binary variable of whether the European economy is in a recession to account for the relation between both economies. Further, we incorporate return on the S&P 500 to evaluate the forecasting power of stock market performance. We computed monthly volatility for the S&P 500 based on daily returns. At this point, it is mainly about not missing any variables that might potentially have strong forecasting power because we perform a feature selection where we filter out less important variables later on.

Variable	Source
US Recession	FRED
Inflation	FRED
Unemployment Rate	FRED
Yield Spread[1]	FRED
Real GDP Growth[2]	FRED
EU Recession	FRED
Industrial Production	FRED
Personal Consumption Expenditures	FRED
Consumer Sentiment	FRED
Macroeconomic Uncertainty 1M ahead estimate	FRED
Macroeconomic Uncertainty 12M ahead estimate	FRED
Financial Uncertainty	FRED
Business Tendency (Manufucturing) Confidence Indicator	FRED
Composite Leading Indicator	FRED
Mortgage Rate (30-year fixed)	FRED
S&P Returns	WRDS[3]
S&P Volatility	WRDS

Table 4: Variables used in the random forest model and their source. 1) Yield spread refers to the 10-year US Treasury Constant Maturity minus the 2-year US Treasury Constant Maturity. 2) The FRED provides real GDP quarterly data, we computed growth rates and extrapolated to monthly data manually. 3) WRDS refers to the Wharton Research Data Services.

In order to decide how many lags to include for all features, we first compute the correlation of the lagged values of the features with the US recession (and the autocorrelation for the US recession). We then include the lags with the highest correlation in our model. After this, we perform hyperparameter optimization for the first time (see next paragraph for an explanation). We then evaluate feature performance. Similar to Jabeur et al. (2021), we use Shapley additive explanations (SHAP) values for this. The idea behind SHAP values is to show the impact of each feature on the model output. For every forecast of the model, the impact of every feature is represented by the SHAP value. This can be both positive and negative, and the absolute value provides us with information on the magnitude of the impact of the feature for that observation.

To select the most important features from all the features we initially included in our model, we perform time series cross-validation in-sample and compute SHAP values for the left-out fold. We then take the average of the absolute values for every feature and then average over the different folds to obtain the impact per feature based on the trainset. The formula to compute SHAP values for every feature j as described by Lundberg et al. (2018) is as follows:

$$\zeta_j = \sum_{S \subseteq N \setminus \{h\}} \frac{|S|! \cdot (M - |S| - 1)!}{M!} \left[f_x(S \cup \{h\}) - f_x(S) \right] \tag{39}$$

Here, we sum over the subset of all features, every time excluding feature h. $\frac{|S|! \cdot (M-|S|-1)!}{M!}$ computes the weight of each subset, where $|S|$ is the cardinality of subset S and M the total number of features. The expression $[f_x(S \cup \{h\}) - f_x(S)]$ calculates the difference in the model's prediction when feature h is included versus when it is excluded, reflecting the feature's impact. In the context of more complex models such as random forests, direct computation of SHAP values as per the mentioned formula can be computationally extensive. This is due to the necessity of evaluating the model's prediction for every possible combination of features, which grows exponentially with the number of features. Therefore, Lundberg et al. (2018) propose an estimation approach, which we use as well. Specifically, we make use of the TreeSHAP algorithm.

Based on the outcome of the SHAP value computation, we select certain features that contribute heavily to the predictions and leave out features that have negligible impact. We do this to increase model interpretability and decrease complexity and chances on overfitting. After having performed the feature selection, we again optimize hyperparameters and then compute out-of-sample SHAP values. This time, next to evaluating the relative importance of the selected features, we also show the relation between the features and the output variables (i.e. high value of a feature might hint on a high or low probability on a recession).

5.3 Hyperparameter Optimization

Probst et al. (2019) state that random forest models work reasonably well with the default parameters in coding programs. However, we still optimize hyperparameters to obtain optimal performance. Probst et al. (2019) also present an overview of the hyperparameters in a random forest model, which are number of trees, number of features considered when looking for the best split, number of observations drawn for each tree, whether observations are drawn with or without replacement, minimum number of observations in a terminal node and splitting criteria in the nodes. Additionally, we also take into account the maximum depth and the minimum number of samples to split an internal node. Number of observations drawn for each tree is set to the entire sample, observations are drawn with replacement and splitting criteria is Gini. The other hyperparameters are optimized with a grid search.

We first optimize hyperparameters in a model that includes all features, then we perform feature selection, and then we again optimize hyperparameters for the model with the reduced number of features to obtain our final optimized hyperparameters. Then, we run the model with selected features and optimal hyperparameters to obtain results and evaluate performance. We optimize hyperparameters both times with time series cross-validation on the train set with k = 3 (so that every period contains both expansion and recession data) and recall of the recession is used as optimization criterion in the grid search. We use recall because it is especially important to not miss a recession with our forecasts. We further elaborate on the importance of this metric in the model evaluation Section.

5.4 Weighting Schemes

We apply multiple weighing schemes, namely the binary, optimized, incremental, recession avoidance and short sell scheme. For the optimized scheme, we do not yet know the weights since they depend on an optimization procedure. The weights of the other schemes are shown in Table 5. The weighting schemes are elaborated on in the sections below.

Probability	< 0.1	0.1 - 0.2	0.2 - 0.3	0.3 - 0.4	0.4 - 0.5	0.5 - 0.6	0.6 - 0.7	0.7 - 0.8	0.8 - 0.9	0.9 - 1.0
WS_B	0 (1.0)	0 (1.0)	0 (1.0)	0 (1.0)	0 (1.0)	1.0 (0)	1.0 (0)	1.0 (0)	1.0 (0)	1.0 (0)
WS_I	0.1 (0.9)	0.2 (0.8)	0.3 (0.7)	0.4 (0.6)	0.5 (0.5)	0.6 (0.4)	0.7 (0.3)	0.8 (0.2)	0.9 (0.1)	1.0 (0)
WS_{RA}	0 (1.0)	0 (1.0)	0 (1.0)	0 (1.0)	0 (1.0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
WS_S	0 (1.0)	0 (1.0)	0 (1.0)	0 (1.0)	0 (1.0)	0 (-1.0)	0 (-1.0)	0 (-1.0)	0 (-1.0)	0 (-1.0)

Table 5: Weights in conservative and risky portfolio (between brackets) for different probability outputs of the random forest model. WS_B, WS_I, WS_{RA} and WS_S represent the binary, incremental, recession avoidance and short sell weighting scheme, respectively. The in-sample optimized weighting scheme is not shown here.

5.4.1 Binary Scheme

Initially, we implement a base model where we invest our full capital (weight 1) in the aggressive model (MV PBR-LS model with k=2 and monthly target return 1.2%) when we predict the economy to be in an expansion and invest our full capital in the conservative model (GMV PBR-GS$^*_{BO}$ with k=3) when we expect the economy to be in a recession. This method is supposed to benefit from times when general stock prices are rising, especially obtaining higher returns than the conservative model. At the same time, this method is expected to minimize losses when stock prices decrease in a recession. We set the threshold for the recession probabilities at 0.5, thereby rounding probabilities above 0.5 to 1 indicating a recession and probabilities below 0.5 to 0, implying a normal state or expansion. A more risk-averse investor would set a lower threshold, whereas a risk-seeking investor might set it higher. Risk preferences however, are captured with the other weighting schemes and therefore we choose to set the threshold here at the logical base value 0.5.

5.4.2 Optimized Scheme

Next to this binary weighting scheme, we also evaluate a more advanced flexible weighting scheme. In order to determine optimal weights per probability interval, we make use of an optimization method where we find weights that optimize the Sharpe ratio of the combined portfolio. Specifically, we use time series cross-validation, similar to Ramos and Oliveira (2016). The idea behind this procedure is that we need to train our model with data that precedes the left-out fold and not vice versa since we deal with time series data. Therefore, in this method one first uses train data and a left-out fold and then extends the train data with successive observations and again uses a left-out fold of the same size as the previous one. Ramos and Oliveira (2016) extend the train set with one observation each time, but they use the procedure to compute MSPE for a forecast. In our case, where we want to construct an optimal weighting scheme, we increase the left-out fold.

The concrete procedure we apply is as follows. We first train the random forest model on 128 observations (September 1976 until April 1987), then obtain probabilities on a recession for the next 100 observations (May 1987 until August 1995) with the trained model. Then, we obtain returns for the conservative and risky portfolio for these 100 observations (taking the FF5 data further back than previously). Then we apply an optimization procedure where we specify probability intervals of 0.05 (so 0-0.05, 0.05-0.1, etc.) and determine optimal combination weights for each interval, with the Sharpe ratio as evaluation criterion. Hence, for every probability interval from the random forest model, we obtain the combination weight that optimizes the Sharpe ratio for our combined portfolio in-sample. Since we use time series cross-validation, we then include the 100 observations in the train set, so we now train the random forest model again with 228 observations (from September 1976 until August 1995) and obtain recession probabilities for the last 100 observations (September 1995 until December 2003). Then, we again optimize combination weights for every probability interval of 0.05. Lastly, for every probability interval, we take the average of the optimal weights from both left-out folds.

If there were no output values for one of the random forest models for a certain probability interval (e.g. no probabilities above 0.95) then we take the optimal weights of the other random forest model, and if there were no output values for both models, then we set the optimal combination weight for this interval to not available. We have chosen relatively large kept-out folds in the time series cross-validation since the main focus is on finding optimal weights given probability outputs from the random forest model. Therefore, we need enough observations and periods that have months of both recession and expansion.

Having obtained the optimal combination weights for probability intervals in-sample, we then use these combination weights to construct a combination portfolio for every month out of sample (January 2004 until December 2013). We again train the random forest model, now with all the 328 observations in the train set (September 1976 until December 2003). We then obtain recession probabilities for the out-of-sample months and form the combination portfolio based on the optimal combination weights that are specified per interval. If the optimal combination weights for a certain interval from the optimization procedure were set to not available, then we invest 1 in the risky portfolio if the probability is smaller than 0.5 and 1 in the conservative portfolio if the recession probability is larger than 0.5.

5.4.3 Incremental, Recession Avoidance and Short Sell Scheme

We evaluate three more investment schemes, as seen in Table 5. The first one is a basic incremental weighting scheme (WS_I in Table) where we increase weights in the conservative portfolio with 0.1 (and decrease weight in the risky portfolio with 0.1) for every 0.1 probability increase. Furthermore, we evaluate two unconventional investment strategies where we release the constraint that combination weights should sum to one at all times. The first is the recession avoidance strategy, where we choose not to invest in either one of the portfolios if we expect there to be a recession in the next period. This makes sense in the way that we also expect the conservative portfolio to make losses in a recession, although they are smaller. In Table 5, this recession avoidance strategy is represented by WS_{RA}. Lastly, we analyse the performance of a portfolio where we short sell the risky portfolio in case of an expected recession and go long in case of an expected expansion. The WS_S portfolio in the table reflects the combination weights in this strategy.

5.5 Model Evaluation

With our random forest model, we want to accurately predict whether the economy will be in a recession since having this wrong might open us up to large potential losses. At the same time we do not want to forecast a recession way too often since this might cause missing out on higher returns. In order to evaluate the performance of the random forest model, we make use of multiple evaluation metrics. We present a confusion matrix that shows how many forecasts are accurate and how many are not. Based on this, we can compute both recall and precision. Recall is given by $\frac{\mathrm{TP}}{\mathrm{TP+FN}}$ where TP represents true positives and FN false negatives. Especially the recall of the recession is a relevant metric since it is harmful if we forecast an expansion and there is a recession because then we invest in the riskier model during a recession. Because of this, we also use the recall of the recession as an optimization criterion in the hyperparameter grid search. If we predict a recession and there turns out to be an expansion, this is not beneficial since the aggressive portfolio would probably obtain a higher return, but it is less harmful since the conservative model will most likely still make a decent return.

