UiA University of Agder

Master

Is the Superior Performance of Optimized Strategies Caused by Data Mining Bias?

MARTIN SYVERT EGERHEI
HAAKON SEBASTIAN OLVIKEN

SUPERVISOR

Valeriy Zakamulin

University of Agder, 2020
Faculty of Business and Law
Department of Economics and Finance

Abstract

There has been a long-standing debate on whether optimized strategies can consistently outperform the naive diversification strategy with statistical significance, initiated by DeMiguel, Garlappi, and Uppal (2009). However, few of the papers contributing to this debate have presented the issue of data mining bias. Hsu, Han, Wu, and Cao (2018) and Yang, Cao, Han, and Wang (2019) corrected for this issue in their papers by applying joint testing. Motivated by the methodology applied by Hsu et al. (2018) and Yang et al. (2019), we aim to evaluate the performance of optimized strategies. To obtain our objective, we compared seven optimized strategies relative to the naive diversification strategy, using US and Norwegian return data. To cope with the issue of data mining bias, we applied the joint tests introduced by White (2000) and Hansen (2005), namely White's Reality Check (WRC) and Superior Predictive Ability (SPA). To measure the performance of the strategies, we used Sharpe ratio and alpha computed in Carhart's four-factor model. The results obtained using US data suggest that the best optimized strategy fails to outperform the naive diversification strategy. However, the results obtained with Norwegian data provides evidence that the best optimized strategy outperforms the naive diversification strategy.

Preface

This thesis has been written as the final part of our Master's Programme in Business Administration at the University of Agder. We would like to express our deepest gratitude to our supervisor, Professor Valeriy Zakamulin. His guidance has been crucial in the process of writing this master thesis, providing us with complex R code and motivating us to further investigate the subject. We would also like to thank the University of Agder for providing us with invaluable insight into economics, especially within the field of finance. Finally, we want to thank family and friends for all the support.

Table of Contents

List of Tables

List of Figures

1 | Introduction

Markowitz (1952) introduced the mean-variance model, which laid the foundation of modern portfolio theory (MPT). The uncertainty related to the model's estimations of the mean returns and the variance-covariance matrix has, however, made the implementation challenging. This has caused the model to produce extreme weights in the portfolio, leading to poor out-of-sample (OOS) performance.

Numerous studies have devoted considerable effort to minimize the estimation uncertainty, in order to improve the mean-variance model. A common solution has been to exclude mean returns from the computation. The basis of this approach is that the variance-covariance matrix is less exposed to estimation errors than estimated mean returns, making it a more reliable parameter. Several strategies that emerged in the aftermath of the mean-variance model have utilized the approach of removing mean return estimate. Clarke, De Silva, and Thorley (2006) proposed the minimum-variance strategy. Choueifaty and Coignard (2008) introduced the maximum diversification strategy to improve the poor diversification of the minimum-variance model. Maillard, Roncalli, and Teiletche (2010) exploited the middle ground between the minimum-variance strategy and equally weighted portfolios developing the equal risk contribution (ERC) portfolio. Asness, Frazzini, and Pedersen (2012) simplified the construction of the ERC portfolio, the optimized strategy is formally known as risk parity.

Despite the improvement in optimized strategies devoted to reduce estimation error, researchers still question whether these strategies add value. DeMiguel et al. (2009) initiated a debate concerning the performance of optimized strategies. The paper evaluated the performance of 14 optimized strategies relative to the naive diversification strategy and concluded that there is no statistically significant difference in performance. The naive diversification strategy allocates wealth equally among all assets available, relying on neither optimization nor parameter estimates (DeMiguel et al., 2009). The simplicity of the strategy has made it a preferred benchmark.

Several researchers have tried to defend the value of optimization in the time after DeMiguel et al. (2009) publication. Kritzman, Page, and Turkington (2010) showed in their paper that the minimum-variance and the mean-variance strategies were capable of outperforming the naive strategy. Kirby and Ostdiek (2012) introduced two new strategies as a counterweight to the mean-variance strategy, namely the volatility timing and the risk-to-reward timing strategies. The timing strategies are constructed to utilize low turnover and is less exposed to estimation risk. Kirby and Ostdiek (2012) demonstrated that their timing strategies were able to outperform the naive diversification strategy, with statistical significance.

There are, however, two major issues with the studies contributing to the debate on optimized strategies, namely the use of Sharpe ratio and individual testing of strategies. Kirby and Ostdiek (2012), Kritzman et al. (2010) and DeMiguel et al. (2009) applied data consisting of monthly portfolio returns from the US stock market, provided by Kenneth French's online library. Zakamulin (2017) presents evidence that low volatility anomalies are present in all of Kenneth French's datasets. The paper argued that for a convincing demonstration of superior performance by optimized strategies to hold, it cannot be by profiting from exposure to known anomalies. Further, Zakamulin (2017) argued that the Sharpe ratio applied in previous papers fails to correct for these anomalies.

Fama and French (1993) introduced alpha in the three-factor model, which can explain the effect of anomalies. Alpha computed in multifactor models adjusts for the low-volatility with the HML factor, correcting for this issue.

Testing multiple strategies individually causes data mining bias. This phenomenon leads researchers to falsely discover superior strategies, known as type I error. (Harvey & Liu, 2015). Hsu et al. (2018) and Yang et al. (2019) reassess the performance of the optimized strategies, by collectively testing strategies to correct for data mining bias. They also found that none of the optimized strategies in their study were able to consistently outperform the naive diversification strategy.

Motivated by Hsu et al. (2018) and Yang et al. (2019), we aim to provide further insight to the discussion initiated by DeMiguel et al. (2009). The objective of our thesis is to replicate and extend previous studies, by reassessing whether optimized strategies outperform the naive diversification strategy using "state of the art" methodology. In order to correct for data mining bias, we apply joint testing which relies on bootstrap. Our thesis extends previous studies by including recent data with a longer time sample using US and Norwegian data.

In this thesis, we measure the OOS performance of seven optimized strategies relative to the naive diversification strategy. The strategies have been evaluated using 18 US and 4 Norwegian datasets, consisting of monthly returns. The performance measures applied are Sharpe ratio, developed by Sharpe (1966), in addition to alpha computed in Carhart (1997) extension of the Fama and French (1993) three-factor model, namely Carhart's four-factor model (FFC4).

The results obtained from the individual tests in our thesis suggests that some optimized strategies performs significantly better than the benchmark. We then corrected for the data mining bias associated with individual testing, by applying the joint tests SPA and WRC. However, when there is a high correlation in returns, which is present in our data, SPA may produce deceptive results. Thus, we ignored the results from the SPA test.

When testing the performance of optimized strategies using US data and WRC, both Sharpe ratio and alpha computed in FFC4 suggest that the performance of optimized strategies is not better than the naive diversification strategy. This indicates that the best optimized strategy fails to outperform the naive diversification strategy in the US market. However, when we test the performance of optimized strategies using WRC and Norwegian data, we obtain evidence suggesting significant superiority from both Sharpe ratio and alpha FFC4. This indicates that the best optimized strategy outperforms the naive diversification strategy in the Norwegian market, with statistical significance. The results from this thesis does not provide an explanation for the differences, but we believe market efficiency may be an explaining factor.

The rest of our thesis is structured in the following way: Section **2** reviews the relevant literature. Section **3** presents the empirical data used in the thesis. Section **4** addresses the methodology relevant for the empirical study we conducted. Section **5** presents the empirical results from our study. Section **6** discusses the results. The final Section (**7**) provides the conclusion of our thesis.

2 | Literature Review

2.1 Modern Portfolio Theory and Optimized Portfolios

Construction of optimized strategies has historically been of interest within academia, and was actualized when Markowitz (1952) derived the optimal rule for diversification among risky assets. The framework developed by Markowitz (1952) is formally known as the mean-variance model, which laid the foundation for MPT. The mean-variance model suggests that investors should allocate wealth across assets, based on the expected returns and the variance-covariance matrix. This theoretical framework formed the efficient frontier, which is a set of optimized portfolios with the largest expected return for a given standard deviation (Merton, 1972). However, a practical implementation of the mean-variance model presents challenges. A minor change in the estimated parameters can lead to significant changes in the portfolio allocation, resulting in extreme weights (Merton, 1980). The process of estimating future forecasts with precision has also proven to be difficult (Chaves, Hsu, Li, & Shakernia, 2011). One of the most commonly used performance measures in MPT is Sharpe ratio, initially introduced by Sharpe (1966) as the reward-to-variability ratio. The measurement is used to obtain an understanding of the return from an investment relative to the risk. However, the ratio is not capable of providing information on whether superior performance occur due to better mean-variance efficiency or established factor premiums (Zakamulin, 2017).

In the aftermath of Markowitz's introduction of the mean-variance model, several optimized strategies have emerged in an attempt to reduce the model's uncertainty. The strategies presented below are some of these strategies, which do not rely on estimating mean returns. The reasoning behind this approach is that estimates of the variance-covariance matrix are less exposed to error estimation. Thus, is the variance-covariance matrix considered to be a more reliable parameter than the mean return estimates (Mausser & Romanko, 2014).

Clarke et al. (2006) introduced the minimum-variance strategy and demonstrated that the strategy produced low risk and a high Sharpe ratio. The strategy is located at the left most tip on the efficient frontier, which gives it the lowest attainable risk. It is tilted towards assets with low volatility, one can therefore consider the minimum-variance strategy as insufficiently diversified (Goldberg, Leshem, & Geddes, 2013). Choueifaty and Coignard (2008) proposed the maximum-diversification strategy. The authors suggested to use the most diversified portfolio to maximize the portfolio's diversification, as a solution to the poor diversification of the minimum-variance strategy. Maillard et al. (2010) considered a combination between minimum-variance and equally weighted portfolios, resulting in the equal risk contribution (ERC) portfolio. This approach aims to distribute risk equally among all assets in the portfolio. To simplify the construction of ERC, one assume the assets to be independent. Asness et al. (2012) suggested a risk parity portfolio based on this simplification of ERC. Risk parity is constructed to be risk averse, and therefore overweights less volatile assets.

2.2 Debate on the Value of Optimization

DeMiguel et al. (2009) initiated a debate concerning the performance of optimized strategies, with their paper comparing performance of optimized strategies relative to the naive diversification strategy. The naive diversification strategy demonstrated good results and in several cases outperformed the optimized strategies (DeMiguel et al., 2009). The paper concluded that there is no statistical evidence of superiority related to optimized strategies.

Kritzman et al. (2010) argued in defence of optimized strategies, by demonstrating that they outperform the naive diversification strategy. However, the study was conducted without comparing statistical differences in the Sharpe ratios. Another study that argued in defence of optimized strategies were Kirby and Ostdiek (2012). The authors introduced two alternative strategies, volatility-timing and risk-to-reward timing. These strategies were developed to mitigate the effect of estimation errors, by focusing on volatility and return from assets. Thus, ignoring the correlation between assets to combine the optimal allocation. They argued that by focusing on the diagonal of the variance-covariance matrix, the strategies can reduce estimation error and therefore outperform the naive diversification strategy. Using Sharpe ratio they demonstrated that these strategies outperformed the naive diversification strategy. The study only used four datasets to evaluate the performance of the optimized strategies, which can lead to insufficient results.