Next to recall, we also look at precision which is given by $\frac{\mathrm{TP}}{\mathrm{TP+FP}}$. This represents how many times we predict a recession or an expansion right as a fraction of the total times we forecast a recession or an expansion. Moreover, we compute the accuracy, which is $\frac{\mathrm{TP\ +\ TN}}{\mathrm{TP+FP+TN+FN}}$. Accuracy however, is slightly less important since we have much more data points in an expansion than in a recession and we particularly want to forecast the recession right. After having determined how well our model forecasts the state of the economy, we can evaluate the performance of our combined investment strategy in terms of Sharpe ratio, return, volatility and turnover. Additionally, we evaluate wealth increase over time.

6 Performance-Based Regularization Results

6.1 Base Models

In Table 6, Sharpe ratios for the different portfolios can be observed. The bold values represent the highest Sharpe ratio for the associated optimization problem. The values between brackets represent the p-values of the HAC test when the corresponding Sharpe ratio is compared with the Sharpe ratio of the SAA portfolio. Since HAC estimators might have poor small-sample properties, p-values of the studentized bootstrap method are also presented in the Appendix in Table 19. These values are more conservative, which is in line with the research of Ledoit and Wolf (2008). Furthermore, for the James Stein (JS) portfolios, taking a target mean equal to zero turned out to perform best in almost all cases. Therefore, all presented James Stein portfolios are with a target mean of zero.

From the Table, we observe that the PBR portfolios have higher Sharpe ratios than the other portfolios for the FF5 and FF10 datasets. However, the difference with the SAA portfolio is only significant for the FF5 PBR-LS and PBR-GS portfolio (three bins) with a monthly target return of 0.8%. The PBR-LS portfolio, where three bins are applied in the k-fold cross-validation with a monthly target return of 1.0% over the FF5 dataset, has the highest Sharpe ratio of all portfolios. This is however partly due to the fact that the mean-variance portfolio with target return 1.0% already has higher Sharpe ratio than the SAA mean-variance portfolios with other target returns and the SAA GMV portfolio.

Often, using three bins in the k-fold cross-validation procedure provides a slightly higher Sharpe ratio. However, the differences are minor and it is not always the case so it is difficult to state that this always supersedes using two bins. Moreover, applying the backtracking line search algorithm or grid search for optimizing the threshold U does not have a large impact on portfolio performance since Sharpe ratios do not differ much. The NS (no short selling) portfolio performs worst in terms of Sharpe ratio for the FF5 and FF10 datasets, which seems logical since it is much more restricted than the other portfolios in finding optimal weights. Furthermore, for the FF10 dataset Sharpe ratios are lower than for the FF5 dataset, which can be expected with more diversification. For the FF10 dataset, no portfolio significantly outperforms its corresponding SAA portfolio.

For the FF49 dataset, we observe that Sharpe ratios are much lower than for the other two datasets. Now, the no short selling portfolio performs relatively well compared to the other portfolios. No portfolio significantly outperforms the SAA portfolio. Hence, it can be concluded that applying PBR works particularly well for a dataset with fewer assets and is not advantageous when a high number of assets is taken into consideration. Therefore, the main takeaway from Table 6 for an investor would be to focus on a dataset with fewer assets and then apply PBR to the SAA portfolio in order to increase the Sharpe ratio by both increasing average return and decreasing portfolio risk. The backtracking line search algorithm seems to deliver slightly higher Sharpe ratios than the grid search, although the differences are small.

In Table 7, the average turnover for the different portfolios is presented. We notice that the turnover values for the FF5 and FF10 industry datasets are approximately similar, and those for the FF49 dataset are much higher, which could be expected due to the much higher number of assets which all have weights in the portfolio. Furthermore, It can be observed that in general the turnovers for the portfolios based on the GMV portfolio are much lower than those for the mean-variance portfolios. This can be explained because the mean-variance portfolios have an additional restriction which might force the portfolios to take more extreme positions as opposed to the GMV portfolio. The turnover values for the PBR portfolios are generally higher than for the SAA portfolios but not much higher, which seems promising since their Sharpe ratios are sometimes significantly higher. Moreover, it can be seen that the turnover values for the no short-selling portfolio are the lowest, which makes sense since weight differences over rebalancing periods are more limited than for the other portfolios.

	FF 5 industry		FF 10 industry		FF 49 industry	
MV, TR = 0.8%						
	2 bins	3 bins	2 bins	3 bins	2 bins	3 bins
PBR - LS	**1.2141** (0.0416)	1.1879 (0.0458)	1.1070 (0.5666)	**1.1084** (0.5426)	0.3353 (0.7445)	0.3341 (0.6872)
PBR - GS	1.1558 (0.0629)	1.1477 (0.04595)	1.0974 (0.4321)	1.1015 (0.4340)	0.3476 (0.9690)	0.3408 (0.7549)
L1	0.9023 (0.4869)	0.8180 (0.2104)	0.9735 (0.3163)	0.9726 (0.2676)	0.4598 (0.3790)	0.4552 (0.3986)
L2	0.9283 (0.6238)	0.9596 (0.6416)	0.9817 (0.3175)	1.0261 (0.6456)	0.3965 (0.5564)	0.3765 (0.7205)
SAA	1.0114		1.0562		0.3493	
NS	0.6255 (0.1145)		0.8332 (0.3474)		**0.5698**(0.4495)	
JS	1.0586 (0.7813)		0.7892 (0.4699)		-0.2435 (0.0498)	
LW	0.9997 (0.2207)		1.0311 (0.2146)		0.4163 (0.5290)	
MV, TR = 1.0%						
	2 bins	3 bins	2 bins	3 bins	2 bins	3 bins
PBR - LS	1.2186 (0.1041)	**1.2306** (0.0541)	1.0502 (0.6215)	**1.0575** (0.5436)	0.2950 (0.8069)	0.2897 (0.6778)
PBR - GS	1.1585 (0.1803)	1.2110 (0.0407)	1.0487 (0.4574)	1.0570 (0.3758)	0.3070 (0.9734)	0.2945 (0.6952)
L1	0.8708 (0.2629)	0.9088 (0.3156)	0.9254 (0.1599)	0.9170 (0.1905)	0.3922 (0.5057)	0.3925 (0.5029)
L2	0.8776 (0.2056)	0.8469 (0.1778)	0.93603 (0.2946)	0.9737 (0.5543)	0.3242 (0.7603)	0.3298 (0.6425)
SAA	1.0731		1.0098		0.3056	
NS	0.5597 (0.0309)		0.5794 (0.1200)		**0.5477**(0.4369)	
JS	0.9269 (0.4518)		0.6517 (0.2985)		-0.2650 (0.0488)	
LW	1.0628 (0.3043)		0.9950 (0.4663)		0.3692 (0.5657)	
MV, TR = 1.2%						
	2 bins	3 bins	2 bins	3 bins	2 bins	3 bins
PBR - LS	**1.1241** (0.0898)	1.1202 (0.0770)	0.9586 (0.6551)	0.9753 (0.4720)	0.2543 (0.8703)	0.2444 (0.6566)
PBR - GS	1.0827 (0.0963)	1.1037 (0.0593)	0.9756 (0.3669)	**0.9787** (0.3193)	0.2509 (0.7892)	0.2474 (0.6338)
L1	0.8357 (0.4168)	0.7799 (0.1330)	0.8263 (0.1760)	0.8535 (0.3840)	0.3013 (0.7599)	0.3110 (0.7067)
L2	0.7745 (0.1286)	0.8371 (0.2315)	0.8572 (0.3401)	0.9109 (0.7998)	0.2717 (0.8362)	0.2705 (0.8338)
SAA	1.0046		0.9270		0.2612	
NS	0.5114 (0.0631)		0.5125 (0.1953)		**0.5118**(0.4267)	
JS	0.8134 (0.2509)		0.5568 (0.2230)		-0.2792 (0.0501)	
LW	0.9956 (0.3344)		0.9209 (0.7796)		0.3219 (0.6026)	
GMV						
	2 bins	3 bins	2 bins	3 bins	2 bins	3 bins
PBR - LS	1.1944 (0.0955)	**1.2178** (0.0572)	1.0512 (0.6297)	1.0754 (0.4271)	0.3757 (0.7386)	0.3761 (0.7009)
PBR - GS	1.1014 (0.3148)	1.1708 (0.0631)	1.0321 (0.5976)	**1.0931** (0.1327)	0.3552 (0.4391)	0.3799 (0.7336)
L1	0.9546 (0.2610)	0.9709 (0.2358)	0.9346 (0.5252)	0.9704 (0.6605)	0.5561 (0.5761)	0.4613 (0.7783)
L2	1.0063 (0.5911)	0.9937 (0.5563)	0.9697 (0.7664)	0.9298 (0.4786)	0.6122 (0.0971)	0.5972 (0.1264)
SAA	1.0296		1.0000		0.3897	
NS	0.8216 (0.0755)		0.8753 (0.4325)		**0.6284**(0.3570)	
LW	1.0186 (0.2002)		0.9840 (0.4170)		0.4752 (0.3960)	
EW	0.5596		0.6033		0.5595	

Table 6: Annualized out-of-sample Sharpe ratios for the portfolios discussed in section 4.1 for three datasets. Values between brackets represent p-values of the HAC test where the difference in Sharpe ratio with the corresponding SAA (sample average approximation) portfolio is evaluated. The highest SR for every optimization problem and dataset is highlighted in bold. TR represents monthly target return, PBR-LS refers to the usage of the line search optimization method for threshold U and PBR-GS to the grid search. Other abbreviations are Lasso (L1), Ridge (L2), no short selling (NS), Ledoit and Wolf (LW) and equally weighted (EW).

	FF 5 industry		FF 10 industry		FF 49 industry	
MV, TR = 0.8%						
	2 bins	3 bins	2 bins	3 bins	2 bins	3 bins
PBR - LS	0.3246	0.3067	0.2720	0.2707	0.7645	0.7543
PBR - GS	0.3434	0.3196	0.3202	0.2912	0.8390	0.7851
L1	0.4434	0.3940	0.3040	0.2948	0.3713	0.3739
L2	0.3577	0.4260	0.2545	0.2852	0.5817	0.6181
SAA	0.2816		0.2400		0.7284	
NS	0.3541		0.2109		0.1960	
JS	0.4437		1.0252		33.4483	
LW	0.2803		0.2256		0.3401	
MV, TR = 1.0%						
	2 bins	3 bins	2 bins	3 bins	2 bins	3 bins
PBR - LS	0.3261	0.3043	0.2835	0.27622	0.8009	0.7930
PBR - GS	0.3427	0.3061	0.3216	0.3036	0.8959	0.8224
L1	0.3590	0.3451	0.2957	0.2753	0.4368	0.4090
L2	0.3173	0.3233	0.2696	0.2726	0.6526	0.6995
SAA	0.2697		0.2467		0.7636	
NS	0.1428		0.1940		0.1706	
JS	0.5984		1.3553		15.1810	
LW	0.2661		0.2287		0.3700	
MV, TR = 1.2%						
	2 bins	3 bins	2 bins	3 bins	2 bins	3 bins
PBR - LS	0.3854	0.3771	0.3612	0.3519	0.8743	0.8601
PBR - GS	0.4033	0.4013	0.4253	0.4011	0.9340	0.8908
L1	0.5616	0.5838	0.4537	0.4190	0.5338	0.5124
L2	0.4723	0.4716	0.3762	0.3651	0.7511	0.7909
SAA	0.3472		0.3355		0.8295	
NS	0.1016		0.1594		0.1971	
JS	0.7862		1.6981		15.3395	
LW	0.3428		0.3112		0.4218	
GMV						
	2 bins	3 bins	2 bins	3 bins	2 bins	3 bins
PBR - LS	0.1349	0.1327	0.1799	0.1778	0.7318	0.7172
PBR - GS	0.1501	0.1386	0.2400	0.1960	0.7876	0.7530
L1	0.1154	0.1031	0.1311	0.1253	0.3805	0.3935
L2	0.0977	0.1445	0.1146	0.0818	0.3229	0.2659
SAA	0.0902		0.1375		0.6938	
NS	0.0389		0.0534		0.0919	
LW	0.0878		0.1222		0.2993	
EW	0.0239		0.0297		0.0397	

Table 7: Turnover values for the portfolios discussed in section 4.1 for three datasets. Abbreviations can be found earlier in the text and in Table 27 in the Appendix.