There are two issues present in the studies that argued in defence of optimized strategies (Kritzman et al., 2010; Kirby & Ostdiek, 2012). The first issue being that the papers only use Sharpe ratio as portfolio performance measurement. The second issue arises when the papers by Kritzman et al. (2010) and Kirby and Ostdiek (2012) proceed to evaluate the performance using individual tests. Evaluating the performance of the strategies individually raises the concern of whether strategies are exposed to data mining bias. In the following Sections (**2.3** and **2.4**) we present methods to cope with these issues.

2.3 Multifactor Models and Asset Pricing Anomalies

Zakamulin (2017) showed that all recent empirical studies surrounding portfolio optimization use the Sharpe ratio as performance measure. Thus, the studies has been conducted without controlling whether the superior performance of these optimized portfolios appears due to exposures to one or several profitable anomalies. However, alpha motivated by multifactor models accounts for various anomalies that can be exploited to influence the Sharpe ratio.

Sharpe-Lintner-Mossin mean-variance equilibrium model of exchange, also referred to as the capital asset pricing model (CAPM), is one of modern capital market theories most critical developments. The model explains the relationship between systematic risk and expected return for an asset. CAPM has been a popular measurement among researchers and is still widely used.

Since CAPM was first introduced, there have been made discoveries of expected return samples that the model was unable to explain. This phenomenon is referred to as market anomalies and are often related to size and value. Fama and French (1993) identified three risk factors in returns from stocks, extending CAPM to Fama and French's three-factor model. The three factors covering the stock market; (i) overall market, (ii) factors related to firm size and (iii) book-to-market equity factors. They claimed that these factors were able to describe the average return anomalies. The three-factor model, however, fails to obtain observation from the cross-sectional variation in momentum portfolios. Carhart (1997) made an extension to the three-factor model, that captures the momentum anomaly identified by Jegadeesh and Titman (1993).

2.4 Data Mining Bias

The issues related to data mining have been a well-known phenomenon within the field of portfolio performance, after it was highlighted by Leamer (1978, 1983). Data mining bias leads researchers to falsely discover superior strategies, known as Type I Error (White, 2000). This suggest that observed superior performance in some instances can be attributed to randomness. The phenomenon is prominent when strategies are tested individually, the number of false discoveries increases with the number of strategies tested. This relationship is known as the false discovery rate (FDR). To correct for data mining bias, one tests the performance of all the strategies collectively, formally known as joint testing.

White (2000) introduced White's Reality Check (WRC) test, in an attempt to cope with data mining bias. WRC provide the framework to collectively test optimized strategies. WRC tests the null hypothesis that the best optimized strategy among other possible strategies does not outperform the benchmark. The joint test that WRC provide has later become a standard procedure in a number of studies. Hansen (2005) later revised the procedure introduced by White (2000) and proposed a new test called SPA. Hansen (2005) suggested two improvements of the WRC test: (i) Normalize the test statistics. Without normalizing one might compare two different things. (ii) Exclude strategies with poor performance, which removes the unfavorable effect they might have in a joint test. Both SPA and WRC are based on the family-wise error rate (FWER), measuring the probability of wrongly discover superior strategies (Yang et al., 2019).

Hsu et al. (2018) reassessed the out-of-sample performance of various optimized strategies, using some advanced tests from WRC, collectively testing all optimized strategies relative to the naive diversification strategy. Their study found that some strategies provided superior performance when using an individual test to evaluate the performance. However, when they corrected for data mining bias using WRC, almost none of the optimized strategies demonstrated superior performance. This demonstrates the importance of controlling for data mining bias. Yang et al. (2019) conducted a similar study, where they evaluated the performance of tactical asset allocation on technical trading rules. They applied advanced extensions of both SPA and WRC in their joint testing, to correct for data mining bias. The paper arrives at the same conclusion as Hsu et al. (2018), that there is no evidence suggesting optimized strategies are superior to the naive diversification strategy.

3 | Data

The data applied in this thesis are monthly return data from US and Norwegian stock portfolios, in addition to the four research factors in Carhart's model: (i) Market return minus risk free rate (MKTRF), (ii) Small-Minus-Big (SMB), (iii) High-Minus-Low (HML) and (iv) Momentum (MOM).

US return data have been retrieved from the online library provided by Kenneth R. French[1]. The start of each US dataset was set to July 1963 in our research, to cope with the different starting times within each dataset. The end was set to December 2019, as this was the most recent available data. We included a total of 18 datasets from US, which are presented in **Table 1**.

[1] Data retrieved from: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

Table 1: Kenneth French Datasets

#	**Dataset**	**N**
1	Portfolios formed on Size	10
2	Portfolios formed on Book-to-Market	10
3	Portfolios formed on Industry	10
4	Portfolios formed on Short-Term-Reversal	10
5	Portfolios formed on Long-Term-Reversal	10
6	Portfolios formed on Market Beta	10
7	Portfolios formed on Variance	10
8	Portfolios formed on Accruals	10
9	Portfolios formed on Residual Variance	10
10	Portfolios formed on Earnings-to-Price	10
11	Portfolios formed on Cash-Flow-to-Price	10
12	Portfolios formed on Dividend Yield	10
13	Portfolios formed on Momentum	10
14	Portfolios formed on Operating Profitability	10
15	Portfolios formed on Investment	10
16	Portfolios formed on Net Share Issues	10
17	Portfolios formed on Size and Book-to-Market	25
18	Portfolios formed on Size and Operating profitability	25

 # is the number of the portfolio in the series. N is the number of portfolios in each dataset. Dataset describes the variable that the datasets are based on.

The Norwegian data applied in the study have been retrieved from the online library provided by Bernt Arne Ødegaard[2]. The number of Norwegian datasets included in our paper were four, due to the limitation of accessible datasets. The datasets are presented in **Table 2**. To cope with the different starting times within each dataset, we set the start of each Norwegian dataset to September 1981. The end was set to December 2019, as this was the most recent available data.

Table 2: Bernt Arne Ødegaard Datasets

#	**Dataset**	**N**
1	Portfolios formed on Size	10
2	Portfolios formed on Book-to-Market	10
3	Portfolios formed on Momentum	10
4	Portfolios formed on Spread	10

 # is the number of the portfolio in the series. N is the number of portfolios in each dataset. Dataset describes the variable that the datasets are based on.

[2]Data retrieved from: `http://finance.bi.no/~bernt/financial_data/ose_asset_pricing_data/index.html`

4 | Methodology

In this section we aim to provide an explanation of the methods applied in our research. We start by presenting the naive diversification strategy (benchmark), followed by the seven optimized strategies we implemented. In the second subsection we introduce our performance measures, Sharpe ratio and alpha computed in FFC4. We then explain the bootstrap methodology. The fourth subsection examines OOS estimation. In the fifth subsection we present hypothesis testing, including both individual- and joint hypothesis testing. Then we proceed to review SPA and WRC test, to correct data mining bias.

4.1 Portfolio Strategies

We have implemented seven optimized strategies, in addition to the naive diversification strategy in this empirical study. These eight strategies are presented in **Table 3**.

Table 3: Portfolio strategies included in the thesis

#	Strategy	Abbreviation
Benchmark:		
0	Naive diversification	$1/N$
Optimized portfolios:		
1	Mean-variance	MV
2	Minimum-variance	MIN
3	Maximum-diversification	MD
4	Equal risk contribution	ERC
5	Risk parity	RP
6	Volatility-timing	VT
7	Risk-reward-timing	RRT

 # is the number of the portfolios. Strategy denotes the strategy name.

We implement two restrictions in the construction of the portfolios. Eq.(1): Setting the sum of all weights equal to 1 (100%), which is a standard assumption in portfolio optimization. Eq.(2): None of the weights can be negative (short sale restriction). The short constraint is implemented to make our research more applicable to real-life scenarios, since portfolio managers usually are not allowed to take short positions.

$$\mathbf{w}'\mathbf{1} = 1, \tag{1}$$

where, $\mathbf{w}$ is a vector of portfolio weights of the risky assets and $\mathbf{1}$ is a vector of ones:

$$\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{pmatrix}, \ \mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

$$\mathbf{w} \geq \mathbf{0}, \tag{2}$$

where, $\mathbf{0}$ is a vector of zeros:

$$\mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

To explain the notations applied in the methodlogy, we use the following example: Assume that the investor invests w_i of his wealth in asset i, the return on the portfolio is then given by:

$$x_p = \sum_{i=1}^{N} w_i x_i, \text{ subject to } \mathbf{w}'\mathbf{1} = 1. \tag{3}$$

We denote the expected return on asset x_i by $E[x_i] = \mu_i$ and the variance by $Var[x_i] = \sigma_i^2$. We also denote $Cov(x_i, x_j) = \sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$ where $\rho_{ij} = \rho_{ji}$ is the correlation coefficient between the returns on asset i and j. The matrix notation of the mean returns on the risky assets and the variance-covariance matrix can be expressed in the following way:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_N \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \dots & \sigma_{NN} \end{pmatrix}.$$

Accordingly, the portfolios mean return and variance are given by:

$$\mu_p = \mathbf{w}'\boldsymbol{\mu}, \quad \sigma_p^2 = \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}.$$

4.1.1 Naive Diversification (Benchmark)

The naive diversification strategy (1/N) given in Eq.(4) is the simplest diversification strategy, and distributes wealth without any optimization. The strategy "naively" distributes wealth equally among all assets accessible, thus it is not affected by estimation errors.

$$w_i = \frac{1}{N}, \tag{4}$$

where w_i denotes the weight of wealth invested in asset i, 1 represents the total wealth and N is the number of assets.

4.1.2 Mean-Variance

The mean-variance model was introduced by Markowitz (1952), where the investor optimize the relation between mean and variance of portfolio returns (DeMiguel et al., 2009). To obtain a desired value of mean returns μ^*, the strategy construct a portfolio with the lowest feasible variance. Thus, we have to identify the minimum-variance that has the mean return of μ^*. In this situation $\mathbf{w}$ is one solution to the quadric problem given by:

$$\min_{\mathbf{w}} \ \frac{1}{2}\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} \ \text{ subject to } \ \mathbf{w}'\boldsymbol{\mu} = \mu^*. \tag{5}$$

However, a problem arises when return from the risk-free asset exceeds the mean of the minimum-variance portfolio. The optimal strategy is then to short sell the risky asset and allocate that capital in the risk-free asset. Another problem occurs when the risk aversion of the investor is such that he prefers to borrow the risk-free asset. Short selling is problematic due to the risk involved in this procedure. Prohibiting short sales cope with these issues and is therefore included in the MV strategy.

To accommodate the issues in Eq.(5), we instead solve the maximization problem in Eq.(6) to obtain the MV portfolio. We maximize the expected utility U of the risky assets $\mathbf{w}$ in the portfolio:

$$\max_{\mathbf{w}} U(\mathbf{w}) = (\boldsymbol{\mu} - \mathbf{1}r_f)'\mathbf{w} + r_f - A\frac{1}{2}\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} \text{ subject to } \mathbf{w}'\mathbf{1} = 1 \text{ and } w_i \geq 0 \ \forall_i, \quad (6)$$

where A is the investor's risk aversion. r_f is the return on the risk-free asset.

4.1.3 Minimum-Variance

The minimum-variance strategy has the unique property of reducing risk by only relying on the variance-covariance matrix. Estimates from the matrix are less exposed to error estimation, which in turn is meant to make the strategy more robust and lead to a more precise and reliable result (Clarke et al., 2006). The MIN portfolio is located farthest to the left in the efficient frontier, which gives it the lowest attainable risk. To derive the weights of the MIN strategy we solve the following minimization problem:

$$\min_{\mathbf{w}} \frac{1}{2}\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} \text{ s.t } \mathbf{w}'\mathbf{1} = 1 \text{ and } w_i \geq 0 \ \forall_i, \quad (7)$$

4.1.4 Maximum-Diversification

Choueifaty and Coignard (2008) introduced the diversification ratio (DR), where the weighted average of standard deviations are divided by the portfolio standard deviation. The motivations behind DR is; the advantage of predicting volatility rather than estimating returns and the poor diversification in the MIN portfolio. DR is given by:

$$\mathrm{DR} = \frac{\mathbf{w}'\boldsymbol{\sigma}}{\sqrt{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}}}, \tag{8}$$

where $\boldsymbol{\sigma} = (\sigma_1, ..., \sigma_N)'$ and σ_i is the standard deviation of asset i. To compute the weights of the MD portfolio, we maximize the diversification ratio subject to the weight constrain. The derivation of the MD portfolio can thus be expressed as:

$$\max_{\mathbf{w}} \frac{\mathbf{w}'\boldsymbol{\sigma}}{\sqrt{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}}} \text{ s.t } \mathbf{w}'\mathbf{1} = 1 \text{ and } w_i \geq 0 \ \forall_i. \tag{9}$$

4.1.5 Equal Risk Contribution

The objective of the strategy is to equally allocate risk among all assets, making each asset contribute equally to the overall portfolio risk. ERC is motivated by observations of 60/40 portfolios consisting of stocks and bonds, where stocks contribute to more than 90% of the risk in the portfolio (Chaves et al., 2011). ERC portfolios are considered to be risk-averse, since it overweight less volatile assets (Asness et al., 2012). Total risk of the portfolio is given by:

$$\sigma_p = \sqrt{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}}. \tag{10}$$

The relative risk contribution (RRC) is given by:

$$RRC_i = \frac{w_i(\boldsymbol{\Sigma}\mathbf{w})_i}{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}}, \tag{11}$$

which satisfy the RRC constraint:

$$\sum_{i=1}^{N} RRC_i = 1. \tag{12}$$

The sum of RRC equals 1 (100%) of the portfolios risk. The goal of the strategy is to distribute risk equally among N assets:

$$RRC_i = \frac{1}{N}. \tag{13}$$

To achieve the objective given in Eq.(13), we need to solve the optimization problem of the ERC portfolio. The problem can be expressed as a Sequential Quadratic Programming (SQP) problem:

$$\min_{\mathbf{w}} \sum_{i=1}^{N}\sum_{j=1}^{N}\Big(w_i(\mathbf{\Sigma w})_i - w_j(\mathbf{\Sigma w})_j\Big)^2 \quad \text{s.t} \;\; \mathbf{w'1} = 1 \text{ and } \mathbf{w} \geq \mathbf{0}\ \forall_i, \tag{14}$$

where $(\mathbf{\Sigma w})_i$ is the i-th weight of vector $\mathbf{\Sigma w}$. A simpler alternative solution to the problem is solving the nonlinear convex optimization problem:

$$\min_{\mathbf{w}} \; \mathbf{w'\Sigma w} \text{ s.t} \sum_{i=1}^{N} \log{(w_i)} \geq c, \tag{15}$$

where $c \geq 0$ represents a random positive constant. However, this solution does not satisfy the budget constraints, therefore the weights have to be normalized after the solution have been obtained.

4.1.6 Risk Parity

In the construction of the risk parity portfolio, the correlation assumptions between assets play an important role. To simplify the construction one assumes that the asset returns are independent. The RP portfolio weights are given by:

$$w_i = \frac{\frac{1}{\sigma_i}}{\sum_{i=1}^{N}\frac{1}{\sigma_i}}, \tag{16}$$

where the weight of asset i is inversely proportional to its standard deviation.

One assumes that the correlation between assets is the same, $\rho_{ij} = \rho$. The correlation matrix diagonal is thus $\rho_{ii} = 1$, which allows for simplification of the ERC portfolios.

4.1.7 Volatility Timing and Risk-Reward Timing

Kirby and Ostdiek (2012) introduced two methods of portfolio optimization, formally known as Volatility-Timing strategy and Risk-to-Reward timing strategy. Both methods are derived from the MV strategy, which suggest a new class of active portfolio strategies. These strategies are outlined to exploit sample information regarding volatility dynamics in order to reduce the effect of estimation risk. The strategy suggests rebalancing the weights monthly, in accordance with the changes in the tuning parameter, that are estimated from the variance-covariance matrix. This allows us to control the sensitivity of the portfolio and measure the portfolio's timing aggressiveness. There are four elements that emphasize the characteristics of the timing strategies: (i) It does not require optimization. (ii) the strategy does not require covariance matrix inversion. (iii) Both strategies assures non-negative weights. (iv) Through volatility changes, the sensitivity of the portfolio weights can be adjusted with a tuning parameter (Kirby & Ostdiek, 2012). The asset weights in the VT strategy are given by:

$$w_i = \frac{(1/\sigma_i^2)^\eta}{\sum_{i=1}^{N}(1/\sigma_i^2)^\eta} \qquad i = 1, 2, \ldots, N, \tag{17}$$

where, $\eta > 0$.

We follow the procedure of Zakamulin (2017) and set $\eta = 4$. The simplicity and long weights leads to parameters being less affected by estimation risk. Kirby and Ostdiek (2012) argue that superior performance occur when increasing the parameter η, since this will reduce the portfolio's transaction and turnover costs.

The RRT strategy emphasizes the same elements as the Volatility-Timing strategy. Kirby and Ostdiek (2012) suggest to include sample information about the dynamics of conditional expected returns. How one operate the RRT strategy with information regarding expected return, is the key difference between the two timing strategies. We use two estimators regarding conditional expected returns. The first estimator is constructed to reduce risk by taking advantage of asset pricing theory. The second estimator is a simple rolling estimator that inflict no forecast assumptions (Kirby & Ostdiek, 2012). The asset weights in RRT are given by the following equation:

$$w_{it} = \frac{\left(\frac{(\mu_i - r_f)^+}{\sigma_i^2}\right)^\eta}{\sum_{i=1}^N \left(\frac{(\mu_i - r_f)^+}{\sigma_i^2}\right)^\eta}, \tag{18}$$

where,

$$(\mu_i - r_f)^+ = \max(\mu_i - r_f, 0). \tag{19}$$

For both strategies, the tuning parameter η is included, where $\eta \geq 0$. This parameter measures the timing of aggressiveness. When $\eta \to 0$ we retrieve the $1/N$ strategy. When $\eta \to \infty$ the weights will approach 1, for the asset with the lowest volatility. If we set $\eta > 1$ we will make up for information lost by ignoring the correlations.

4.2 Performance Measures

4.2.1 Sharpe Ratio

The Sharpe ratio was first introduced by Sharpe (1966). The objective of the Sharpe ratio is to measure the relation between mean and standard deviation of the excess return. The ratio aims to provide an understanding of the risk related to the excess return (Auer & Schuhmacher, 2013). The simplicity of comparing performance of portfolios with various risk exposure, have made the Sharpe Ratio a popular model. The ratio can be expressed in the following way:

$$SR = \frac{\mu_i - r_f}{\sigma_i} \tag{20}$$

Although the measure is widely used and recognized within financial research, it has some limitations. For instance, if the returns are not normally distributed, the results may be deceiving. Jobson and Korkie (1981) developed a hypothesis test for SR with the null hypothesis:

$$H_0 : SR_k \leq SR_0. \tag{21}$$

Memmel (2003) revised the test and simplified the derivation of the ratio (DeMiguel et al., 2009).

We test the null hypothesis developed by Jobson and Korkie (1981), in order to evaluate which portfolio that achieves the highest Sharpe ratio:

$$z = \frac{SR_k - SR_0}{\sqrt{\frac{1}{T}\left[2(\mathbf{1} - \rho) + \frac{1}{2}(SR_k^2 + SR_0^2 - 2SR_k SR_0 \rho^2)\right]}}, \tag{22}$$

where SR_k and SR_0 denotes the Sharpe ratios of the optimized and the benchmark strategy. T denotes the sample size. ρ denotes the correlation coefficient over a sample given by T. The z-test is distributed asymptotically as a standard normal.

4.2.2 Carhart's Four-Factor Model

Alpha is one of the most common performance measures within finance (Gerber & Hens, 2006). Alpha motivated by CAPM (single-factor model) is estimated using Ordinary Least Squares (OLS) and consists of one market factor. Namely, the market excess return.

Fama and French (1993) believe there are two additional factors that have to be taken into account, to explain the size and value anomalies. The additional factors are SMB and HML. The idea behind these factors are that small/value stocks are riskier than large/growth stocks and thus they provide a special risk premium. SMB aims to imitate the risk factor in returns linked to size, while HML is supposed to imitate the risk factor in returns related to the book-to-market equity (Fama & French, 1993). Their model is referred to as the Fama-French three-factor model (FF3). Carhart (1997) made an extension to the three-factor model introducing the four-factor model (FFC4), which can obtain observation from the cross-sectional variation in momentum portfolios. The model capture Jegadeesh and Titman (1993) momentum factor, in addition to the factors the three-factor model capture. The FFC4 model is given by:

$$R_{i,t} = \alpha_i + \beta_i R_{M,t} + s_i SMB_t + h_i HML_t + p_i PR1YR_t + \epsilon_{i,t}, \tag{23}$$

where, R_{it} denote the return of a portfolio in excess of r_t. $PR1YR$ denote returns in a one year momentum.

The multiple factor model is estimated using OLS, then tested with the null hypothesis:

$$H_0 : \alpha_k \leq \alpha_0, \tag{24}$$

in order to find and evaluate the significance of the alpha. The test of the null hypothesis is given by:

$$z = \frac{\alpha_k - \alpha_0}{\sqrt{\frac{1}{T}\left(\sigma_k^2 - 2\rho\sigma_k\sigma_0 + \sigma_0^2\right)}}, \tag{25}$$

where α_k and α_0 denotes the alpha values of the optimized and benchmark strategies. T denotes the sample size. ρ denotes the correlation coefficient between the residuals, obtained in the linear regression. σ_k and σ_0 denotes the standard errors. The z-test is distributed asymptotically as a standard normal.

4.3 Bootstrap

The implementation of the bootstrap method depends crucially on whether the return is assumed dependent or independent. If the data is assumed to be dependent, a method called block bootstrap is applied, introduced by Hall (1985). Hall (1985) and Kunsch (1989) bootstrap methods operate with overlapping blocks of data to calculate estimates of distribution and parameters. Bootstrapping do not require any parameters of the probability distribution, and effectively utilize observations from minor sample sizes (Cogneau & Zakamouline, 2013).

When conducting the block bootstrap method with dependent data, it is important to keep the dependency intact in observations. In accordance to the block bootstrap we need to take a set or block of observations into account when calculating statistics or parameter estimates.

We follow the approach introduced by Politis and Romano (1994), called the stationary block bootstrap method. The block lengths are random and the resample will be stationary. For the stationary block bootstrap the length of each block is developed from geometric distribution. This type of simulation is referred to as "geom", and is the process of generating the lengths of the blocks in the resampling operation. In order to choose the block length we can use an approach by Politis and White (2004) with the correction by Patton, Politis, and White (2009).