6.2 Bounded Optimization Problem

Until now, we have not restricted weights to form optimal portfolios. In Table 8, Sharpe ratios are shown for the different portfolios in both the mean-variance context with monthly target return 1.0% and the GMV context for the FF5 dataset. This time however, the weights are bounded to be between (-1,1). The MV context with a target return of 1.0% is chosen to further evaluate because it is superior in terms of Sharpe ratio to the other MV portfolios. Additionally, the GMV portfolio is also interesting to look into further because of its sole focus on variance.

It can be observed that the SAA portfolios do both have slightly higher Sharpe ratios than their unbounded versions. In the mean-variance context, the PBR portfolios do not perform much better than their unbounded versions, and they do not significantly outperform the bounded MV SAA portfolio. In the GMV context however, the PBR portfolios do perform better and now both PBR portfolios significantly outperform the SAA portfolio when three bins are used for the cross-validation. Therefore, the PBR portfolio seems especially promising in a bounded GMV context. In Table 8, turnover values for the bounded portfolios are shown as well. It can be observed that turnover values for the bounded GMV portfolio and its variations do not differ much from their unbounded versions. Turnover values for the mean-variance portfolio and its variations however, are larger for the bounded alternative.

FF 5 industry				
MV, TR = 1.0%	Sharpe Ratio		Turnover	
	2 bins	3 bins	2 bins	3 bins
PBR - LS	1.1649 (0.6468)	1.2386 (0.1665)	0.3950	0.3719
PBR - GS	1.1932 (0.1802)	**1.2464** (0.0848)	0.3951	0.3489
L1	0.8346 (0.1568)	0.8065 (0.1071)	0.4522	0.4569
L2	0.8687 (0.0690)	0.8191 (0.0236)	0.3589	0.3707
SAA	1.1136		0.3121	
NS	0.5597 (0.5584)		0.1428	
JS	0.3752 (0.0002)		1.0467	
LW	1.1054 (0.1805)		0.3110	
GMV				
	2 bins	3 bins	2 bins	3 bins
PBR - LS	1.2168 (0.0721)	1.2411 (0.0468)	0.1300	0.1232
PBR - GS	1.1470 (0.1235)	**1.2538** (0.0431)	0.1486	0.1388
L1	0.9507 (0.1575)	0.9676 (0.1150)	0.1078	0.1019
L2	1.0063 (0.4075)	0.9937 (0.4273)	0.0977	0.1445
SAA	1.0410		0.0861	
NS	0.8216 (0.0559)		0.0389	
LW	0.9841 (0.6532)		0.1222	

Table 8: Sharpe ratios and turnover values for the portfolios from section 4.1 with individual asset weights now bounded by (-1,1).

6.3 Additional Optimization Methods for Threshold U

We have now evaluated a wide variety of unbounded portfolios and some bounded portfolios. Since the PBR method performed especially well for the bounded GMV portfolio on the FF5 dataset, we further analyse this particular optimization problem. Firstly, next to the already implemented optimization methods backtracking line search and grid search, we implement a random search (RS) for optimization of the threshold U. Secondly, we slightly modify our model by releasing the restriction that $w'\alpha$ should be greater than zero, which is used for computing the lower bound on the threshold until now (see equation (23) in Section 4). This restriction was needed because for unbounded weights and large datasets, minimizing $w'\alpha$ can lead to strong negative values. Then taking this value to the power four provides us with a very high lower bound, sometimes even higher than the upper bound, which makes the PBR framework useless. For the bounded version and with 5 assets, this is not the case and therefore we can release the constraint. In Table 9, the bounded PBR portfolios with different optimization methods are presented, both when the positive restriction on $w'\alpha$ is applied and when it is released (asterisk).

It can be observed that for the portfolio where grid search is applied and the positive constraint on $w'\alpha$ is released, the highest Sharpe ratio is obtained. It is the highest Sharpe ratio until now of all evaluated portfolios, and it significantly outperforms the corresponding GMV SAA portfolio. The random search portfolios do not deliver higher Sharpe ratios than the other optimization methods, but p-values are slightly lower, which is related to the characteristics of the returns and of the HAC test. Releasing the restriction increases the Sharpe ratio moderately, only for the PBR-GS portfolio the increase is a bit larger with 0.025. Interesting to see is that all PBR portfolios with three bins outperform the SAA portfolio, whereas all portfolios with two bins do not significantly outperform the SAA portfolio, implying that using three bins in the k-fold cross-validation leads to better portfolio performance in this context. In the Table, the turnover values for the different portfolios are shown as well, and it can be observed that the turnover values for the random search method are slightly higher than for the other two optimization methods and the turnover value for the GMV SAA portfolio is lowest of all.

FF 5 industry				
GMV	Sharpe Ratio		Turnover	
	2 bins	3 bins	2 bins	3 bins
PBR - LS_{BO}	1.2168 (0.0721)	1.2411 (0.0468)	0.1300	0.1232
PBR - LS^*_{BO}	1.2172 (0.0749)	1.2431 (0.04602)	0.1306	0.1229
PBR - GS_{BO}	1.1470 (0.1235)	1.2538 (0.0431)	0.1486	0.1388
PBR - GS^*_{BO}	1.1487 (0.1863)	**1.2788**(0.0359)	0.1576	0.1327
PBR - RS_{BO}	1.1698 (0.0989)	1.2419 (0.0343)	0.2107	0.1605
PBR - RS^*_{BO}	1.1517 (0.1414)	1.2440 (0.0284)	0.1976	0.1648
SAA_{BO}	1.0410		0.0861	

Table 9: Sharpe ratios and turnover values for the bounded GMV PBR portfolios based on backtracking line search (LS), grid search (GS) and random search (RS) optimization. $*$ marks that the $w'\alpha > 0$ restriction is released for the corresponding portfolio and BO represents bounded weights.

6.4 Mean-Variance Optimization with 1/N Target Return

Earlier, we evaluated mean-variance portfolios with monthly target returns 0.8%, 1.0% and 1.2%. Now we have computed a portfolio that takes into account a time-varying monthly target return, which is equal to the expected return on the 1/N portfolio. The results can be observed in Table 10. It can be observed that the regular SAA portfolio has a higher Sharpe ratio than the mean-variance SAA portfolios with a monthly target return of 0.8% and 1.2% in Table 6. However, the Sharpe ratio is lower than for the mean-variance portfolio with a target return of 1.0%. Interesting to see in Table 10 is that there are major improvements from the SAA portfolio when implementing PBR. This time, the unbounded PBR-LS method delivers the highest Sharpe ratio when 3 bins are used. Furthermore, four of the eight portfolios significantly outperform the SAA portfolio.

When we further look at Table 10, we find that turnover values for the SAA portfolios are lower than turnover values for the mean-variance portfolios with a fixed target return (Tables 7 and 8). Also, the PBR portfolios have lower turnovers than their counterparts with fixed target return. Turnovers are still higher than for the GMV-based PBR portfolios, however. In Figure 9 in the Appendix, monthly target returns over time are presented. It can be observed that they are initially high (1.1%), and then there is a steep decline (to 0.1%) caused by relatively low returns on the FF5 assets between 1998 and 2008. After that, there is again a rise in the target return (to 0.8%).

FF 5 industry				
MV	Sharpe Ratio		Turnover	
	2 bins	3 bins	2 bins	3 bins
PBR - LS	1.2472 (0.0243)	**1.2613** (0.0277)	0.2220	0.2237
PBR - LS_{BO}	1.2094 (0.0699)	1.2156 (0.1065)	0.2613	0.2821
PBR - GS	1.1238 (0.1117)	1.1779 (0.0334)	0.2196	0.2347
PBR - GS_{BO}	1.1071 (0.1198)	1.2111 (0.0140)	0.2546	0.2235
SAA	1.0326		0.1648	

Table 10: Sharpe ratios and turnover values for MV PBR portfolios with time-varying target return equal to the expected return on the 1/N portfolio. The SAA portfolio is shown as well for comparison purposes.

6.5 Model Performance with Transaction Costs

We also evaluated the impact of transaction costs on our best-performing models and some benchmark portfolios. Specifically, we evaluate the 1/N, bounded SAA GMV and unbounded SAA MV with 1/N target return portfolios as benchmarks. Furthermore, we analyse the impact of transaction costs on our best-performing portfolios, which are the bounded PBR-GS portfolio with released restriction and the unbounded PBR-LS portfolio with 1/N target return.

In Table 11, results are shown with p-values when the Sharpe ratio is compared to the Sharpe ratio of the 1/N portfolio. As discussed earlier, other research has found that it is difficult to outperform the 1/N portfolio when transaction costs are included. In our case, the 1/N portfolio has a much lower

Sharpe ratio when transaction costs are not included, and hence, all four portfolios outperform the 1/N portfolio significantly. Including transaction costs does decrease the Sharpe ratios of the four portfolios more than that of the 1/N portfolio because of the higher turnover, but the portfolios still all significantly outperform the 1/N portfolio. In Table 23 in the Appendix, the results are presented with studentized bootstrap p-values instead of HAC p-values.

We also compared our two best-performing portfolios with their SAA counterparts when transaction costs were included and evaluated if the differences in Sharpe ratio were still significant. The results are shown in Table 12. Note that the Sharpe ratios are exactly the same, but the p-values between brackets differ because now the portfolios are compared to the SAA portfolios. It can be observed that the GMV PBR-GS$^*_{BO}$ with 3 bins does not significantly outperform its SAA counterpart anymore when transaction costs are included. The MV PBR-LS$_{EW}$ portfolio still outperforms its SAA counterpart when transaction costs are included, despite an increase in p-values.

FF 5 industry						
Portfolio	SR excl. TC		TO		SR incl. TC	
	2 bins	3 bins	2 bins	3 bins	2 bins	3 bins
GMV PBR - GS$^*_{BO}$	1.1488 (0.0050)	**1.2788**(0.0043)	0.1576	0.1327	1.0850 (0.0081)	**1.2263**(0.0057)
MV PBR - LS$_{EW}$	1.2472 (0.0019)	1.2613 (0.0022)	0.2220	0.2237	1.1565 (0.0058)	1.1668 (0.0067)
1/N	0.5596		0.0239		0.5396	
SAA GMV$_{BO}$	1.0410 (0.0029)		0.0861		1.0115 (0.0032)	
SAA MV$_{EW}$	1.0326 (0.0008)		0.1648		0.9584 (0.0036)	

Table 11: Sharpe ratios (SR) before and after transaction costs (TC) are shown. Also turnover (TO) is presented. The *EW* subscript represents a target return equal to the 1/N portfolio. HAC p-values are shown versus the 1/N portfolio.

FF 5 industry						
Portfolio	SR excl. TC		TO		SR incl. TC	
	2 bins	3 bins	2 bins	3 bins	2 bins	3 bins
GMV PBR - GS$^*_{BO}$	1.1488 (0.1863)	**1.2788**(0.0359)	0.1576	0.1327	1.0850 (0.3843)	**1.2263**(0.0559)
MV PBR - LS$_{EW}$	1.2472 (0.0243)	1.2613 (0.0277)	0.2220	0.2237	1.1565 (0.0340)	1.1668 (0.0418)
1/N	0.5596		0.0239		0.5396	
SAA GMV$_{BO}$	1.0410		0.0861		1.0115	
SAA MV$_{EW}$	1.0326		0.1648		0.9584	

Table 12: Sharpe ratios before and after transaction costs are shown. Also turnover is presented. The *EW* subscript represents a target return equal to the 1/N portfolio. HAC p-values are shown versus the corresponding SAA portfolio.