4.4 Out-Of-Sample Testing

We apply backtesting to test the performance of strategies. The procedure simulate strategies using historical data, which allows us to simultaneously test several strategies. This gives an understanding of how the strategies performed in a given time period. We divide the time sample into "in-sample" (prior to simulation) segment and "out-of-sample" (simulation) segment, in order to simulate a real trading scenario. The data in the in-sample period t is used to estimate the parameters applied in the out-of-sample simulation. By only allowing in-sample data to be used, the issue of "peeking" is eliminated.

The parameters used in the OOS simulation are estimated using a lookback period, which consists of a training period and a rolling lookback window. Training period is the time prior to the simulation, also known as in-sample period. When the simulation starts, the training period moves forward in the same frequency as the simulation, this procedure is known as rolling lookback window. The rolling lookback window estimates parameters in a given period ahead of the OOS estimations. The length of the training period and the rolling lookback window period can vary (between three and twenty years is recommended), however, there is no scientific justification for how long these periods actually should be. We have chosen five years, based on the sample time in our study and the precedence within portfolio research. It is common to assume that returns are time-invariant and slowly changing, which allows us to use the sample mean and variance provided by the lookback period

4.5 Hypothesis Testing and Data Mining Bias

False discoveries of profitable strategies have been a known phenomenon within financial research. Some of these false rejections have been caused by the individual hypothesis test, when more than one hypothesis was tested (Type I Error). Joint hypothesis testing aims to cope with the issue of false rejections in individual hypothesis test.

When conducting a individual test, a significance level of 5% for a individual strategy is sufficient to reject the null hypothesis. However, this criterion is not sufficient for joint hypothesis testing. This is because we expect that some of the variables will be significant at a 5%-level, just by chance, when testing several strategies. The observed outperformance can thus be explained by the two factors:

$$\text{Observed outperformance} = \text{True outperformance} + \text{Randomness}. \tag{26}$$

Further, we assume that the performance of the optimized strategies are equal to the benchmark performance, the observed outperformance can thus be given by:

$$\text{Observed outperformance} = \text{Randomness}. \tag{27}$$

Given that Eq.(27) is true, the z test statistic is normally distributed $z \sim \mathcal{N}(0, 1)$. The p-value of a individual test is given by Eq.(28), where we assume that all returns are independent. The generalization of the returns distribution allow us to study returns that are correlated.

$$p^S = Pr(z > z_{1-\alpha}), \tag{28}$$

where $z_{1-\alpha}$ denotes the $1 - \alpha$ quantile of $z \sim \mathcal{N}(0, 1)$. Given a significance level of $\alpha = 0.05$, the probability of false discovery is 5% in each test. If the researcher has studied many strategies and only selects the best strategy, the p-value is likely to overestimate the significance of the strategy.

In order to evaluate the overestimation of the best strategy, we assume that the researcher only presents the best strategy out of N strategies and the test statistics are independent for these N strategies. The p-value of obtaining at least one significant value by chance, when testing N strategies, is given by:

$$p^N = 1 - Prob(z_1 < z_{1-\alpha}; z_2 < z_{1-\alpha}; ...; z_N < z_{1-\alpha}) = 1 - (1 - p^S)^N, \qquad (29)$$

where z_k, $k \in [1, N]$, denotes the value of the z test statistic for strategy k. If p^S is 5% and N is 100, then p^N is 99.4%. This means that there is close to 100% likelihood that one or more strategies outperform the benchmark, just by chance. It is therefore reasonable to assume that the best selected strategy benefited from randomness, given that true outperfomance is equal to zero. We employ the joint tests SPA and WRC to correct for data mining bias.

4.5.1 White's Reality Check

White (2000) introduced an approach to conduct joint tests, known as White's reality check (WRC). This test allows the researcher to simultaneously test multiple strategies, while maintaining a low probability of falsely discovering profitable strategies. For simplicity, we only demonstrate the WRC framework using Sharpe ratio, however, the same approach applies for alpha. We measure the outperformance of each optimized strategy:

$$f_k = SR_k - SR_0, \qquad (30)$$

where SR_k denotes Sharpe ratio of optimized strategy k, and SR_0 denotes Sharpe ratio of the $1/N$ strategy. f_k denotes the outperformance measure. We want to find the strategy with the greatest observed outperformance:

$$\bar{f} = \max_{k=1,...,N} f_k, \qquad (31)$$

where $\bar{f}$ denotes the best strategy. We want to check if the best optimized strategy can outperform our benchmark strategy. We use the null hypothesis: that the best strategy does not outperform the benchmark:

$$H_0 : \bar{f} \leq 0. \tag{32}$$

We need to know the probability distribution of $\bar{f}$, in order to test the null hypothesis. In the computation of WRC we implement a geometric block bootstrap to find the probability distribution of $\bar{f}$. When conducting the bootstrap, we resample the excess returns to the optimized strategies and the excess returns to the naive strategy simultaneously. We denoted the resample series of strategy k by $r^*_{j,k}$. The series provided by the benchmark is denoted by $r^*_{j,0}$. Where j indicates the repetition number in the bootstrap. Note($*$) denotes that r^* is a resampled (bootstraped) version of r.

We compute the Sharpe ratios for each strategy after implementing the bootstrap, thus $SR^*_{j,k}$ denotes Sharpe ratio of strategy k, and $SR^*_{j,0}$ denotes the Sharpe ratio of the $1/N$ strategy. To compute the outperformance we use:

$$f^*_{j,k}{=}SR^*_{j,k}{-}SR^*_{j,0}, \tag{33}$$

where $f^*_{j,k}$ denotes the bootstrapped outperformance measure between the k optimized strategy and the benchmark. The strategy with the greatest observed outperformance is given by:

$$\bar{f}^*_j = \max_{k=1,\ldots,N} \left(f^*_{j,k} - f_k\right), \tag{34}$$

where we subtract the observed outperformance (f_k) from the resampled observed outperformance ($f^*_{j,k}$), in order to adjust our computation of the outperformance to conform the null hypothesis. The collection of $\bar{f}^*_j$ defines the probability distribution of $\bar{f}$. To calculate the p-value, we have to check how many times $\bar{f}^*_j$ is observed higher than $\bar{f}$. The computation of the p-value is given by:

$$P_{WRC} = \sum_{j=1}^{J} \frac{1 \bar{f}^*_j > \bar{f}}{J}, \tag{35}$$

where 1 denotes the indicator function that takes the value of one, if there are observations of the condition being satisfied.

4.5.2 Superior Predictive Ability

Hansen (2005) introduced the SPA test, which extends WRC. Hansen added two improvements opposed to the WRC. The first improvement the author suggested is to normalize the test statistics. Without normalizing the test statistics, we may compare two different models that are incomparable. Kosowski, Timmermann, Wermers, and White (2006) address and confirm this problem in their paper as well. Hansen suggest to substitute f_k with:

$$z_k = \frac{f_k}{\hat{\sigma}_{fk}}, \tag{36}$$

where $\hat{\sigma}_{fk}$ denotes the standard deviation of f_k. This allows us to formulate a null hypotheses of interest, that the statistics of standard deviation and mean can not exceed zero:

$$H_0 : \bar{z} \leq 0, \tag{37}$$

where,

$$\bar{z} = \max_{k=1,\dots,N} z_k. \tag{38}$$

The second improvement Hansen suggested in his paper was to exclude poor models from the test. Hansen argued that poor models could manipulate the result in the WRC. When conducting tests of various strategies, one will not be able to obtain the results from a strategy that produce good results, among strategies that produce poor results. In order to determine which strategies that produce poor results, a threshold value is set. The statistic threshold value A is given by:

$$A = -\sqrt{2ln(ln(n))}, \tag{39}$$

where n denotes the number of strategies. In our thesis, we apply two performance measures, namely Sharpe ratio and alpha. We compare each strategy to the threshold value, based on which performance measure we use, and exclude the strategies where z_k is below the threshold value A. Thus, the optimized strategy is excluded from the test if:

$$z_k < A. \tag{40}$$

5 | Empirical Results

In this section we present the results obtained from OOS simulated strategies, using US and Norwegian monthly return data. The parameters in the empirical study are estimated using a training period and a rolling lookback period of five years (60 months), based on limitations in sample time and previous research. It is common to assume that return distributions are time-invariant and slowly changing, which allows us to use the sample mean and variance provided by the lookback period.

We apply the joint tests SPA and WRC, which are conducted with Sharpe ratio and alpha computed in FFC4. The results are provided by testing the null hypothesis $H_0 : \bar{f} \leq 0$, that the best optimized strategy fail to outperform the $1/N$ strategy.

5.1 Individual Tests

The Sharpe ratios of all strategies and p-values from the individual hypothesis test $H_0 : SR_k \leq SR_0$ are presented in **Table 4**. From this table we observe that about 40% (51 out of 126) of the optimized strategies in the US market achieve significantly better Sharpe ratio than the benchmark. In the Norwegian markets about 10% (3 out of 28) of the optimized strategies outperform the benchmark, in terms of Sharpe ratio. The differences in the US and Norwegian data may be caused by the number of datasets applied, or other market factors Sharpe ratio fails to measure.

Carhart's (1997) four-factor model accounts for four market factors (MKTMRF, HML, SMB and MOM) when measuring the alpha value. These factors may be able to explain some of the performance of optimized strategies, that the Sharpe ratio fails to measure. Alpha values from all strategies and p-values from the individual hypothesis test $H_0 : \alpha_k \leq \alpha_0$ are presented in **Table 5**. We observe that about 18% (23 out of 126) of the optimized strategies in the US market achieve significantly higher alpha than the benchmark. In the Norwegian markets about 20% (6 out of 28) of the optimized strategies significantly outperform the benchmark.

Comparing the Sharpe ratio and alpha from the US market, we see that there is a (23/51) 55% reduction in significant results. However, there is (6/3) 100% increase in significant results in the Norwegian market. The differences are likely caused by the different factors applied by the models.

Using individual test to find the best performing strategies have been a standard procedure within financial research (DeMiguel et al., 2009; Kirby & Ostdiek, 2012). This methodology has been used to produce the results found in this section. However, this approach has several weaknesses pointed out in the methodology in Section **4.5**. The main issue being false discovery of profitable strategies, known as data mining bias.

Table 4: Annualized Sharpe ratios and p-values from $H_0 : SR_k \leq SR_0$.