6.6 Wealth Accumulation

From all the portfolios we evaluated, quite a few PBR portfolios significantly outperform their corresponding SAA portfolio. The best-performing portfolio in terms of Sharpe ratio is the bounded PBR-GS* portfolio (SR 1.2788 in Table 9). The second best-performing portfolio is the unbounded PBR-LS portfolio with target return equal to the expected return of the 1/N portfolio (SR 1.2613 in Table 10). Therefore, we plotted the accumulated wealth of both portfolios in Figure 2, complemented with three benchmark portfolios. These are the unbounded GMV SAA portfolio, the MV SAA portfolio with target return of the 1/N portfolio and the 1/N portfolio. In the plot, we start with 1 dollar at the beginning of January 2004 and multiply this with monthly returns to end up with a final amount of wealth at the end of December 2013.

It can be observed that the GMV PBR-GS portfolio and the MV PBR-LS portfolio end up with much higher wealth than the benchmark portfolios. It is interesting to see that all portfolios perform almost similar until 2008 and then the PBR portfolios clearly have a lower dip during the financial crisis, which indicates that restricting the uncertainty in the sample variance seems to work in uncertain periods with high volatility and low stock returns. Then, all portfolios perform almost similarly, and in the last year, the PBR portfolios increase faster in wealth than the SAA portfolios. Therefore, we can conclude from the graph that the PBR portfolios perform better during uncertain periods and do perform equal or better during periods where the stock market is performing well in general.

Additionally, we also plotted the portfolios when transaction costs are incorporated. The graph can be seen in Figure 3. As opposed to the situation without transaction costs, now the GMV PBR-GS portfolio ends up with the highest wealth due to a lower turnover than the competing MV PBR-LS portfolio. Although the difference is still small, we can conclude that the turnover and corresponding transaction costs do have an impact on optimal portfolio choice.

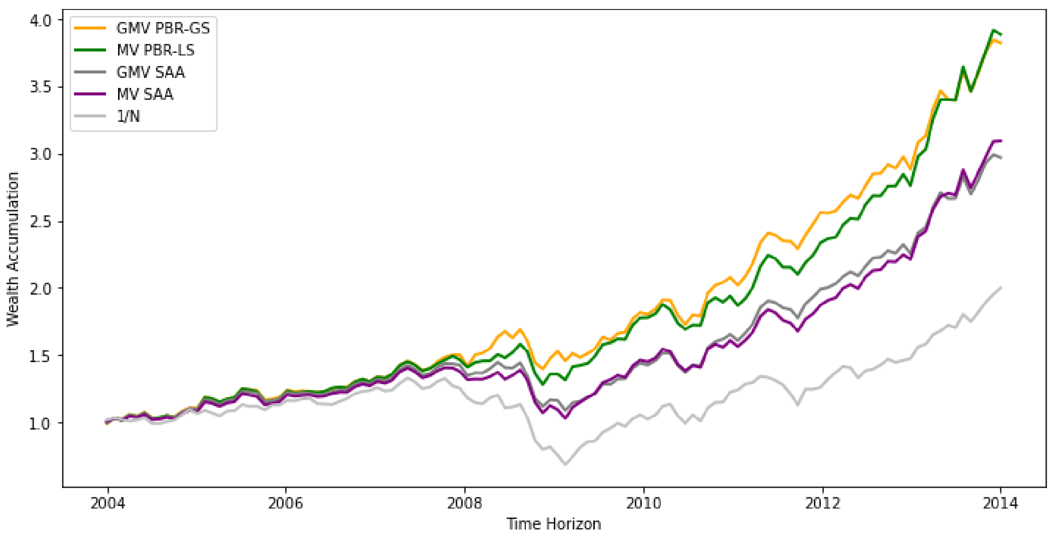

Figure 2: Visual presentation on wealth when 1 dollar is invested in January 2004 for different portfolios before transaction costs. The GMV PBR-GS portfolio is the PBR-GS$^*_{BO}$ portfolio in Table 9 (SR 1.2788). The MV PBR-LS portfolio has target return equal to the 1/N portfolio (SR 1.2613 in Table 10). The unbounded GMV SAA, MV SAA (with target return of 1/N portfolio) and 1/N portfolios are shown for comparison purposes.

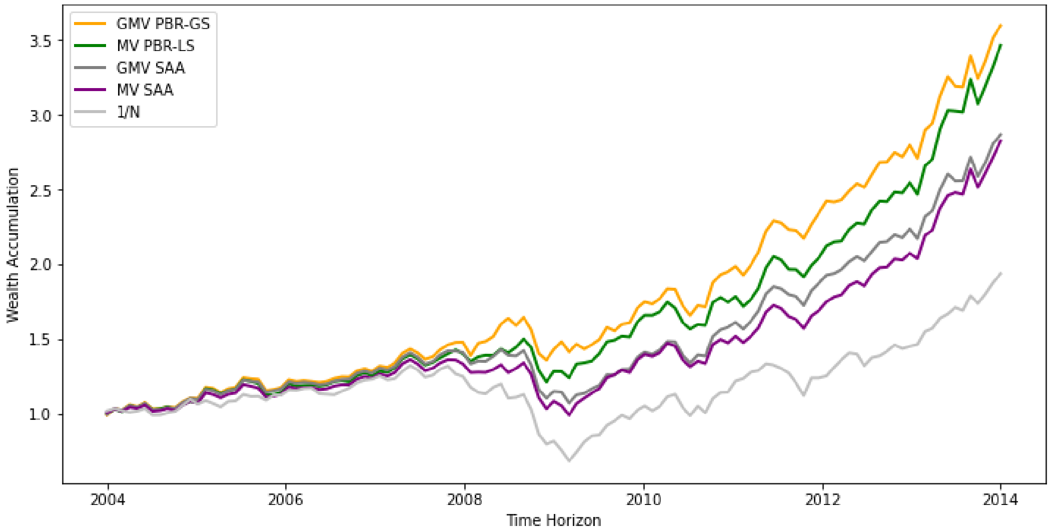

Figure 3: Visual presentation on wealth when 1 dollar is invested in January 2004 for different portfolios after transacion costs. The GMV PBR-GS portfolio is the PBR-GS$^{*}_{BO}$ portfolio in Table 9 (SR 1.2788). The MV PBR-LS portfolio has target return equal to the 1/N portfolio (SR 1.2613 in Table 10). The unbounded GMV SAA, MV SAA (with target return of 1/N portfolio) and 1/N portfolios are shown for comparison purposes.

6.7 Threshold Analysis

Since the implementation of restricting the sample variance of the sample variance is not a widely used concept yet in portfolio management, it is interesting to further look into the characteristics of the optimal threshold over time. In order to do this, we analyse the two best-performing PBR portfolios from the previous Section. In Figure 4, the vertical axis represents the optimal value for the threshold U. The line is effectively created from 120 monthly optimal thresholds from January 2004 until December 2013.

It can be observed that for the GMV PBR-GS portfolio, the thresholds are initially higher after which they make a steep decline and are much lower than for the MV PBR-LS portfolio. From 2010 onwards they take on similar values. The differentation in threshold until 2010 could be due to the difference in optimization methods. The backtracking line search algorithm does evaluate a more narrow range than the grid search whereas the backtracking line search algorithm does evaluate its range more thoroughly. This could lead to the initial differences. Additionally, the imposed target return for the mean-variance portfolio can also result in differences in optimal thresholds. Specifically for the period around 2008, this seems to be the case. There, the mean-variance portfolio still aims for a specified target return and thus does not allow for extremely low thresholds, whereas the GMV portfolio does not have this and directly creates very low thresholds in this period. In the wealth accumulation plot, it could also be observed that the GMV PBR-GS portfolio did have a smaller dip than the MV PBR-LS portfolio which might be due to the lower thresholds.

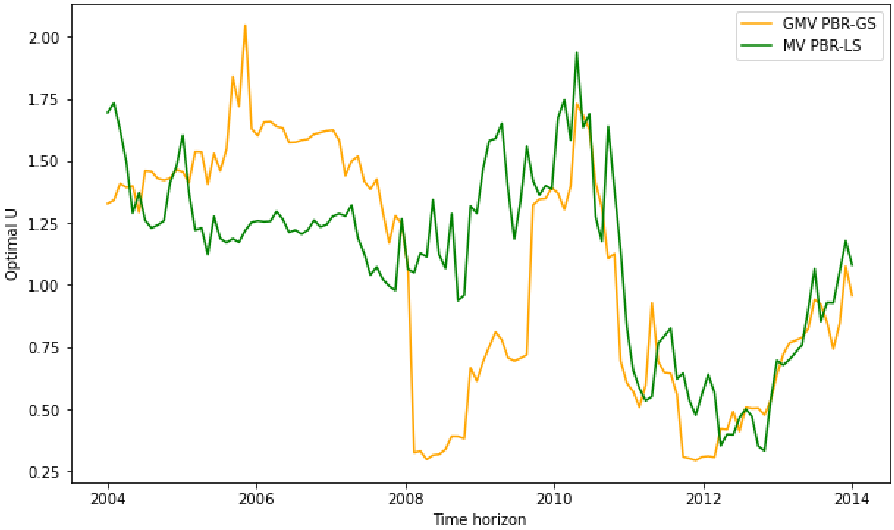

Figure 4: Optimal thresholds over time for the GMV PBR-GS$^*_{BO}$ portfolio (SR 1.2788 in Table 9) and the unbounded MV PBR-LS portfolio with target return of 1/N portfolio (SR 1.2613 in Table 10).

7 Model Combination Results

Based on the previous section, we choose the GMV PBR-GS$^*_{BO}$ with k=3 for the conservative portfolio because this is the PBR version of the GMV portfolio (thus double conservative) and attains the highest Sharpe ratio of all portfolios. For the more aggressive model, we take the unbounded MV PBR-LS model with k=2 and monthly target return 1.2% because it achieves a high average return compared to the other portfolios and also a much higher average return than the conservative portfolio. Below, we first present results for the recession prediction, after which we show results for the combined models that invest in these two portfolios.

7.1 Feature and Hyperparameter Evaluation

For all included macroeconomic variables, correlation is computed with the US recession variable for 24 lags of every macroeconomic variable. Additionally, autocorrelation is computed for the US recession with 24 lags. The variables are not transformed, except for some variables that are transformed to monthly data instead of weekly or quarterly and the GDP variable is transformed to growth rates instead of quarterly state. Table 26 in the Appendix contains the (auto)correlations and based on these, initial input variables for the random forest model are selected. It can be observed that for most variables, the correlation declines with the number of lags, especially after three or four lags. Therefore, for most variables three or four lagged values are included in the model.

Variables for which other lagged values are included are unemployment (lag 1, 2 and 24 are included since lag 24 also has a relatively high negative correlation), personal consumption expenditures (lags 22, 23 and 24 are included since absolute correlation increases with number of lags) and macroeconomic uncertainty estimate 12 months ahead (lag 12) since this estimate provides information 12 months ahead. For yield spread, we only include three lags despite the correlation increasing with the number of lags. This is because our data for this variable does not go further back. A detailed representation of which lags are included for every feature can be observed in the last row of Table 26. These are all included variables in the base model before feature selection is performed to make the model more interpretable and less vulnerable to overfitting.

For the model with all these features, optimized hyperparameters are: max depth (5), number of features per split (0.5 of the total), minimum samples per leaf (4), minimum samples per split (10), and optimal number of trees (100). After obtaining these optimal hyperparameters for the model with all features, we evaluated the impact of each feature and an overview of this can be seen in Figure 5, where all features with a SHAP value smaller than 0.001 are hidden because of their insignificance. We observe that the US recession with lag 1 attains by far the highest SHAP value, followed by the composite leading indicator. After that, there is a gap in terms of the magnitude of the SHAP value. Business tendency with one lag, mortgage with three lags, US recession with two lags, composite leading indicator with two lags and inflation with one lag all take on SHAP values around 0.01 and after that the features take on even lower values. Therefore, we include all features until inflation with lag one in our model. We do this because adding more features does not add much in terms of predictive power and can potentially cause overfitting and lead to more model complexity and less interpretability.