	Naive	MV	Min	MD	ERC	RP	VT	RRT
US Data:								
Portfolios formed on Size	0.407	0.425 (0.815)	0.407 (0.99)	0.392 (0.629)	0.413 (0.25)	0.413 (0.214)	0.428 (0.309)	0.45 (0.49)
Portfolios formed on Book-to-Market	0.479	0.454 (0.696)	0.485 (0.876)	0.434 (0.094)	0.484 (0.172)	0.486 (0.074)	0.499 (0.091)	0.524 (0.356)
Portfolios formed on Industry	0.463	0.434 (0.743)	0.533 (0.35)	0.484 (0.666)	**0.494** (0.016)	**0.488** (0.008)	**0.534** (0.047)	0.441 (0.787)
Portfolios formed on Short-Term-Reversal	0.395	0.475 (0.217)	0.399 (0.921)	**0.321** (0.001)	0.399 (0.473)	0.403 (0.176)	0.416 (0.2)	**0.523** (0.026)
Portfolios formed on Long-Term-Reversal	0.476	0.483 (0.913)	0.53 (0.165)	0.442 (0.175)	0.485 (0.059)	0.486 (0.059)	0.503 (0.088)	0.535 (0.294)
Portfolios formed on Market Beta	0.407	0.332 (0.326)	0.537 (0.119)	0.45 (0.243)	**0.442** (0.002)	**0.438** (0.002)	**0.507** (0.009)	0.437 (0.671)
Portfolios formed on Variance	0.353	0.316 (0.62)	0.515 (0.066)	0.289 (0.122)	**0.402** (0.001)	**0.403** (0.00)	**0.5** (0.002)	0.458 (0.151)
Portfolios formed on Accruals	0.407	0.452 (0.464)	0.489 (0.057)	0.42 (0.466)	**0.42** (0.001)	**0.419** (0.001)	**0.449** (0.003)	0.477 (0.175)
Portfolios formed on Residual Variance	0.349	0.441 (0.205)	**0.506** (0.031)	**0.261** (0.006)	**0.388** (0.001)	**0.39** (0.00)	**0.473** (0.001)	**0.487** (0.043)
Portfolios formed on Earnings-to-Price	0.464	0.441 (0.706)	0.466 (0.958)	0.451 (0.61)	0.466 (0.389)	0.468 (0.179)	0.473 (0.33)	0.494 (0.556)
Portfolios formed on Cash-Flow-to-Price	0.456	0.522 (0.304)	0.469 (0.744)	0.451 (0.85)	0.458 (0.232)	0.459 (0.16)	0.466 (0.239)	0.527 (0.168)
Portfolios formed on Dividend Yield	0.47	0.389 (0.265)	0.521 (0.412)	0.455 (0.722)	**0.487** (0.026)	**0.484** (0.007)	**0.525** (0.013)	0.452 (0.771)
Portfolios formed on Momentum	0.359	**0.553** (0.017)	0.439 (0.06)	0.375 (0.615)	**0.381** (0.011)	**0.384** (0.004)	**0.427** (0.005)	0.508 (0.034)
Portfolios formed on Operating Profitability	0.383	0.399 (0.789)	**0.487** (0.002)	0.349 (0.095)	**0.396** (0.00)	**0.397** (0.00)	**0.424** (0.00)	0.418 (0.483)
Portfolios formed on Investment	0.451	0.447 (0.949)	0.49 (0.318)	0.45 (0.941)	**0.466** (0.001)	**0.466** (0.001)	0.489 (0.006)	0.509 (0.285)
Portfolios formed on Net Share Issues	0.351	**0.483** (0.04)	0.391 (0.387)	**0.404** (0.021)	**0.367** (0.001)	**0.364** (0.001)	**0.394** (0.005)	**0.475** (0.028)
Portfolios formed on Size and Book-to-Market	0.391	**0.547** (0.007)	**0.538** (0.00)	0.375 (0.301)	**0.418** (0.00)	**0.419** (0.00)	**0.473** (0.00)	**0.585** (0.00)
Portfolios formed on Size and Operating profitability	0.406	0.502 (0.082)	**0.494** (0.003)	**0.375** (0.036)	**0.42** (0.00)	**0.421** (0.00)	**0.456** (0.00)	**0.523** (0.021)
Norwegian Data:								
Portfolios formed on Size	1.318	1.497 (0.097)	1.268 (0.571)	**1.447** (0.044)	1.339 (0.106)	1.304 (0.183)	1.268 (0.257)	1.379 (0.488)
Portfolios formed on Book-to-Market	0.769	0.636 (0.133)	0.772 (0.973)	0.761 (0.856)	0.778 (0.382)	0.775 (0.484)	0.789 (0.559)	0.664 (0.154)
Portfolios formed on Momentum	0.777	0.814 (0.655)	0.703 (0.278)	0.801 (0.589)	0.773 (0.645)	0.767 (0.263)	0.74 (0.25)	0.791 (0.852)
Portfolios formed on Spread	1.073	1.294 (0.058)	1.117 (0.566)	**1.243** (0.007)	**1.108** (0.005)	1.08 (0.457)	1.075 (0.945)	1.212 (0.165)

P-values in parentheses. Significant Sharpe ratios at a 5%-level are marked with ***bold***

Table 5: Annualized alphas from OLS estimation computed in FFC4 and p-values from $H_0 : \alpha_k \leq \alpha_0$.

	Naive	MV	Min	Max	ERC	RP	VT	RRT
US Data:								
Portfolios formed on Size	-0.063	0.97	0.208	-0.218	0.007	0.001	0.168	0.894
		(0.397)	(0.73)	(0.084)	(0.524)	(0.113)	(0.124)	(0.316)
Portfolios formed on Book-to-Market	0.115	-1.345	0.048	-0.25	0.195	0.203	0.333	-0.137
		(0.113)	(0.905)	(0.342)	(0.142)	(0.118)	(0.218)	(0.712)
Portfolios formed on Industry	0.919	0.475	1.67	1.168	1.079	1.075	1.439	0.199
		(0.709)	(0.354)	(0.69)	(0.3)	(0.187)	(0.209)	(0.488)
Portfolios formed on Short-Term-Reversal	0.093	1.53	-0.231	**-1.066**	0.033	0.078	0.081	1.766
		(0.165)	(0.522)	(0.002)	(0.419)	(0.842)	(0.957)	(0.052)
Portfolios formed on Long-Term-Reversal	0.46	0.058	0.818	0.083	0.518	0.522	0.598	0.806
		(0.672)	(0.504)	(0.34)	(0.405)	(0.408)	(0.523)	(0.653)
Portfolios formed on Market Beta	0.094	-2.013	1.207	0.536	0.331	0.318	0.779	-0.341
		(0.084)	(0.205)	(0.413)	(0.11)	(0.081)	(0.118)	(0.66)
Portfolios formed on Variance	-0.552	-2.087	0.903	**-1.811**	-0.172	**-0.133**	**0.647**	-0.001
		(0.156)	(0.131)	(0.044)	(0.058)	(0.022)	(0.036)	(0.567)
Portfolios formed on Accruals	0.69	0.622	1.224	0.586	**0.797**	**0.797**	1.032	0.906
		(0.941)	(0.33)	(0.694)	(0.025)	(0.025)	(0.051)	(0.77)
Portfolios formed on Residual Variance	-0.382	0.189	0.955	**-2.053**	-0.123	**-0.077**	**0.572**	0.435
		(0.585)	(0.085)	(0.001)	(0.078)	(0.035)	(0.028)	(0.356)
Portfolios formed on Earnings-to-Price	0.664	-0.779	0.227	0.644	0.65	0.65	0.601	0.323
		(0.094)	(0.398)	(0.958)	(0.639)	(0.687)	(0.58)	(0.608)
Portfolios formed on Cash-Flow-to-Price	0.519	0.348	0.276	0.293	0.497	0.503	0.459	0.552
		(0.851)	(0.65)	(0.569)	(0.423)	(0.597)	(0.582)	(0.962)
Portfolios formed on Dividend Yield	0.373	-0.725	0.874	0.277	0.505	0.49	0.849	-0.026
		(0.254)	(0.494)	(0.857)	(0.166)	(0.075)	(0.07)	(0.627)
Portfolios formed on Momentum	0.39	0.485	0.445	0.10	0.317	0.395	0.437	-0.209
		(0.92)	(0.924)	(0.537)	(0.514)	(0.962)	(0.876)	(0.449)
Portfolios formed on Operating Profitability	-0.279	-0.419	**0.984**	-0.541	**-0.135**	**-0.121**	**0.181**	0.031
		(0.879)	(0.005)	(0.386)	(0.002)	(0.001)	(0.001)	(0.667)
Portfolios formed on Investment	0.492	-0.631	0.455	0.191	0.568	0.571	0.664	0.314
		(0.213)	(0.936)	(0.271)	(0.118)	(0.103)	(0.254)	(0.81)
Portfolios formed on Net Share Issues	-0.478	**1.637**	-0.225	0.028	-0.362	-0.383	-0.176	1.09
		(0.028)	(0.681)	(0.122)	(0.105)	(0.101)	(0.158)	(0.051)
Portfolios formed on Size and Book-to-Market	-0.172	1.194	**1.075**	-0.68	**0.08**	**0.09**	**0.583**	**1.851**
		(0.17)	(0.027)	(0.098)	(0.002)	(0.001)	(0.001)	(0.018)
Portfolios formed on Size and Operating profitability	-0.072	1.118	**1.013**	-0.455	**0.084**	**0.093**	**0.436**	1.42
		(0.236)	(0.033)	(0.185)	(0.007)	(0.009)	(0.002)	(0.095)
Norwegian Data:								
Portfolios formed on Size	0.113	**0.2**	0.119	**0.144**	0.116	0.109	0.104	**0.154**
		(0.00)	(0.687)	(0.001)	(0.135)	(0.053)	(0.249)	(0.005)
Portfolios formed on Book-to-Market	0.008	-0.005	0.018	0.009	0.009	0.008	0.012	-0.007
		(0.515)	(0.511)	(0.87)	(0.484)	(0.664)	(0.498)	(0.35)
Portfolios formed on Momentum	0.013	0.036	0.008	0.028	0.012	0.01	0.005	0.026
		(0.21)	(0.687)	(0.141)	(0.437)	(0.095)	(0.191)	(0.425)
Portfolios formed on Spread	0.066	**0.13**	0.084	**0.104**	**0.071**	0.066	0.066	**0.108**
		(0.00)	(0.138)	(0.00)	(0.004)	(0.816)	(0.985)	(0.008)

P-values in parentheses. Significant alphas computed in FFC4 at a 5%-level are marked with ***bold***

5.2 Adjusting for Data Mining Bias

In much of the previously conducted research on the field, the z-values from individual tests have been applied to measure the performance of each strategy. However, this method does not adjust for data mining, making the results vulnerable to data mining bias. In this thesis we gather all z-statistics and test them collectively with the null hypothesis $H_0 : \bar{f} \leq 0$. Significant values from a individual test do therefore not necessarily mean that the strategy is superior to the benchmark, due to the weaknesses in this procedure. We have therefore used the two methods of joint testing, SPA and WRC. In the WRC test we include all test-statistics, but in the SPA test we set a threshold value at -1.5. This procedure excludes all strategies with a test-statistic below -1.5.

Figure 1 shows the bootstrapped distribution of max-$\bar{z}/\bar{f}$ test statistics, while using US return data. **Figure 2** presents equivalent results using Norwegian return data. Using **Panel A** in **Figure 1** as example for the significant results: we see that the mode of the bootstrapped $\bar{z}$ test statistics is located around 2.2% per month, but varies from about 1% to about 5%. The best performing optimized strategy is marked with the dotted red line at about 4%, which is well out in the right tail and gives a p-value of 0.006 ($< \alpha = 0.05$). Thus, we reject H_0. For the cases with insignificant results, we use **Panel C** in **Figure 1** as example: The mode of the bootstrapped alphas computed with FFC4 is located around 3% per month, but varies from about 0.3% to about 6.5%. The best performing strategy is located around 3.2%, and provides a p-value of 0.468 ($> \alpha = 0.05$). Thus, we fail to reject H_0.