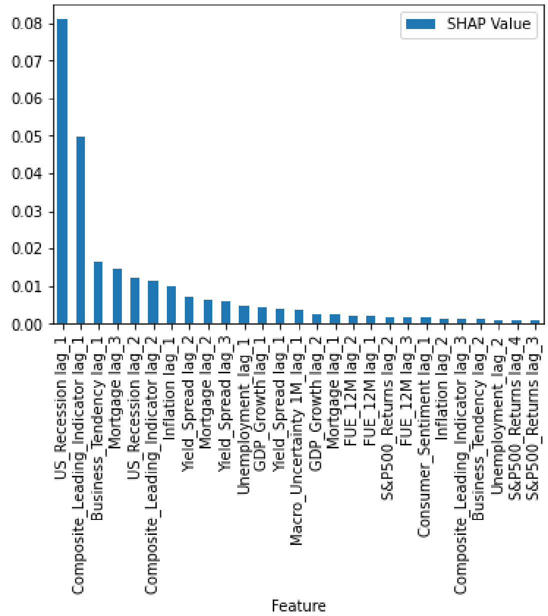

Figure 5: Average absolute SHAP values of the evaluated features obtained with in sample k-fold cross-validation with k=3. Features with SHAP value smaller than 0.001 are not shown.

Now that the features are selected, hyperparameters are again optimised, and the evaluated grid and final values are shown in Table 13. We slightly adjusted the grid that was used in the hyperparameter optimisation before feature selection by adding smaller or larger values if the smallest or largest value in the grid was selected. This time for the number of trees, maximum depth and minimum samples in a leaf the minimum value of the grid is selected. However, we choose not to evaluate smaller values for the number of trees because 50 trees is already a relatively low number. Also, the maximum depth is already relatively low. The minimum samples in a leaf cannot be reduced further with value 1.

Hyperparameter	Grid	Optimal value
Number of trees	50, 100, 200, 300, 400, 500	50
Maximum depth	4, 5, 10, 15, None	2
Minimum number of samples for split	2, 5, 10, 15, 20, 25, 30, 35	20
Minimum number of samples in a leaf	1, 2, 4, 6, 8	1
Maximum number of features evaluated for a split	'auto', 'sqrt', 'log2', 0.3, 0.5	'sqrt'

Table 13: Optimal hyperparameters for the random forest model after feature selection.

With the model with the seven selected features and optimal hyperparameters, predictions on whether the US economy is in a recession the next month are obtained. In Figure 6, a SHAP plot is shown with the relation of each feature with the dependent variable, the US recession. Every dot represents an out-of-sample forecast and the impact of the feature on it. A red dot represents a high value of the feature, and a blue dot a low value. Hence, we observe that a previous recession is strongly related to a recession in the upcoming month, whereas a high value for the composite leading indicator (representing short-term economic activity relative to the long-term level) would probably mean that the economy will not be in a recession the next month. For the mortgage rate, it is difficult to see whether high or low values do make a big difference. For the business tendency low values would again imply a recession and for inflation the impact on predictions is relatively small regardless of whether inflation is high or low.

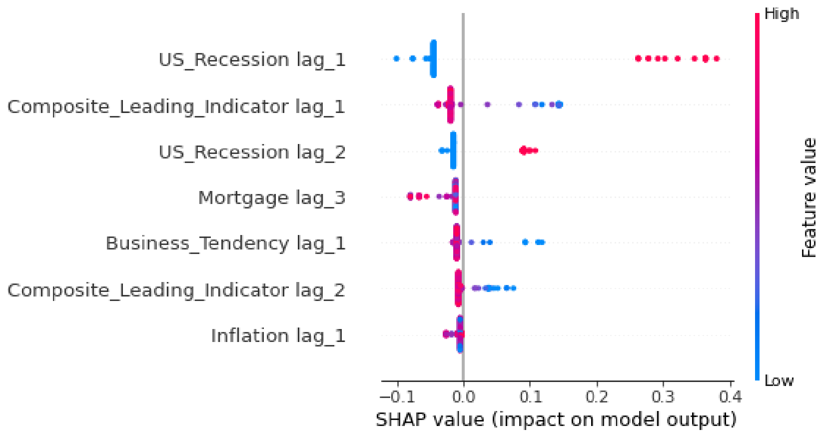

Figure 6: SHAP values for the selected features. Every dot represents an out-of-sample prediction and the corresponding impact of the feature.

7.2 Performance Evaluation

In Table 14, the confusion matrix is shown for the predictions made with the random forest model. Especially important for our purpose is the recall of the recession, which takes on value 94.7%. This is important because if we predict the economy to be in an expansion and invest with our less conservative model and the economy turns out to be in a recession, we are vulnerable to high losses. With a recall of 94.7% this only happens one time out of 19 recession months and losses are therefore minimized well. At the same time, it is not the case that our random forest model predicts a recession too often because precision there is also 94.7% since there is only one time we predict a recession when there turns out to be an expansion. Additionally, the recall and precision for an expansion are also very high with 99%. Overall accuracy takes on value 98.3%. Hence, taking into account all performance measures, we can conclude that our model performs well in forecasting whether there will be a recession or expansion next month.

	Actual Expansion	Actual Recession	Precision
Predicted Expansion	100	1	99.0%
Predicted Recession	1	18	94.7%
Recall	99.0%	94.7%	

Table 14: Confusion matrix of the random forest model for predicting whether the US economy will be in a recession next month or not. Overall accuracy of the model is 98.3%.

The main gain of the random forest model is that it provides us with a percentage probability of a recession for every month, thereby allowing for tailored weighted portfolio combinations based on these percentages. In contrast, using the binary US recession variable would not allow us to construct portfolios with specific weighting schemes related to the magnitude of the probability on a recession. In Table 15, performance statistics for our five combined models, the two separate portfolios and the benchmark mean-variance SAA portfolio with a target return of 1.2% are shown. It can be observed that without transaction costs the binary combined model already has a much higher Sharpe ratio than the conservative portfolio, which is our best-performing single portfolio (SR 1.3557 vs. 1.2788). Next to that, we see that finding optimal combination weights for specified probability intervals in-sample and using these out of sample, further improves Sharpe ratio. Moreover, we see that the recession avoidance strategy obtains the highest Sharpe ratio of all (excluding transaction costs). This makes sense since during a recession both separate portfolios do obtain negative returns for many months, so not investing at all would theoretically be a successful strategy.

It is noteworthy that the first four combined portfolios all significantly outperform the MV SAA portfolio in terms of Sharpe ratio for both the case with and without transaction costs. Simultaneously, they attain higher Sharpe ratios than the separate portfolios that they are constructed from when transaction costs are not included. Two combined portfolios (with optimized and recession avoidance weights) also take on higher Sharpe ratios than both individual counterparts when transaction costs are included. In the case of included transaction costs, the combined model with optimized weights (CM WS_O) performs best of all portfolios in this research. Therefore, we can conclude that the concept of taking into account the economic state when investing in a combination of portfolios proves to be beneficial for portfolio performance. Studentized bootstrapped p-values for the Sharpe ratios can be found in Table 25 in the Appendix.

	Excluding Transaction Costs				Including Transaction Costs		
Portfolio	SR	Mean	Vol.	TO	SR	Mean	Vol.
MV SAA	1.0046	14.9251	14.8574	0.3472	0.8738	13.0974	14.9887
MV PBR - LS	1.1241 (0.0898)	17.4244	15.5010	0.3854	0.9934 (0.0762)	15.4874	15.5906
GMV PBR - GS^*_{BO}	1.2788 (0.1532)	14.0969	11.0238	0.1327	1.2263 (0.0585)	13.5955	11.0862
CM WS_B	1.3557 (0.0310)	18.1815	13.4114	0.3773	1.2155 (0.0313)	16.2844	13.3968
CM WS_O	1.3956 (0.0200)	16.1830	11.5959	0.2248	**1.3010**(0.0075)	15.1603	11.6529
CM WS_I	1.2935 (0.0116)	17.4616	13.4998	0.3401	1.1673 (0.0085)	15.7775	13.5158
CM WS_{RA}	**1.4146**(0.0246)	16.9758	12.0002	0.3729	1.2589 (0.0281)	15.0945	11.9902
CM WS_S	1.0623 (0.8971)	16.5272	15.5585	1.5005	0.4728 (0.3612)	7.7905	16.4757

Table 15: Annualized Sharpe ratio, mean return, volatility and turnover for the MV SAA (monthly TR 1.2%), two separate portfolios and combined models (CM) with weighting schemes binary (WS_B), optimized (WS_O), incremental (WS_I), recession avoidance (WS_{RA}) and short sell (WS_S). The MV PBR-LS portfolio has target return 1.2%. HAC p-values for the differences in Sharpe ratio between the MV SAA (TR 1.2%) portfolio and corresponding portfolio are shown in brackets.

In Figure 7, a wealth accumulation plot including transactions costs can be observed for the combined portfolio with binary weighting scheme, the two indvidiual portfolios and the 1/N portfolio. Here the benefit of the combined portfolio can be observed very well. Before the financial crisis of 2008-2009 the combined portfolio is fully invested in the more risky MV PBR-LS portfolio and benefits from this by increasing wealth faster than the more conservative GMV PBR-GS portfolio. Then, during the financial crisis, the combined model takes on a position in the conservative model and has less wealth decrease due to this. Then, after the crisis, it invests again in the riskier model and obtains high returns as a result. At the end of the time period, it even attains higher wealth than the aggressive MV PRB-LS model and much higher wealth than the GMV PBR-GS model. In Figure 10 in the Appendix, the same graph is shown without transaction costs.

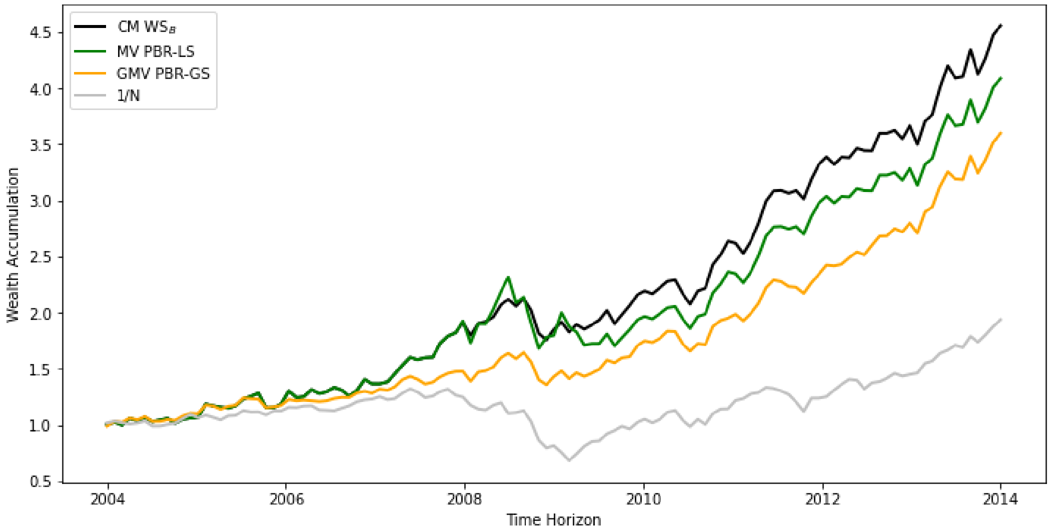

Figure 7: Wealth accumulation plot including transaction costs for the combined model with binary weighting scheme (CM WS$_B$), MV PBR-LS (monhtly TR = 1.2%), GMV PBR-GS$^*_{BO}$ ($k = 3$) and the 1/N portfolio.