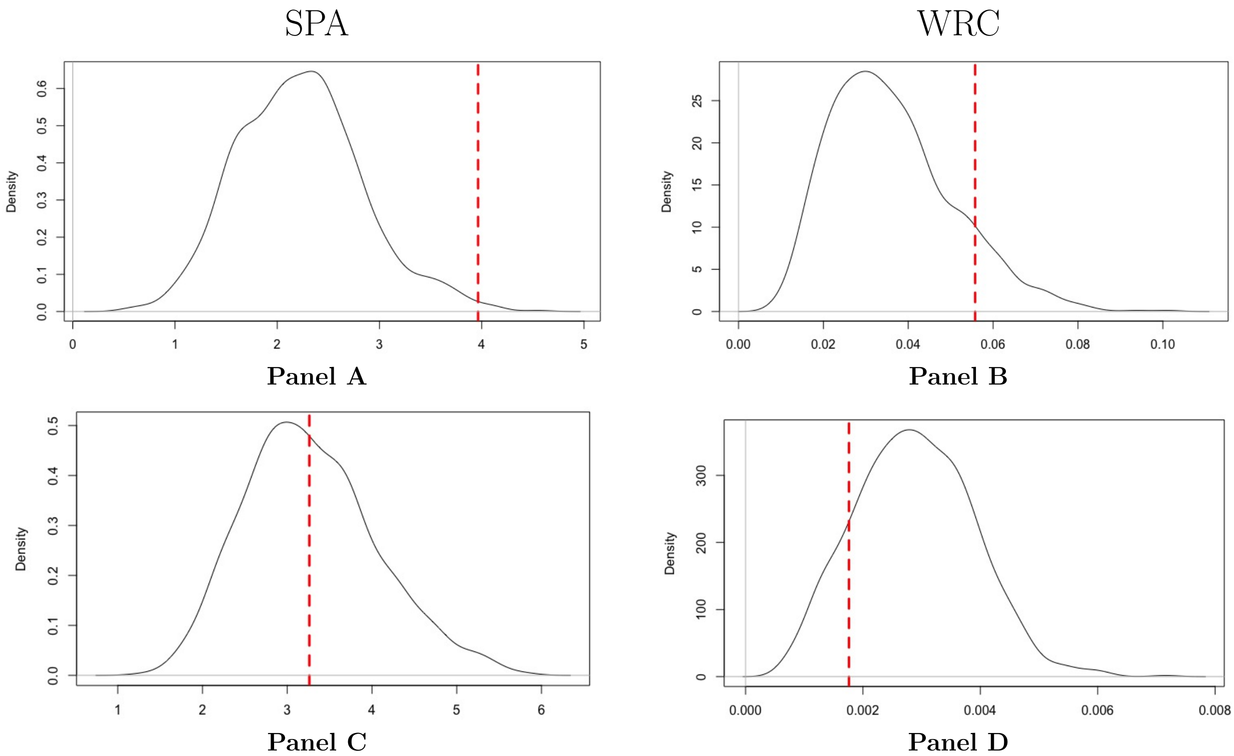

Figure 1: Panel A and C present the results from the SPA test, which is based on z-statistics. Panel B and D present the results from the WRC test, which is based on f-statistics. The panels plot a bootstrapped distribution of estimates of the max-$\bar{z}/\bar{f}$ statistics. The results are obtained by applying monthly return data from 18 US datasets in the time period July 1963 to December 2019. X-axis presents the monthly performance of the test statistics, while the Y-axis presents the probability distribution. The dotted red line shows the location of the best performing optimized strategy. R denotes the number of resamples in the bootstrap.

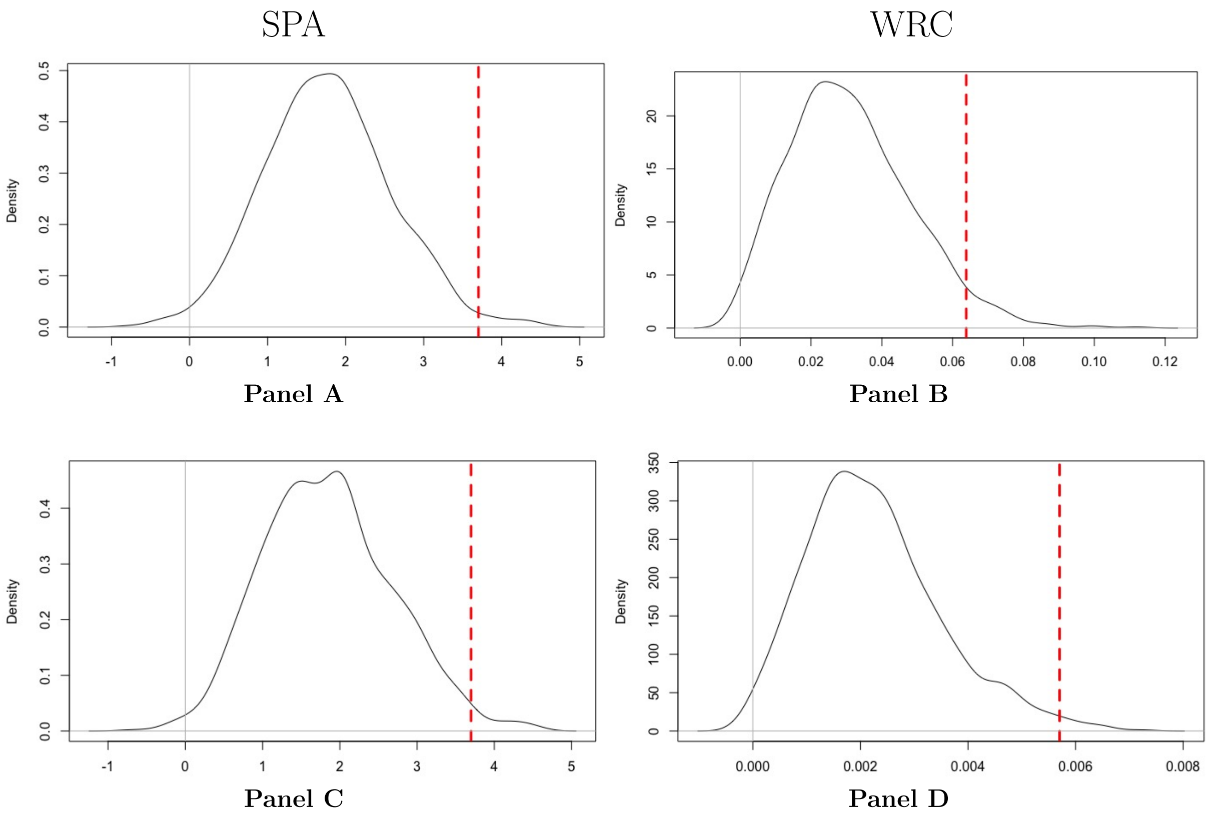

Figure 2: Panel A and C present the results from the SPA test, which is based on z-statistics. Panel B and D present the results from the WRC test, which is based on f-statistics. The panels plot a bootstrapped distribution of estimates of the max-$\bar{z}/\bar{f}$ statistics. The results are obtained by applying monthly return data from four Norwegian datasets in the sample period September 1981 to December 2019. X-axis presents the monthly performance of the test statistics, while the Y-axis presents the probability distribution. The dotted red line shows the location of the best performing optimized strategy.

Table 6: P-values from SPA and WRC.

	US		Norway	
Datasets	18		4	
	SR	Alpha	SR	Alpha
SPA	**0.006**	0.468	**0.031**	**0.015**
WRC	0.101	0.857	**0.039**	**0.013**

*Significant p-values are marked with **bold**.*

The results from **Figure 1** and **2** are presented in **Table 6**. From the table we see that the p-values from the SPA test are significant in all instances, except alpha FFC4 in US. The p-values from the WRC test are significant in the Norwegian data, but not in the US data. The number of Norwegian datasets is important to notice when considering the results.

There is an issue concerning the SPA test being dependent on t-statistics (z-statistics). When returns are highly correlated ($\rho \rightarrow 1$), the standard error of estimation located in the denominator of a t-statistic becomes substantially small Eq.(22 and 25). This results in a very high t-statistic, which means that significant differences may not be economically significant (i.e. two portfolios performs economically insignificant, although the p-values suggest a significant difference). This phenomenon is present in **Table 4**, were we see that for instance, in the US dataset formed in accruals, the ERC strategy significantly outperforms the $1/N$ strategy at a 5%-level. However, the differences is Sharpe ratio are marginal, namely 0.013 (ERC = 0.42 vs $1/N = 0.407$).

Table 7 shows the correlation between the returns from optimized strategies and the benchmark. The results from this table suggests that high correlations are present, with the majority of the correlations above 0.9. ERC and RP have the highest correlations among the strategies with most correlations above 0.999. This suggests that the results provided by the SPA test might be influenced by the high correlation in returns, wrongly causing significant results. WRC, however, is not dependent on t-statistics and thus not influenced by the high correlations. We therefore emphasize the results provided by the WRC test.

Based on the results provided by the WRC test extracted from **Table 6**, we draw the following conclusion: There is evidence suggesting that the best optimized strategy in the US data fails to outperform the $1/N$ strategy (fail to reject H_0 at

5%). In the Norwegian data, however, there is significant evidence suggesting that the best optimized strategy outperform the $1/N$ strategy (reject H_0 at 5%).

Table 7: Correlations between optimized strategies and the naive strategy and p-values from $H_0 : \rho = 0$.

	MV	Min	Max	ERC	RP	VT	RRT
US Data:							
Portfolios formed on Size	**0.858** (0.00)	**0.905** (0.00)	**0.975** (0.00)	**0.999** (0.00)	**0.999** (0.00)	**0.99** (0.00)	**0.902** (0.00)
Portfolios formed on Book-to-Market	**0.9** (0.00)	**0.962** (0.00)	**0.983** (0.00)	**1.00** (0.00)	**1.00** (0.00)	**0.996** (0.00)	**0.939** (0.00)
Portfolios formed on Industry	**0.79** (0.00)	**0.862** (0.00)	**0.94** (0.00)	**0.996** (0.00)	**0.998** (0.00)	**0.968** (0.00)	**0.822** (0.00)
Portfolios formed on Short-Term-Reversal	**0.895** (0.00)	**0.967** (0.00)	**0.987** (0.00)	**0.999** (0.00)	**0.999** (0.00)	**0.994** (0.00)	**0.918** (0.00)
Portfolios formed on Long-Term-Reversal	**0.892** (0.00)	**0.962** (0.00)	**0.985** (0.00)	**0.999** (0.00)	**0.999** (0.00)	**0.994** (0.00)	**0.92** (0.00)
Portfolios formed on Market Beta	**0.852** (0.00)	**0.822** (0.00)	**0.965** (0.00)	**0.997** (0.00)	**0.997** (0.00)	**0.963** (0.00)	**0.871** (0.00)
Portfolios formed on Variance	**0.855** (0.00)	**0.802** (0.00)	**0.956** (0.00)	**0.995** (0.00)	**0.995** (0.00)	**0.942** (0.00)	**0.864** (0.00)
Portfolios formed on Accruals	**0.907** (0.00)	**0.953** (0.00)	**0.993** (0.00)	**1.00** (0.00)	**1.00** (0.00)	**0.995** (0.00)	**0.933** (0.00)
Portfolios formed on Residual Variance	**0.864** (0.00)	**0.866** (0.00)	**0.974** (0.00)	**0.997** (0.00)	**0.997** (0.00)	**0.965** (0.00)	**0.882** (0.00)
Portfolios formed on Earnings-to-Price	**0.9** (0.00)	**0.963** (0.00)	**0.983** (0.00)	**1.00** (0.00)	**1.00** (0.00)	**0.998** (0.00)	**0.937** (0.00)
Portfolios formed on Cash-Flow-to-Price	**0.895** (0.00)	**0.962** (0.00)	**0.982** (0.00)	**1.00** (0.00)	**1.00** (0.00)	**0.998** (0.00)	**0.933** (0.00)
Portfolios formed on Dividend Yield	**0.868** (0.00)	**0.902** (0.00)	**0.958** (0.00)	**0.998** (0.00)	**0.999** (0.00)	**0.988** (0.00)	**0.899** (0.00)
Portfolios formed on Momentum	**0.895** (0.00)	**0.967** (0.00)	**0.987** (0.00)	**0.999** (0.00)	**0.999** (0.00)	**0.994** (0.00)	**0.918** (0.00)
Portfolios formed on Operating Profitability	**0.895** (0.00)	**0.967** (0.00)	**0.987** (0.00)	**0.999** (0.00)	**0.999** (0.00)	**0.994** (0.00)	**0.918** (0.00)
Portfolios formed on Investment	**0.9** (0.00)	**0.963** (0.00)	**0.992** (0.00)	**1.00** (0.00)	**1.00** (0.00)	**0.995** (0.00)	**0.925** (0.00)
Portfolios formed on Net Share Issues	**0.895** (0.00)	**0.947** (0.00)	**0.987** (0.00)	**0.999** (0.00)	**1.00** (0.00)	**0.994** (0.00)	**0.92** (0.00)
Portfolios formed on Size and Book-to-Market	**0.916** (0.00)	**0.962** (0.00)	**0.994** (0.00)	**0.999** (0.00)	**0.999** (0.00)	**0.994** (0.00)	**0.935** (0.00)
Portfolios formed on Size and Operating profitability	**0.923** (0.00)	**0.977** (0.00)	**0.995** (0.00)	**1.00** (0.00)	**1.00** (0.00)	**0.997** (0.00)	**0.935** (0.00)
Norwegian Data:							
Portfolios formed on Size	**0.835** (0.00)	**0.887** (0.00)	**0.942** (0.00)	**0.998** (0.00)	**0.998** (0.00)	**0.97** (0.00)	**0.891** (0.00)
Portfolios formed on Book-to-Market	**0.876** (0.00)	**0.905** (0.00)	**0.969** (0.00)	**0.998** (0.00)	**0.999** (0.00)	**0.98** (0.00)	**0.914** (0.00)
Portfolios formed on Momentum	**0.89** (0.00)	**0.928** (0.00)	**0.969** (0.00)	**0.999** (0.00)	**0.999** (0.00)	**0.984** (0.00)	**0.913** (0.00)
Portfolios formed on Spread	**0.799** (0.00)	**0.911** (0.00)	**0.942** (0.00)	**0.998** (0.00)	**0.999** (0.00)	**0.984** (0.00)	**0.851** (0.00)