In Figure 8, wealth accumulation plots including transaction costs for the five combined portfolios are shown. It can be observed that the binary weighting scheme attains the highest wealth at the end of the period, although the optimized weighting scheme provides a higher Sharpe ratio (see Table 15). This is due to the fact that the higher Sharpe ratio of the optimized weighting scheme portfolio is mainly a result of its very low volatility. Therefore, it depends on the preferences of an investor which combined portfolio he chooses. A more risk-averse investor would lean towards the optimized weights combined portfolio, whereas an investor seeking higher returns would invest in the binary weights combined portfolio. Lastly, we observe that the idea of short selling the more risky portfolio during recession periods does not deliver promising results due to both moderate performance to begin with and the addition of high transaction costs. In Figure 11 in the Appendix, the same graph is shown without transaction costs.

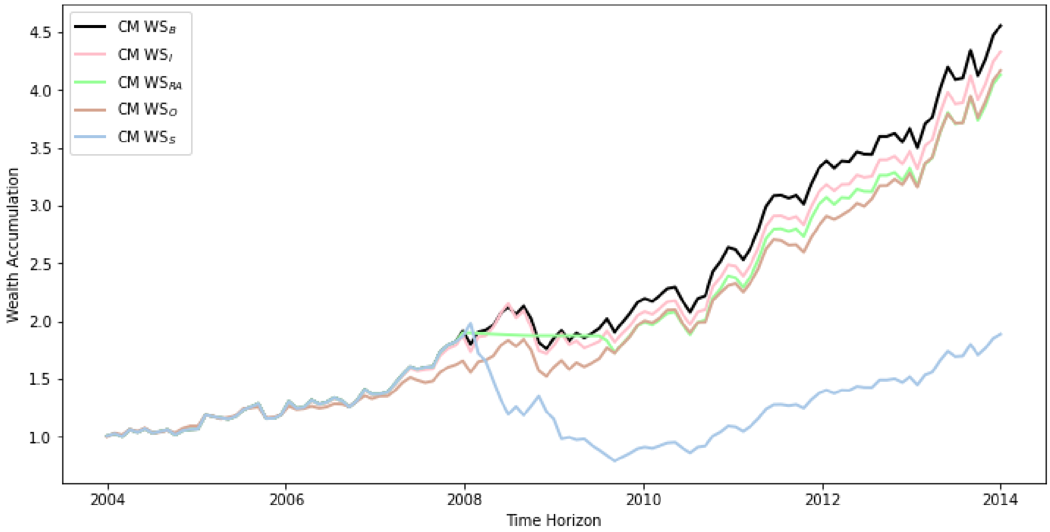

Figure 8: Wealth accumulation plot including transaction costs for the five combined models (CM) with binary (B), optimized (O), incremental (I), recession avoidance (RA) and short sell (S) weighting schemes.

8 Conclusion

This research consists of two major parts where we analyse portfolio performance on monthly returns from Fama French datasets. The first part investigates whether restricting the uncertainty in the sample variance in a mean-variance framework improves portfolio performance. The restricting threshold on this uncertainty is optimized for every month through PBR, which implies using a k-fold cross-validation optimization procedure with regard to the Sharpe ratio. Several additions to this PBR portfolio are evaluated, namely bounded portfolio weights, releasing an optimization restriction, different optimization methods for the uncertainty threshold and setting the return of the 1/N portfolio as a moving target return for the mean-variance portfolio. Moreover, the addition of transaction costs on portfolio performance is analysed. We find that the GMV PBR-GS$^*_{BO}$ portfolio with three bins in the cross-validation attains the highest Sharpe ratio, closely followed by the MV PBR-LS model with 1/N moving target return and three bins in the cross-validation. Both portfolios significantly outperform their SAA counterparts and the 1/N portfolio. After including transaction costs, the MV PBR-LS model still outperforms its MV SAA counterpart.

The second part of this research focuses on forecasting economic recessions and related investment strategies. This is especially interesting in combination with the first part since the PBR portfolios perform particularly well during times of economic downturn. In order to forecast the probability on a US recession for the next month, we make use of a random forest model with mostly macroeconomic variables as input. We optimize hyperparameters for the random forest model and evaluate SHAP values to analyse feature importance. Then, based on the probability output of a recession from the random forest model, we invest in a conservative and/or a more risky portfolio from the first part. We evaluate five different combination weighting schemes for recession probabilities. In the random forest model, we find that a couple of evaluated features have the majority of predictive power, from which the recession and the composite leading indicator in the previous month are the most important. Furthermore, our model performs well in forecasting a recession with high recall and accuracy scores. Considering the combined portfolios, most of the portfolios attain higher Sharpe ratios than their individual counterparts. The combined portfolio with an optimized weighting scheme through in-sample time series cross-validation turns out to deliver the highest Sharpe ratio of all evaluated portfolios, including transaction costs. Therefore, it can be concluded that the PBR method improves portfolio performance and combining PBR portfolios based on the forecasted state of the economy further enhances performance.

For further research, it would be interesting to investigate the PBR method on daily returns, since then more data can be taken into account. The HAC p-values for testing differences in Sharpe ratio are then more reliable, which might be a limitation of this research. Moreover, for forecasting the economic state we now took into account a binary variable (recession or not). We strived to also evaluate a variable that represented more states of the economy to be more flexible, but we were not able to find data on this. However, using a proxy like the yield curve might be interesting to further look into in the future. Lastly, for the combined portfolio a suggestion would be to evaluate performance when an even more aggressive portfolio is taken into account for the risky model.

References

Gah-Yi Ban, Noureddine El Karoui, and Andrew EB Lim. Machine learning and portfolio optimization. *Management Science*, 64(3):1136–1154, 2018.

James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(2), 2012.

Stephen P Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

Michael W Brandt. Portfolio choice problems. In *Handbook of financial econometrics: Tools and techniques*, pages 269–336. Elsevier, 2010.

Leo Breiman. Random forests. *Machine learning*, 45:5–32, 2001.

Roger G Clarke and Harindra de Silva. State-dependent asset allocation. *Journal of Portfolio Management*, 24(2):57, 1998.

A Craig MacKinlay and L'uboš Pástor. Asset pricing models: Implications for expected returns and portfolio selection. *The Review of financial studies*, 13(4):883–916, 2000.

Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The review of Financial studies*, 22(5):1915–1953, 2009.

Victor DeMiguel, Alberto Martin-Utrera, and Francisco J Nogales. Size matters: Optimal calibration of shrinkage estimators for portfolio selection. *Journal of Banking & Finance*, 37(8):3018–3034, 2013.

Graham Elliott and Allan Timmermann. Optimal forecast combination under regime switching. *International Economic Review*, 46(4):1081–1102, 2005.

Periklis Gogas, Theophilos Papadimitriou, Maria Matthaiou, and Efthymia Chrysanthidou. Yield curve and recession forecasting in a machine learning framework. *Computational Economics*, 45:635–645, 2015.

Shingo Goto and Yan Xu. Improving mean variance optimization through sparse hedging restrictions. *Journal of Financial and Quantitative Analysis*, 50(6):1415–1441, 2015.

Sami Ben Jabeur, Salma Mefteh-Wali, and Jean-Laurent Viviani. Forecasting gold price with the xgboost algorithm and shap interaction values. *Annals of Operations Research*, pages 1–21, 2021.

Philippe Jorion. Bayes-stein estimation for portfolio analysis. *Journal of Financial and Quantitative analysis*, 21(3):279–292, 1986.

Raymond Kan and Guofu Zhou. Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, 42(3):621–656, 2007.

Chris Kirby and Barbara Ostdiek. It's all in the timing: simple active portfolio strategies that outperform naive diversification. *Journal of financial and quantitative analysis*, 47(2):437–467, 2012.

Oliver Ledoit and Michael Wolf. Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance*, 15(5):850–859, 2008.

Olivier Ledoit and Michael Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance*, 10(5):603–621, 2003.

Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2):365–411, 2004.

Olivier Ledoit and Michael Wolf. The power of (non-) linear shrinking: A review and guide to covariance matrix estimation. *Journal of Financial Econometrics*, 20(1):187–218, 2022.

Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

Scott M Lundberg, Gabriel G Erion, and Su-In Lee. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*, 2018.

Kathryn L Lunetta, L Brooke Hayward, Jonathan Segal, and Paul Van Eerdewegh. Screening large-scale association study data: exploiting interactions using random forests. *BMC genetics*, 5:1–13, 2004.

Henri Nyberg. Dynamic probit models and financial variables in recession forecasting. *Journal of Forecasting*, 29(1-2):215–230, 2010.

Philipp Probst, Marvin N Wright, and Anne-Laure Boulesteix. Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: data mining and knowledge discovery*, 9(3):e1301, 2019.

Patrícia Ramos and José Manuel Oliveira. A procedure for identification of appropriate state space and arima models based on time-series cross-validation. *Algorithms*, 9(4):76, 2016.

Lloyd S Shapley et al. A value for n-person games. 1953.

Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, 25, 2012.

Charles Stein. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, volume 3, pages 197–207. University of California Press, 1956.

Jun Tu and Guofu Zhou. Markowitz meets talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, 99(1):204–215, 2011.

9 Appendix

9.1 Data Tables

FF 5 industry					
	Consumer	Manufacturing	Hi-Tech	Health	Other
Mean	9.0260	10.8270	7.9420	7.8550	4.3180
SD	12.8992	16.2952	17.3585	13.1642	19.6724
SR	0.6997	0.6644	0.4575	0.5967	0.2195

Table 16: Annualized descriptive statistics for excess returns on the 5 industry portfolios for the period ranging from January 2004 until December 2013.

FF 10 industry										
	Cnsmr Nond.	Cnsmr Dur.	Manuf.	Energy	Hi-Tech	Telecom	Shops	Health	Utilities	Other
Mean	9.436	7.069	10.596	13.410	7.789	9.148	9.505	7.806	9.441	4.299
SD	11.6190	29.6112	18.3668	20.8271	18.7207	15.4913	14.2531	13.1857	12.7804	19.4636
SR	0.8121	0.2387	0.5769	0.6439	0.4161	0.5905	0.6669	0.5920	0.7387	0.2209

Table 17: Annualized descriptive statistics for excess returns on the 10 industry portfolios for the period ranging from January 2004 until December 2013.

FF 49 industry					
	Min.	Q_1.	Median	Q_3	Max.
Mean	-2.202	13.354	12.587	13.395	16.016
SD	34.3804	43.9289	25.1281	22.0474	17.2434
SR	-0.064	0.304	0.5009	0.6076	0.9288

Table 18: Annualized descriptive statistics for excess returns on the 49 industry portfolios for the period ranging from January 2004 until December 2013.