P-values in parentheses. Significant correlations at a 5%-level are marked with ***bold***

6 | Discussion

DeMiguel et al. (2009), Zakamulin (2017) and several others have demonstrated that optimized strategies fail to consistently outperform the $1/N$ strategy with statistical significance. However, a number of papers like Kritzman et al. (2010) and Kirby and Ostdiek (2012) have argued in defence of the optimized strategies, claiming their superiority. The evidence on either side of the debate have mainly been based on individual tests, t-statistics and Sharpe ratio. The t-statistics are, however, questionable in the research of portfolio performance, due to the presence of high correlation in returns (presented in **Table 7**). Individual hypothesis testing is also a subject for criticism, due to its vulnerability to data mining bias. Sharpe ratio is a questionable performance measurement, due to its limitations mentioned in Section **4.2.1**. This issue might be what causes the following difference in Sharpe ratio and alpha computed in FFC4: The majority of the Sharpe ratios presented in **Table 4** are insignificant when tested with $H_0 : SR_k \leq SR_0$. However, the number of significant observations is still about 120% higher than in **Table 5** ($H_0 : \alpha_k \leq \alpha_0$).

We aim to cope with these issues by applying SPA and WRC test, in addition to alpha computed in FFC4. However, the SPA test is based on t-statistics, which makes it vulnerable to high correlation in returns. The results obtained through WRC are therefore emphasized.

From the WRC test in **Table 6** we see that Sharpe ratio and alpha computed in FFC4 are insignificant when using US return data. These results support the null hypothesis H_0 at a 5%-level, indicating that the best optimized strategy fails to outperform the $1/N$ strategy in the US market. However, when we apply Norwegian return data to WRC, both Sharpe ratio and alpha FFC4 yields significant p-values. Thus, we obtain evidence suggesting that we reject the null hypothesis H_0 at a 5%-level. This implies that the best optimized strategy outperform the benchmark in the Norwegian market. Comparing the results obtained with US data and Norwegian data, we observe that there is a clear difference in the outcomes.

What causes this difference is not explained by our results, but we have some ideas of what the cause could be. The size of the economies may affect the market efficiency, since US have a significantly larger economy than Norway, the stocks in Norway might be more exposed to "wrong pricing". The number of datasets included may also have an impact. However, these are only guesses with no scientific grounding.

The use of joint test is a relatively new and pristine method of processing data, although there are some papers that already have applied this methodology, like Hsu et al. (2018) and Yang et al. (2019). These papers applied both SPA and WRC in their research and arrived at the same conclusion as our paper when using US return data. Namely, that there is no evidence suggesting that optimized strategies are superior to the $1/N$ strategy. We were unable to find papers conducting a similar study with Norwegian data. Thus, we find our study to extend previous research by applying joint testing methodology to Norwegian return data, provided by Bernt Arne Ødegaard. We also use a very recent time sample, that we are yet to observe in other studies.

The low-volatility anomalies in Kenneth French's datasets, pointed out by Zakamulin (2017), may affect the results provided by the Sharpe ratio. Alpha computed in FFC4 adjusts for the low-volatility with the HML factor, correcting for this issue. In this case, however, the results from Sharpe ratio and alpha FFC4 were equal in terms of significance. Thus, we chose to ignore the flaws in Sharpe ratio when considering our results. We also believe it is important to mention the few number of Norwegian datasets included, which was inevitable due to the limitations in accessible datasets.

The cost of rebalancing the portfolio each month and other expenses related to a real-market scenario are ignored in this thesis. This limits our results to only apply in a theoretical market scenario, however, we still believe that our contribution is valuable in the study of portfolio performance and the discussion initiated by DeMiguel et al. (2009).

Going forward, there are several interesting aspects to further investigate, which was out of the scope of this thesis. Adding more optimized strategies and adjusting the time horizon may give interesting results. Applying other datasets could also yield new interesting insight. We limited our study to Carhart's four-factor model. Expanding the research to include Fama-French five research factors would also be an interesting extension in future studies.

7 | Conclusion

The motivation behind our thesis has been a long-standing debate within academia, on whether optimized portfolios can significantly outperform the $1/N$ strategy, initiated by DeMiguel et al. (2009). We noticed that there are several weaknesses in many of the papers contributing to this debate, which we aimed to correct in our thesis. Mainly, the use of individual hypothesis tests and Sharpe ratio as the only measurement.

In this thesis, we have measured the performance of seven optimized strategies relative to the $1/N$ strategy, adjusting for data mining bias. The issues of data mining are corrected by applying the joint test WRC. The results obtained in our thesis were two-folded: First, using return data from the US market, the WRC test provides insignificant results at 5%-level from both Sharpe ratio and alpha FFC4. This indicates that the best optimized strategy fails to outperform the $1/N$ strategy in the US market. Thus, we fail to reject H_0 when using US return data. Second, the WRC test conducted with Norwegian data provides, however, significant results. Both Sharpe ratio and alpha FFC4 provides significant p-values at a 5%-level, which indicates that the best optimized strategy is statistically significant superior to the $1/N$ strategy. Thus, we reject H_0 when using Norwegian return data.

What causes the difference in results when applying US return data opposed to Norwegian return data, are not explained by the results obtained in our thesis. However, a possible explanation may be the differences in market efficiency in the respective countries, causing stock prices to deviate from their "real" value based on all available information.

References

Asness, C. S., Frazzini, A., & Pedersen, L. H. (2012). Leverage aversion and risk parity. *Financial Analysts Journal*, *68*(1), 47–59.

Auer, B. R., & Schuhmacher, F. (2013). Performance hypothesis testing with the Sharpe ratio: The case of hedge funds. *Finance Research Letters*, *10*(4), 196–208.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, *52*(1), 57–82.

Chaves, D., Hsu, J., Li, F., & Shakernia, O. (2011). Risk parity portfolio vs. other asset allocation heuristic portfolios. *Journal of Investing*, *20*(1), 108–118.

Choueifaty, Y., & Coignard, Y. (2008). Toward maximum diversification. *Journal of Portfolio Management*, *35*(1), 40–51.

Clarke, R. G., De Silva, H., & Thorley, S. (2006). Minimum-variance portfolios in the us equity market. *Journal of Portfolio Management*, *33*(1), 10–24.

Cogneau, P., & Zakamouline, V. (2013). Block bootstrap methods and the choice of stocks for the long run. *Quantitative Finance*, *13*(9), 1443–1457.

DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial studies*, *22*(5), 1915–1953.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*, 3–56.

Gerber, A., & Hens, T. (2006). Modelling alpha-opportunities within the CAPM. *Available at SSRN 966279*.

Goldberg, L. R., Leshem, R., & Geddes, P. (2013). Restoring value to minimum variance. *Forthcoming in Journal of Investment Management*.

Hall, P. (1985). Resampling a coverage pattern. *Stochastic processes and their applications*, *20*(2), 231–246.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics*, *23*(4), 365–380.

Harvey, C. R., & Liu, Y. (2015). Backtesting. *Journal of Portfolio Management*, *42*(1), 13–28.

Hsu, P.-H., Han, Q., Wu, W., & Cao, Z. (2018). Asset allocation strategies, data snooping, and the 1/n rule. *Journal of Banking & Finance*, *97*, 257–269.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of finance*, *48*(1), 65–91.

Jobson, J. D., & Korkie, B. M. (1981). Performance hypothesis testing with the Sharpe and Treynor measures. *Journal of Finance*, *36*(4), 889–908.

Kirby, C., & Ostdiek, B. (2012). It's all in the timing: simple active portfolio strategies that outperform naive diversification. *Journal of Financial and Quantitative Analysis*, *47*(2), 437–467.

Kosowski, R., Timmermann, A., Wermers, R., & White, H. (2006). Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis. *Journal of Finance*, *61*(6), 2551–2595.

Kritzman, M., Page, S., & Turkington, D. (2010). In defense of optimization: the fallacy of 1/n. *Financial Analysts Journal*, *66*(2), 31–39.

Kunsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 1217–1241.

Leamer, E. E. (1978). *Specification searches: Ad hoc inference with nonexperimental data* (Vol. 53). Wiley New York.

Leamer, E. E. (1983). Let's take the con out of econometrics. *American Economic Review*, *73*(1), 31–43.

Maillard, S., Roncalli, T., & Teiletche, J. (2010). The properties of equally weighted risk contribution portfolios. *Journal of Portfolio Management*, *36*(4), 60–70.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, *7*(1), 77–91.

Mausser, H., & Romanko, O. (2014). Computing equal risk contribution portfolios. *IBM Journal of Research and Development*, *58*(4), 5–1.

Memmel, C. (2003). Performance hypothesis testing with the Sharpe ratio. *Finance Letters*, *1*(1).

Merton, R. C. (1972). An analytic derivation of the efficient portfolio frontier. *Journal of Financial and Quantitative Analysis*, *7*(4), 1851–1872.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, *8*(4), 323–361.

Patton, A., Politis, D. N., & White, H. (2009). Correction to "automatic block-length selection for the dependent bootstrap" by D. Politis and H. White. *Econometric Reviews*, *28*(4), 372–375.

Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical association*, *89*(428), 1303–1313.

Politis, D. N., & White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, *23*(1), 53–70.

Sharpe, W. F. (1966). Mutual fund performance. *Journal of Business*, *39*(1), 119–138.

White, H. (2000). A reality check for data snooping. *Econometrica*, *68*(5), 1097–1126.

Yang, J., Cao, Z., Han, Q., & Wang, Q. (2019). Tactical asset allocation on technical trading rules and data snooping. *Pacific-Basin Finance Journal*, *57*.

Zakamulin, V. (2017). Superiority of optimized portfolios to naive diversification: Fact or fiction? *Finance Research Letters*, *22*, 122–128.

Appendix

Reflection Note 1

This reflection note is written as a part of the master program in Business Administration at University of Agder. The objective of this reflection note is to preset an overview of the insight I have gathered throughout my studies and in particular the master thesis. The thesis has been written by Haakon Sebastian Olviken in companionship with me.

Markowitz (1952) introduced the mean-variance model that laid the foundation of the modern portfolio theory (MTP). This work actualized the research of optimized strategies since the 50's Lately, DeMiguel et al. (2009) initiated a debate regarding optimized strategies, by presenting evidence that optimized strategies fail to outperform the naïve diversification strategy. Kritzman et al. (2010), (Kirby & Ostdiek, 2012) and several others have later defended the outperformance of optimized strategies. However, there are several limitations with the papers mentioned; they use Sharpe ratio as performance measurement and single hypothesis test to support/reject their null hypothesis. If the returns are not normally distributed, the results may be deceiving when applying Sharpe ratio. Data snooping have been a known phenomenon in the research of strategies, causing the researcher to falsely discover profitable strategies (Type I Error). Single hypothesis tests do not correct for this issue, making it a vulnerable test. Yang et al. (2019) and (Hsu et al., 2018) solved these issues by implementing alpha motivated by Carhart (1997) four-factor model (FF4) and joint hypothesis testing.

We aim to contribute to this discussion with our thesis, by applying the methods used in the papers by Yang et al. (2019) and Hsu et al. (2018) to new datasets. Namely, joint hypothesis testing and bootstrapping. Our paper concerns the performance of seven optimized portfolio strategies relative to a passive benchmark (naïve diversification). The analysis of the performance is conducted with US-/and

Norwegian datasets, consisting of portfolio returns. The naïve-diversification strategy is independent of any estimates based on future returns or other uncertainties, which makes it preferable benchmark. The results provided in thesis support the findings from Yang et al. (2019) and Hsu et al. (2018) in the US data, providing evidence that optimized strategies fail to outperform the naïve benchmark. However, our results suggest that optimized strategies are significant superior when using Norwegian data. Our results fail to explain the reason for this difference, but market efficiency in the respective markets may be a part of the explanation.

International Trends

We analyze data representing stock returns from US and Norway, which is deeply influenced by the international economics. The state of the world economics affects both the market factors and stocks returns, making the data applied in our thesis heavily dependent on international economic trends. The trends within the financial market is formally known as bull (rising) and bear (recession) markets. The trend in the last decade (since the financial crisis) has been low interest rates, even negative in some instances, causing many to invest in the stock market. This is one of several specific trends influencing both the Norwegian and US stock markets, within the sample period; US: 1963-2019, Norwegian: 1981-2019.

Innovation

The thesis provides methods that are "state of the art" within the research of optimized strategies performance, namely bootstrapping and the joint tests Reality Check from White (2000) (WRC) and Superior Predictive Availability from Hansen (2005) (SPA). Previously, researchers tested the optimized strategies individually, which was related to great probability of wrongly discovering superior strategies. Quick implementation of new methodology is important in this line of research, in order to achieve broad academic foundation for new findings and correct errors. Going forward, continuous exploration of new optimized strategies may also lead new discoveries within this field. Sharpe ratio and FF4 are both acknowledged performance measures, but they also have known flaws. Development of more precise performance measures that improves upon the flaws of these models, would probably contribute to a better understanding of which strategies that are most profitable. We use the advanced statistical programming language R to conduct our research, which allows to quickly adopt new methodology. The optimized strategies (algorithms) are also getting more complex as new techniques are developed, which makes this an interesting topic to follow. The "easy" access to large amounts of data online also helps speed the process of testing the performance of newly emerged strategies.

Responsibility

There are several aspects concerned with responsibility that has been raised during the production of the thesis. In the writing process we have been very careful in referencing all sources used, such that we do not take any credit for work conducted by others. Data snooping is a known phenomenon in the research of profitable strategies, leading to false discoveries of profitable strategies. I have taken a number of precautions to cope with the issue of data snooping in our thesis; (i) Picking 18 arbitrary US datasets to avoid selecting preferable datasets. However, this was not possible with the Norwegian data, due to the limitation of 4 available datasets. (ii) WRC and SPA are tests that are implemented to cope with the issue of false discoveries. Although we are trying to reduce the likelihood of errors in our discoveries, there is still a 5% chance of false discoveries, caused by the test procedure (p-value $> \alpha = 0.05$).

There is also a well-known fact that large financial corporation have self-interest in the discovery of profitable strategies that beats the benchmark, in order to make customers pay for the active ones. This may give resistance to papers demonstrating that "passive" strategies achieve the same performance as active strategies. We have tried our best to be independent of any outside influence towards our study, but we cannot exclude the possibility that we have been somewhat influenced.

Summary

The thesis has been a great possibility for me to apply all the knowledge I have accumulated through the master's program. I believe that the studies have made ready to enter work life, with enough resources to be successful in a high competent profession. Working with the master thesis can for sure be all-consuming at times, making it easy to forget the wider impact. Discussing the three concepts international trends, innovation and responsibility in this reflection note truly helped me gain a wider understanding of the related repercussions. In particular the responsibility related to academic work is something that I am deeply committed to, as research is so much about discovering the truth. Intentionally biased research is weakening the reputation of academia, as well as staging the progression in research. I believe that continuous research, leading to increased knowledge is beneficial to the society, making our contribution important in that sense as well.

Kristiansand, 02.06.2020

Reflection Note 2

We have evaluated the performance of seven optimized strategies, relative to the naive diversification strategy. We use the naive strategy as benchmark, since it is easy to implement and is commonly used as benchmark among researchers within the field. (DeMiguel et al., 2009). The strategy allocate wealth equally among N assets, and is not affected by estimation errors. We use advanced methods, namely, joint testing and bootstrapping to test the optimized strategies collectively. In order to handle data snooping bias, also referred to as false discovery, we implement White's reality check (WRC) and superior predictive ability test (SPA) to correct this issue. Introduced by White (2000) and (Hansen, 2005). We evaluated the performance of the strategies across datasets we have gathered through online libraries provided by Kenneth French and Bernt Arne Ødegaard. The datasets consist of monthly returns form the US and Norwegian market. We use 18 datasets from the US market, with the monthly returns from the period 1963 to 2019. For the Norwegian market we use four data sets, with monthly return from the period 1981 to 2019. Our objective with this thesis is to contribute to the ongoing debate regarding optimized strategies initiated by DeMiguel et al. (2009). The debate addresses whether optimized strategies consistently outperform the naive diversification or not. With our empirical studies, we have provided new evidence that contribute to the debate, with the methods used in this thesis. We extend previous studies by including more datasets with a longer period from the US- and Norwegian market. Our thesis conclude that optimized strategies do not consistently outperform the naive diversification strategy (benchmark).

Further, my reflection note will include a reflection of internationalisation, innovation and responsibility. These factors are the key concepts in the School of Business and Law's mission statement and strategy.

Internationalisation

Our thesis is subject to several international factors. First, the thesis utilize historical data with monthly returns of the US market, provided by Kenneth French's online library. In addition, we use data from the Norwegian market return. Both are used in the computation of generating empirical results. The market data we use is dependent on international movements in the global economy. Which will affect the results of our thesis.

There are researchers worldwide that provide us with new perspectives upon the subject. We use terminologies that are used globally to assure our contribution regarding the subject is understood everywhere. We naturally adapt our research to other researchers by writing in English, and by providing an empirical approach of our conducted demonstration. This ensure that whomever would be capable to reassess our results, assuming they have access to the necessary tools. Our results can therefore be tested on an international level.

Innovation

In our thesis regarding modern portfolio theory, we involve great numbers of data in the research process. Our thesis is categorized as quantitative, our approach have to handle a lot of data. We use the latest technology available to produce results. We use the data program R to test our methods, and present our thesis using latex. We use different combination of methods in our approach, involving joint testing and the use of bootstrapping. The results we achieve with these methods and the new data (market returns from 2019/2020) we have gathered, will give a result we have never seen before. Hopefully our results will be a valid contribution to research on the subject.

Since the first optimized strategy was developed in the 1950s, a lot of effort have been put into improving the allocation strategy. In the last decade, modern portfolio theory have received a lot of attention, and new strategies have emerged. This is a result of the volatile global markets, leading investors to evaluate capital allocation more than ever before. In recent litterateur the value of allocation strategies are debated, whether optimal strategies add value or not to the portfolio. With our approach and methods, we argue that optimized strategies do not consistently add value. The contribution we have made can provide new insight to the ongoing debate. Our evidence for the discovery will be published, and available for everyone to read.

The technology we have access to, gives us more data power, and allow us to compute number and data much faster. The rapid development of technology and data power can result in new research methods emerging due to the technological advancement. Exploring various markets can give new insight to the ongoing discussion. We observe that the majority of researcher use market data from the US. We used data from the Norwegian market and the US market, to understand how different results variate in separate markets, not to compare the two individual markets.

Responsibility

Optimized strategies primarily focuses on how to allocate capital among assets, in order to achieve high return. The methods neglect other information characterizing the asset. Our thesis is purely based on a theoretical representation of the market return. Our studies do not suggest an approach for ethical trading, however this is an important topic to consider before allocating wealth.

If an investor intend to implement these strategies in a functional financial market, one have to consider the cost of maintaining an optimized portfolio strategy. This relates especially to portfolio managers, who manage other people's wealth. Our thesis suggest that optimization do not add any particular value. One can therefor question if the additional cost of running an optimized portfolio is worth it, or if it is ethical. Based on preferences from each individual investor, the investor should consider what they believe are necessary in a portfolio, in order for them to make investments.

Our contribution consist of new information and previous knowledge. The knowledge we have obtained while conducting our research have clearly been referred to, through the entire thesis. We are responsible for the information in the thesis, and theory we have gathered from other researchers will be accredited by referring to the original authors.

Summary and Conclusion

This final section is brief summarization and conclusion on my discussion. The courses we had during our master degree have equipped us with the necessary tools in order for us to write this master thesis. With advanced methodological approaches, we want to make a contribution to the ongoing debate initiated by DeMiguel et al. (2009). We have to adapted our presentation of our thesis, so everyone can understand our contribution. We used market data on a national and international level, to provide evidence for our empirical results. The data we use are affected by changes in the global economy, thus influencing our result. Using the latest of technology gives us more data power, and allow us to use new approaches to test our hypothesis. This produce results, that give new insight to portfolio optimization. Everyone can construct portfolios using the strategies we have evaluated, it is not said that only corporations can use the strategies to invest. However, the strategies primarily focus on how to allocate wealth, therefore investors have to make their own personal opinion on which investments to make.

Kristiansand, 02.06.2020