9.2 PBR Results

	FF 5 industry		FF 10 industry		FF 49 industry	
MV, TR = 0.8%						
	2 bins	3 bins	2 bins	3 bins	2 bins	3 bins
PBR - LS	**1.2141** (0.0644)	1.1879 (0.0790)	1.1070 (0.6016)	**1.1084** (0.5914)	0.3353 (0.7800)	0.3341 (0.7256)
PBR - GS	1.1558 (0.1080)	1.1477 (0.0964)	1.0974 (0.4336)	1.1015 (0.4776)	0.3476 (0.9708)	0.3408 (0.7828)
L1	0.9023 (0.5888)	0.8180 (0.3450)	0.9735 (0.4216)	0.9726 (0.3620)	0.4598 (0.5112)	0.4552 (0.5472)
L2	0.9283 (0.7402)	0.9596 (0.7256)	0.9817 (0.3132)	1.0261 (0.7222)	0.3965 (0.5866)	0.3765 (0.7412)
SAA	1.0114		1.0562		0.3493	
NS	0.6255 (0.1494)		0.8332 (0.4180)		**0.5698** (0.5080)	
JS	1.0586 (0.7770)		0.7892 (0.5092)		-0.2435 (0.0528)	
LW	0.9997 (0.2512)		1.0311 (0.2736)		0.4163 (0.5450)	
MV, TR = 1.0%						
	2 bins	3 bins	2 bins	3 bins	2 bins	3 bins
PBR - LS	1.2186 (0.1446)	**1.2306** (0.0942)	1.0502 (0.6516)	**1.0575** (0.5964)	0.2950 (0.8532)	0.2897 (0.7520)
PBR - GS	1.1585 (0.2244)	1.2110 (0.0972)	1.0487 (0.4812)	1.0570 (0.4136)	0.3070 (0.9760)	0.2945 (0.7082)
L1	0.8708 (0.4494)	0.9088 (0.4948)	0.9254 (0.2878)	0.9170 (0.4660)	0.3922 (0.6162)	0.3925 (0.6092)
L2	0.8776 (0.3608)	0.8469 (0.3894)	0.93603 (0.2754)	0.9737 (0.7186)	0.3242 (0.7892)	0.3298 (0.6476)
SAA	1.0731		1.0098		0.3056	
NS	0.5597 (0.0494)		0.5794 (0.1660)		**0.5477** (0.4766)	
JS	0.9269 (0.4880)		0.6517 (0.3654)		-0.2650 (0.0578)	
LW	1.0628 (0.3516)		0.9950 (0.5808)		0.3692 (0.5574)	
MV, TR = 1.2%						
	2 bins	3 bins	2 bins	3 bins	2 bins	3 bins
PBR - LS	**1.1241** (0.1418)	1.1202 (0.1192)	0.9586 (0.6744)	0.9753 (0.4954)	0.2543 (0.8660)	0.2444 (0.6399)
PBR - GS	1.0827 (1.1434)	1.1037 (0.1188)	0.9756 (0.3736)	**0.9787** (0.3596)	0.2509 (0.8238)	0.2474 (0.6430)
L1	0.8357 (0.5594)	0.7799 (0.2478)	0.8263 (0.2850)	0.8535 (0.4924)	0.3013 (0.7964)	0.3110 (0.7360)
L2	0.7745 (0.2044)	0.8371 (0.3322)	0.8572 (0.3258)	0.9109 (0.8456)	0.2717 (0.8180)	0.2705 (0.8736)
SAA	1.0046		0.9270		0.2612	
NS	0.5114 (0.0888)		0.5125 (0.2276)		**0.5118**(0.4924)	
JS	0.8134 (0.3024)		0.5568 (0.3182)		-0.2792 (0.0670)	
LW	0.9956 (0.3780)		0.9209 (0.7976)		0.3219 (0.5608)	
GMV						
	2 bins	3 bins	2 bins	3 bins	2 bins	3 bins
PBR - LS	1.1944 (0.1262)	**1.2178** (0.0838)	1.0512 (0.6844)	1.0754 (0.4744)	0.3757 (0.7606)	0.3761 (0.7222)
PBR - GS	1.1014 (0.3736)	1.1708 (0.1062)	1.0321 (0.6196)	**1.0931** (0.1704)	0.3552 (0.5062)	0.3799 (0.7552)
L1	0.9546 (0.2610)	0.9709 (0.3026)	0.9346 (0.5412)	0.9704 (0.6604)	0.5561 (0.5992)	0.4613 (0.7958)
L2	1.0063 (0.6282)	0.9937 (0.6018)	0.9697 (0.7684)	0.9298 (0.4952)	0.6122 (0.1332)	0.5972 (0.1524)
SAA	1.0296		1.0000		0.3897	
NS	0.8216 (0.0900)		0.8753 (0.4608)		**0.6284** (0.4120)	
LW	1.0186 (0.2712)		0.9840 (0.4510)		0.4752 (0.4264)	
EW	0.5596		0.6033		0.5595	

Table 19: Same table as Table 6, but with studentized bootstrap p-values instead of HAC p-values between brackets. Explanation: Annualized out-of-sample Sharpe ratios for the portfolios discussed in section 4.1 for three datasets. Abbreviations can be found in Table 27.

FF 5 industry		
MV, TR = 1.0%	Sharpe Ratio	
	2 bins	3 bins
PBR - LS	1.1649 (0.6890)	1.2386 (0.2230)
PBR - GS	1.1929 (0.2258)	**1.2464**(0.1842)
L1	0.8346 (0.3608)	0.8065 (0.3160)
L2	0.8687 (0.2408)	0.8191 (0.1766)
SAA	1.1136	
NS	0.5597 (0.5604)	
JS	0.3752 (0.0026)	
LW	1.1054 (0.2196)	
GMV		
	2 bins	3 bins
PBR - LS	1.2168 (0.0984)	1.2411 (0.0768)
PBR - GS	1.1470 (0.1828)	**1.2538** (0.1076)
L1	0.9507 (0.2110)	0.9676 (0.1316)
L2	1.0063 (0.4556)	0.9937 (0.4800)
SAA	1.0410	
NS	0.8216 (0.0660)	
LW	0.9841 (0.6554)	

Table 20: Same table as table 8, but with studentized bootstrap p-values instead of HAC p-values.

Explanation: Sharpe ratios for the portfolios from section 4.1 with individual asset weights now bounded by (-1,1).

FF 5 industry		
GMV	Sharpe Ratio	
	2 bins	3 bins
PBR - LS_{BO}	1.2168 (0.0984)	1.2411 (0.0768)
PBR - LS^*_{BO}	1.2172 (0.0954)	1.2431 (0.0692)
PBR - GS_{BO}	1.1470 (0.1828)	1.2538 (0.1076)
PBR - GS^*_{BO}	1.1487 (0.2642)	**1.2788**(0.0976)
PBR - RS_{BO}	1.1698 (0.1758)	1.2419 (0.0884)
PBR - RS^*_{BO}	1.1517 (0.1774)	1.2440 (0.0826)
SAA	1.0410	

Table 21: Same table as Table 9, but with studentized bootstrap p-values instead of HAC p-values.

Explanation: Sharpe ratios for the bounded GMV PBR portfolios based on backtracking line search (LS), grid search and random search (RS) optimization. $*$ marks that the $w'\alpha > 0$ restriction is released for the corresponding portfolio and BO represents bounded weights.

FF 5 industry		
MV	Sharpe Ratio	
	2 bins	3 bins
PBR - LS	1.2472 (0.0436)	**1.2613**(0.0480)
PBR - LS_{BO}	1.2094 (0.1096)	1.2156 (0.1340)
PBR - GS	1.1238 (0.1982)	1.1779 (0.0614)
PBR - GS_{BO}	1.1071 (0.1574)	1.2111 (0.0564)
SAA	1.0326	

Table 22: Same as Table 10, but with studentized bootstrap p-values instead of HAC p-values.
Explanation: Sharpe ratios for MV PBR portfolios with time-varying target return equal to the expected return on the 1/N portfolio. The SAA portfolio is shown as well for comparison purposes.

FF 5 industry						
Portfolio	SR excl. TC		TO		SR incl. TC	
	2 bins	3 bins	2 bins	3 bins	2 bins	3 bins
GMV PBR - GS^*_{BO}	1.1488 (0.0128)	**1.2788**(0.0090)	0.1576	0.1327	1.0850 (0.0264)	**1.2263**(0.0150)
MV PBR - LS_{EW}	1.2472 (0.0036)	1.2613 (0.0042)	0.2220	0.2237	1.1565 (0.0126)	1.1668 (0.0128)
1/N	0.5596		0.0239		0.5396	
SAA GMV_{BO}	1.0410 (0.0098)		0.0861		1.0115 (0.0180)	
SAA MV_{EW}	1.0326 (0.0024)		0.1648		0.9584 (0.0106)	

Table 23: Same as Table 11, but with studentized bootstrap p-values instead of HAC p-values.
Explanation: Sharpe ratios (SR) before and after transaction costs (TC) are shown. Also turnover (TO) is presented. The EW subscript represents time-varying target return equal to the 1/N portfolio. p-values are shown versus the 1/N portfolio.

FF 5 industry						
Portfolio	SR excl. TC		TO		SR incl. TC	
	2 bins	3 bins	2 bins	3 bins	2 bins	3 bins
GMV PBR - GS^*_{BO}	1.1488 (0.2642)	**1.2788**(0.0976)	0.1576	0.1327	1.0850 (0.4544)	**1.2263**(0.1266)
MV PBR - LS_{EW}	1.2472 (0.0436)	1.2613 (0.0480)	0.2220	0.2237	1.1565 (0.0512)	1.1668 (0.0708)
1/N	0.5596		0.0239		0.5396	
SAA GMV_{BO}	1.0410		0.0861		1.0115	
SAA MV_{EW}	1.0326		0.1648		0.9584	

Table 24: Same as Table 12, but with studentized bootstrap p-values instead of HAC p-values.
Explanation: Sharpe ratios (SR) before and after transaction costs (TC) are shown. Also turnover (TO) is presented. The EW subscript represents time-varying target return equal to the 1/N portfolio. p-values are shown versus the corresponding SAA portfolio

.

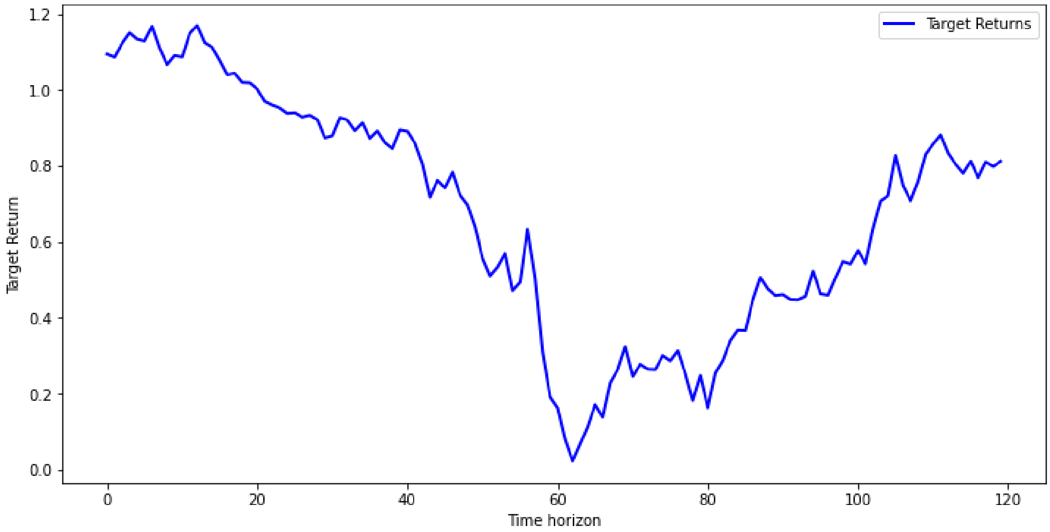

Figure 9: Target returns for the mean-variance portfolio with target return equal to the expected return of the 1/N portfolio over time. The forecast period ranges from January 2004 until December 2013, so 120 months.

9.3 Model Combination Results

	Excluding Transaction Costs				Including Transaction Costs		
Portfolio	SR	Mean	Vol.	TO	SR	Mean	Vol.
MV SAA	1.0046	14.9251	14.8574	0.3472	0.8738	13.0974	14.9887
MV PBR - LS	1.1241 (0.1418)	17.4244	15.5010	0.3854	0.9934 (0.1232)	15.4874	15.5906
GMV PBR - GS^*_{BO}	1.2788 (0.1768)	14.0969	11.0238	0.1327	1.2263 (0.0900)	13.5955	11.0862
CM WS_B	1.3557 (0.0878)	18.1815	13.4114	0.3773	1.2155 (0.0998)	16.2844	13.3968
CM WS_O	1.3956 (0.0558)	16.1830	11.5959	0.2248	**1.3010**(0.0286)	15.1603	11.6529
CM WS_I	1.2935 (0.0532)	17.4616	13.4998	0.3401	1.1673 (0.0638)	15.7775	13.5158
CM WS_{RA}	**1.4146**(0.0597)	16.9758	12.0002	0.3729	1.2589 (0.0708)	15.0945	11.9902
CM WS_S	1.0623 (0.8858)	16.5272	15.5585	1.5005	0.4728 (0.4466)	7.7905	16.4757

Table 25: Same as Table 15, but with studentized bootstrap p-values instead of HAC p-values.
Explanation: Annualized Sharpe ratio, mean return, volatility and turnover for the MV SAA (monthly TR 1.2%), two separate portfolios and combined models (CM) with weighting schemes binary (WS_B), optimized (WS_O), incremental (WS_I), recession avoidance (WS_{RA}) and short sell (WS_S). The MV PBR-LS portfolio has target return 1.2%. p-values for the differences in Sharpe ratio between the MV SAA (TR 1.2%) portfolio and corresponding portfolio are shown in brackets.

Lag	US Recession Autocorr.	Inflation	Unemp.	Yield Spread	GDP Growth	EU Recession	Industrial Prod.	Personal Cons. Exp.	Cons. Sent.	Macro Unc. 1M	Macro Unc. 12M	Financial Unc.	Business Tendency	Comp. Leading Ind.	S&P 500 Returns	Mortg. Rate	S&P 500 Volatility
1	0.890	0.396	0.166	-0.221	-0.530	0.234	-0.108	-0.099	-0.475	0.591	0.558	0.287	-0.643	-0.626	-0.085	0.393	0.134
2	0.781	0.391	0.121	-0.290	-0.510	0.197	-0.099	-0.101	-0.445	0.569	0.552	0.282	-0.588	-0.584	-0.095	0.408	0.114
3	0.671	0.387	0.079	-0.349	-0.466	0.159	-0.092	-0.101	-0.403	0.540	0.541	0.272	-0.529	-0.529	-0.121	0.410	0.089
4	0.562	0.382	0.044	-0.383	-0.371	0.121	-0.085	-0.102	-0.361	0.512	0.526	0.257	-0.475	-0.466	-0.152	0.397	0.076
5	0.452	0.378	0.016	-0.397	-0.304	0.083	-0.080	-0.103	-0.327	0.491	0.512	0.239	-0.428	-0.400	-0.120	0.380	0.063
6	0.343	0.373	-0.008	-0.418	-0.263	0.045	-0.075	-0.104	-0.295	0.478	0.502	0.223	-0.383	-0.334	-0.143	0.368	0.050
7	0.233	0.370	-0.031	-0.450	-0.254	0.028	-0.071	-0.106	-0.274	0.466	0.492	0.206	-0.332	-0.269	-0.121	0.357	0.036
8	0.151	0.369	-0.049	-0.471	-0.244	0.011	-0.068	-0.108	-0.235	0.458	0.482	0.189	-0.277	-0.207	-0.109	0.343	0.008
9	0.069	0.372	-0.063	-0.479	-0.179	-0.007	-0.065	-0.110	-0.208	0.454	0.473	0.175	-0.224	-0.151	-0.098	0.329	0.007
10	0.041	0.374	-0.074	-0.482	-0.134	-0.005	-0.063	-0.112	-0.179	0.444	0.465	0.163	-0.178	-0.101	-0.041	0.315	0.009
11	0.013	0.379	-0.084	-0.486	-0.095	-0.004	-0.061	-0.115	-0.168	0.436	0.459	0.156	-0.144	-0.060	-0.071	0.298	0.014
12	0.013	0.383	-0.091	-0.481	-0.089	-0.002	-0.061	-0.117	-0.159	0.429	0.453	0.157	-0.120	-0.027	-0.013	0.280	0.021
13	0.013	0.392	-0.092	-0.482	-0.101	-0.001	-0.061	-0.120	-0.155	0.423	0.449	0.161	-0.105	-0.002	-0.011	0.263	0.031
14	0.012	0.398	-0.093	-0.471	-0.105	0.001	-0.062	-0.122	-0.157	0.419	0.446	0.168	-0.092	0.017	-0.016	0.244	0.045
15	0.012	0.399	-0.098	-0.466	-0.116	0.002	-0.061	-0.125	-0.167	0.409	0.442	0.179	-0.075	0.032	0.040	0.237	0.042
16	0.011	0.400	-0.107	-0.468	-0.089	0.004	-0.061	-0.128	-0.169	0.396	0.434	0.188	-0.055	0.046	0.017	0.233	0.064
17	0.011	0.400	-0.113	-0.456	-0.063	0.005	-0.061	-0.131	-0.168	0.384	0.424	0.193	-0.033	0.058	0.019	0.221	0.089
18	0.038	0.406	-0.117	-0.437	-0.038	-0.012	-0.061	-0.134	-0.162	0.366	0.409	0.193	-0.014	0.069	0.059	0.206	0.081
19	0.037	0.410	-0.126	-0.412	0.018	-0.030	-0.062	-0.137	-0.175	0.343	0.390	0.189	0.001	0.077	0.055	0.189	0.070
20	0.037	0.411	-0.131	-0.396	0.060	-0.048	-0.063	-0.140	-0.168	0.319	0.370	0.182	0.012	0.083	0.063	0.175	0.046
21	0.036	0.405	-0.136	-0.383	0.082	-0.066	-0.063	-0.143	-0.165	0.299	0.348	0.170	0.019	0.085	0.028	0.158	0.032
22	0.036	0.398	-0.142	-0.362	0.059	-0.084	-0.065	-0.147	-0.163	0.282	0.324	0.153	0.023	0.084	0.059	0.139	0.021
23	0.035	0.389	-0.146	-0.342	0.040	-0.102	-0.067	-0.150	-0.165	0.264	0.300	0.141	0.020	0.079	0.082	0.119	0.000
24	0.035	0.379	-0.150	-0.312	0.009	-0.120	-0.069	-0.154	-0.162	0.248	0.276	0.132	0.012	0.071	0.053	0.096	0.005
# Lags Included	1-4	1-3	1, 2, 24	1-3	1-3	1-3	1-3	22-24	1-3	1	12	1-3	1-3	1-3	1-4	1-3	1-3

Table 26: Correlation of lagged features with non-lagged US recession. For US recession autocorrelation is shown. The last row shows the included lags for every feature in our random forest model before feature selection. Abbreviations represent the following: Unempl. (Unemployment), Industrial Prod. (Industrial Production), Personal Cons. Exp. (Personal Consumption Expenditures), Cons. Sent. (Consumer Sentiment), Macro Unc. 1M/12m (Macro Uncertainty Estimate 1M/12M ahead), Financial Unc. (Financial Uncertainty), Comp. Leading. Ind. (Composite Leading Indicator), Mortg. Rate (Mortgage Rate).

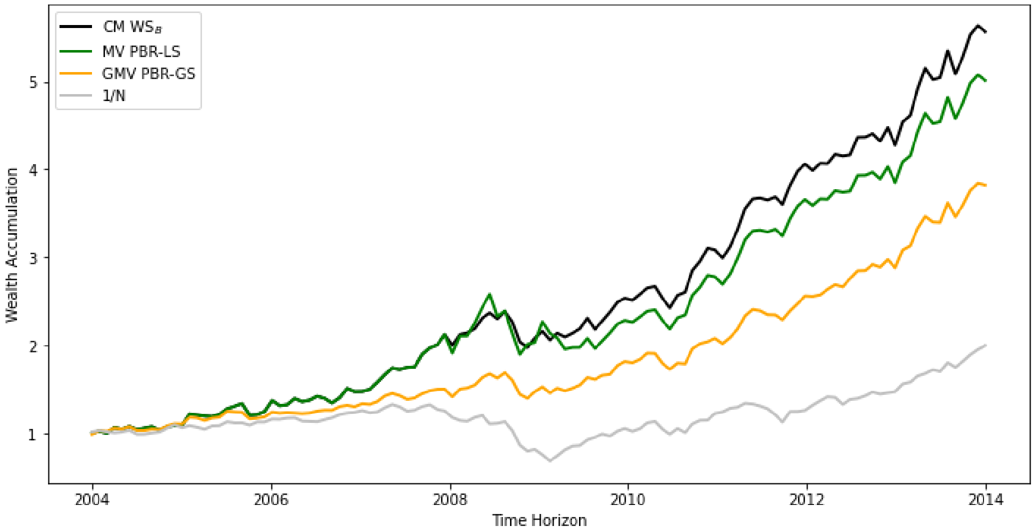

Figure 10: Wealth accumulation plot excluding transaction costs for the combined model with binary weighting scheme (CM WS$_B$), MV PBR-LS ($TR = 1.2$), GMV PBR-GS$^*_{BO}$ ($k = 3$) and the 1/N portfolio.

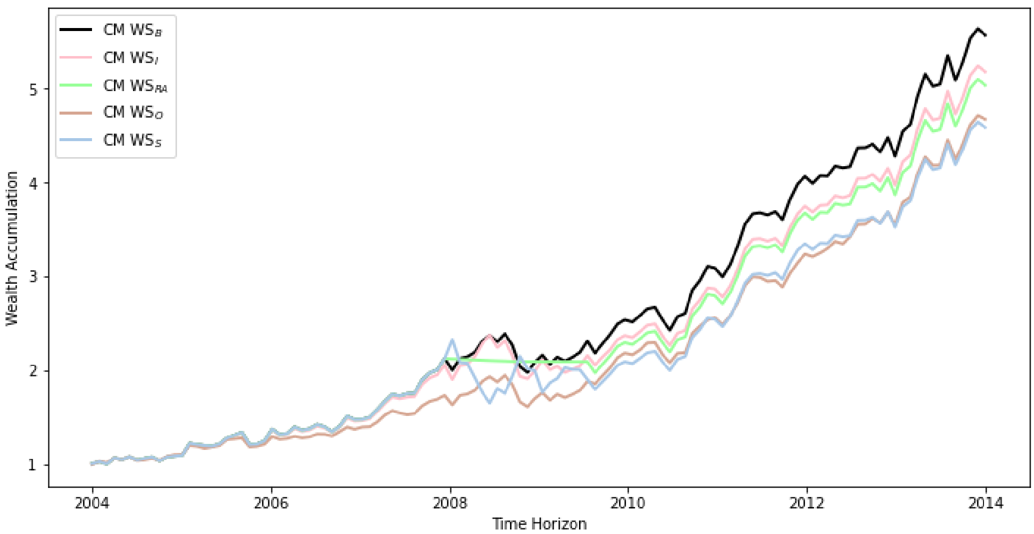

Figure 11: Wealth accumulation plot excluding transaction costs for the five combined models (CM) with binary (B), optimized (O), incremental (I), recession avoidance (RA) and short sell (S) weighting schemes.

9.4 Abbreviations

Abbreviation	Explanation
PBR	Performance-based regularization
FF5	Fama French 5 industry portfolios
FF10	Fama French 10 industry portfolios
FF49	Fama French 49 industry portfolios
Cnsmr nond.	Consumer nondurables
Cnsmr dur.	Consumer durables
SD	Standard deviation
SR	Sharpe ratio
CV	Cross-validation
TR	Target return
TO	Turnover
TC	Transaction costs
SHAP	Shapley additive explanation
GMV	Global minimum variance
MV	Minimum variance
SAA	Sample average approximation
PBR - LS	Performance-based regularization with backtracking line search for optimizing threshold U
PBR - GS	Performance-based regularization with grid search for optimizing threshold U
PBR - RS	Performance-based regularization with random search for optimizing threshold U
L1	Lasso
L2	Ridge
NS	No short selling
JS	James Stein
LW	Ledoit and Wolf
EW	Equally weigthed portfolio (1/N)
PBR - LS_{BO}	Portfolio with PBR - LS applied and weights bounded on (-1,1)
PBR - LS^*_{BO}	Portfolio with PBR - LS applied, weights bounded on (-1,1) and released restriction $w'\alpha > 0$
PBR - LS_{EW}	Mean variance portfolio with target return of the 1/N portfolio and PBR - LS applied
RF	Random forest
CM WS_B	Combined model weighting scheme binary
CM WS_O	Combined model weighting scheme optimized
CM WS_I	Combined model weighting scheme incremental
CM WS_{RA}	Combined model weighting scheme recession avoidance
CM WS_S	Combined model weighting scheme short selling

Table 27: Abbreviations on the left side with their full counterparts on the right side.